

Generative AI Chatbots in ESLEFL Education. A Systematic Review of Educational Opportunities, Learning Outcomes, Challenges, and Future Directions (2021–2026)

Generative AI Chatbots in ESL/EFL Teaching: A Systematic Review of Educational Opportunities, Learning Outcomes, Challenges, and Future Perspectives (2021–2026)

Marlon Rubén Barcia Moreira¹ marlon.barcia@unesum.edu.ec <https://orcid.org/0000-0001-6170-195X>

María Eugenia Ampuero Intriago² maria.ampuero@utm.edu.ec <https://orcid.org/0000-0003-2206-3187>

Letty Jaqueline Saltos Rodríguez³ letty.saltos@utm.edu.ec <https://orcid.org/0000-0003-2900-5289>

Celene Margarita Casierra Párraga⁴ celene.casierra@utm.edu.ec <https://orcid.org/0000-0003-0473-4207>

Martha Narcisa Loor Fernández⁵ martha.loor@utm.edu.ec <https://orcid.org/0000-0003-2013-4543>

Abstract

GenAI (Generative Artificial Intelligence) chatbots have transformed English as a Second Language (ESL) education and English as a Foreign Language (EFL) education to provide personalized learning, feedback, and interaction with students in both fields. However, there is a lack of research in the field that supports such approaches in education. In this systematic review, we sought to find and analyze existing research on the use of generative AI chatbots in ESL/EFL education, including the educational opportunities and learning outcomes, implementation challenges, emerging trends, research gaps, and future directions. We conducted a systematic

¹ Universidad Estatal del Sur Manabí. Ecuador.

² Universidad Técnica de Manabí – Instituto de Lenguas Modernas. Ecuador.

³ Universidad Técnica de Manabí – Instituto de Lenguas Modernas. Ecuador.

⁴ Universidad Técnica de Manabí – Instituto de Lenguas Modernas. Ecuador.

⁵ Universidad Técnica de Manabí – Instituto de Lenguas Modernas. Ecuador.

exploratory study based on Scopus and Web of Science to identify and exclude questions and answers and to qualitatively synthesize a number of peer-reviewed literature. We found that generative AI chatbots can enhance writing, grammar, vocabulary, speaking, engagement of learners, autonomy, and personalized feedback in the classroom and are particularly effective in teacher-driven education. But issues of academic integrity, algorithmic bias, data privacy, unequal access, AI literacy, and teacher professional development are still hurdles in fulfilling this and are still preventing sustainable implementation. Generative AI chatbots are more of a partner than a substitute for teachers and need to be integrated with human-centered education and ethical governance, and evidence-based teaching strategies to make them the best classroom assistants in order to promote the teaching and learning of the English language in the future.

Resumen

Los chatbots de GenAI (Inteligencia Artificial Generativa) han transformado la educación de inglés como segundo idioma (ESL) y la educación de inglés como lengua extranjera (EFL) para proporcionar un aprendizaje personalizado, retroalimentación e interacción con los estudiantes en ambos campos. Sin embargo, hay una falta de investigación en el campo que respalde tales enfoques en la educación. En esta revisión sistemática, buscamos encontrar y analizar la investigación existente sobre el uso de chatbots de IA generativa en la educación de ESL/EFL, incluyendo las oportunidades educativas y los resultados de aprendizaje, los desafíos de implementación, las tendencias emergentes, las lagunas de investigación y las direcciones futuras. Realizamos un estudio exploratorio sistemático basado en Scopus y Web of Science para identificar y excluir preguntas y respuestas y para sintetizar cualitativamente una serie de literatura revisada por pares. Encontramos que los chatbots de IA generativa pueden mejorar la escritura, la gramática, el vocabulario, la expresión oral, la participación de los estudiantes, la autonomía y la retroalimentación personalizada en el aula, y son particularmente efectivos en la educación dirigida por el profesor. Pero los problemas de integridad académica, sesgo algorítmico, privacidad de datos, acceso desigual, alfabetización en IA y desarrollo profesional de los docentes siguen siendo obstáculos para lograr esto y aún impiden una implementación sostenible. Los chatbots de IA generativa son más un socio que un sustituto de los maestros y

necesitan integrarse con una educación centrada en el ser humano y una gobernanza ética, así como con estrategias de enseñanza basadas en evidencia para convertirlos en los mejores asistentes de aula y promover la enseñanza y el aprendizaje del idioma inglés en el futuro.

Keywords: AI chatbots; ESL/EFL education; Generative artificial intelligence; Language learning outcomes; Systematic review.

Palabras clave: Educación ESL/EFL; chatbots de IA; inteligencia artificial generativa; resultados del aprendizaje de idiomas; revisión sistemática.

Introduction

The rapid development of generative artificial intelligence (GenAI) is revolutionising education and has allowed for more adaptable, interactive and personalized learning experiences in a wide range of domains. Generative AI chatbots such as ChatGPT in English as a Second Language (ESL) and English as a Foreign Language (EFL) education are very powerful in supporting language practice, providing immediate feedback, facilitating autonomous learning and engaging learners. The ability to generate context-sensitive responses, mimic real-life communication and personalize teaching support has positioned these technologies at the forefront of contemporary computer-assisted language learning (CALL), and research from 2021 to 2026 has been highlighting the benefits of AI chatbots in teaching language skills and raising questions about their pedagogical development, ethical design, teacher preparedness and educational sustainability (Guan et al., 2024; Kundu & Bej, 2025; Law, 2024; Liu et al., 2026).

Despite a growing body of empirical research on generative AI chatbots in ESL/EFL education, the literature is still fragmented and covers multiple educational domains, language skills, technological platforms, and methodologies. While many studies have observed more productive writing, speaking, vocabulary acquisition, and learner autonomy, there are also concerns related to academic integrity, algorithmic bias, data privacy, unequal technological access, and teacher training issues. Moreover, the rapid development of chatbot technologies has far surpassed the development of evidence to integrate the educational opportunities in the context of pedagogical obstacles and implementation challenges. As a

result, teachers, researchers, and policymakers lack a well-rounded view of how to use generative AI chatbots in English language education to increase learning success and provide ethical, equitable, and sustainable teaching (Law, 2024; Steele, 2023; Tan et al., 2025).

Recent studies confirm that generative AI chatbots enhance language learning through individualized instruction, adaptive feedback, and learner-centered pedagogies. The evidence is unanimous: AI-enabled teaching systems not only enhance the language skill development across the four language domains but also increase motivation, engagement, vocabulary acquisition, grammatical accuracy, and self-directed learning for students. AI tools help students receive immediate answers, personalized explanations, and continuous practice in an environment where students can't work consistently in the classroom. However, AI is not to replace teachers, who are very much needed to help students scaffold learning, track progress, and engage in meaningful relationships (Kundu & Bej, 2025; Liu et al., 2026; Torres & Kahveci, 2025).

One of the most prominent trends in teaching about AI chatbots is that of second language writing. The researchers found that students are more likely to use ChatGPT during brainstorming, idea generation, drafting, and language revision with better writing skills, grammatical accuracy, and confidence in composing. Similarly, prompt engineering has been shown to significantly influence the quality of AI feedback, as well as that the right prompts are intended for learners and are in the context of learners' linguistic ability to compose and in educational contexts, so guidance on revision is better. This evidence suggests that the effectiveness of teaching is not just a function of chatbots, but also of the way teachers, students, and AI systems are taught to interact with each other (Algobaei & Alzain, 2026; Hwang et al., 2025).

Research has also found that generative AI has an impact on grammar acquisition and language accuracy when integrated into teacher-led instruction. Instead of being a tool for teaching students as a self-sufficient tutor, AI chatbots are best equipped to teach students through structured guidance, reflective activities and consistent support for a learning environment. The research on English tense acquisition shows that guided AI-assisted

learning has proven superior performance to independent chatbot use and conventional instruction and the need for scaffolding, critical prompting, student engagement and feedback literacy is critical. Such results further support sociocultural perspectives of language learning to teach and guide children on how to interact with digital technologies and not to be overtaken by them (Lukešová & Jennings, 2026). In addition to formal classroom instruction, generative AI has offered students autonomy, informal and cooperative language learning. Meta-analytic studies show that AI-supported informal digital learning enhances English language proficiency and self-regulation through speaking skills and language assessment in general. Similarly, AI-enhanced telecollaborative settings promote authentic communication, intercultural competency, digital literacy and problem solving within a diverse learning universe by enabling learners to participate in meaningful tasks in a geographically distributed way. While students appreciate AI as a cognitive support tool, they also value teacher feedback in a way that is based on emotion, motivation, and context, which indicates that successful language learning relies on human-AI collaboration rather than technological substitution (Guan et al., 2024; Mont et al., 2026).

In addition to these educational opportunities, increasing research also highlights the ethical, technological, and pedagogical problems associated with chatbot use. Issues include inaccurate information generation, hallucinated responses, algorithmic bias, academic dishonesty, lack of critical thinking, privacy risks, too much reliance on students to learn, and poor access to digital resources. In addition, institutional readiness often lags behind technological innovation, as teachers report little to no knowledge of AI, limited career development opportunities, and a lack of responsibility in the classroom to take charge of AI. Therefore, many authors have proposed that a comprehensive framework of institutional policies, ethical standards, transparent assessment, and teacher education be implemented to support the development of generative AI to help enhance educational quality in the learning process while avoiding the negative side effects (Al Harrasi et al., 2025; Law, 2024; Steele, 2023; Tan et al., 2025).

Moreover, current literature is littered with research gaps that need to be filled in. Although many studies show positive learning outcomes, there is significant heterogeneity in research

design, research participants, educational levels, instructional interventions, outcome measures, and theoretical frameworks. Most studies focus on higher education and writing and relatively little on primary education, listening, intercultural communication, longitudinal learning trajectories, and cross-cultural comparisons. Moreover, very few reviews cover pedagogical applications, educational effectiveness, implementation problems, ethical implications, and emerging technological trends at the same time in a systematic way. It is essential to address these gaps so as to make evidence-based recommendations that will inform future research, classroom practice, and educational policy concerning generative AI chatbots in English language education.

As generative AI technologies are rapidly being implemented into language education, the recent scientific evidence is in need of a comprehensive synthesis. The current reviews, on the other hand, address a single aspect of AI adoption (e.g., writing support, technological effectiveness, or teacher perceptions) without an overall assessment of pedagogical possibilities, learning outcomes, implementation difficulties, ethical considerations, and future developments. This systematic review will provide teachers, researchers, curriculum designers, and policymakers with a well-rounded evidence base for making informed decisions. Such a synthesis may help to ensure responsible implementation of technology, help teachers shape their careers, identify research priorities, and develop pedagogy for AI-enhanced English language learning environments.

In this systematic review, we reviewed peer-reviewed scientific literature published between 2021 and 2026 regarding the integration of generative AI chatbots in ESL/EFL education. This review focuses on studies that examine pedagogical applications, learning outcomes, learner and teacher perceptions, educational opportunities, implementation challenges, ethical considerations and new technological trends in various educational settings and language skills. By combining empirical data from various methodological perspectives, this review aims to identify and address common findings, methodological limitations, research gaps and future directions which are absent from studies unrelated to English language education and generative AI chatbot technologies. By providing a broad understanding of the existing

knowledge and research trends we have in place and setting the agenda for future research and education innovation.

This systematic review aims to systematically identify, analyze, and synthesize the scientific evidence published between 2021 and 2026 on the use of generative AI chatbots in ESL/EFL education, their educational opportunities, learning outcomes, implementation challenges, and emerging trends, while identifying research gaps and future directions to support evidence-based English language teaching and learning. Finally, this systematic review article is based on the research project "Technological Tools Under the SAMR Model." Universidad Estatal del Sur de Manabí as a case study.

Methodology

Based on the methodological framework proposed by Manchado Garabito et al. (2009), I have prepared the following methodology section for a high impact systematic exploratory review. It follows the protocol described by the authors, including the definition of inclusion criteria, information sources, search strategy, study selection, data extraction, bibliometric analysis, and qualitative synthesis.

Research Design

This study used a systematic exploratory review (scoping review) based on the methodological framework proposed by Manchado Garabito et al. (2009), and this method is chosen because it allows us to systematically identify, organize and synthesize current research findings and research theories with very different methodologies, educational contexts and various outcome measures. Unlike traditional systematic reviews to answer a specific research question, exploratory systematic reviews provide an overview of the existing knowledge base, research trends, methodology characteristics and other aspects of research that should be addressed and the research direction for future research (Manchado Garabito et al., 2009). Consequently, this review aims to systematically identify, analyze and synthesize scientific evidence on the use of generative AI chatbots in ESL/EFL education published between 2021 and 2026 which can be applied in the classroom, educational opportunities, learning outcomes, implementation challenges and further research directions.

Research Question

The review was guided by the following research question:

The research was guided by the following research question: What scientific evidence has been published between 2021 and 2026 about the use of generative AI chatbots in ESL/EFL education, and what pedagogical applications, learning outcomes, educational opportunities, implementation challenges, research gaps, and future trends emerge from this body of literature?

Eligibility Criteria

Following the recommendations of Manchado Garabito et al. (2009), explicit inclusion and exclusion criteria were established before initiating the search process to ensure methodological transparency and reproducibility.

Inclusion criteria

Studies were included when they met all of the following conditions:

- Published between **2021 and 2026**.
- Peer-reviewed journal articles.
- Indexed in **Scopus** or **Web of Science (WoS)**.
- Written in English.
- Focused on the application of **generative AI chatbots** (e.g., ChatGPT, Gemini, Claude, Copilot, or comparable conversational AI systems) in **ESL/EFL education**.
- Reported empirical, theoretical, review, or mixed-method evidence concerning pedagogical applications, learning outcomes, teacher or learner perceptions, educational opportunities, ethical issues, implementation challenges, or future directions.

Exclusion criteria

The following documents were excluded:

- Conference papers, editorials, dissertations, book chapters, preprints, reports, and non-peer-reviewed publications.
- Studies unrelated to English language teaching.
- Studies focusing exclusively on general artificial intelligence without chatbot applications.
- Publications lacking sufficient methodological information or full-text availability.
- Duplicate records retrieved from multiple databases.

Information Sources

In accordance with Manchado Garabito et al. (2009), two internationally recognized multidisciplinary bibliographic databases were selected because of their broad coverage of high-quality educational research:

- **Scopus**
- **Web of Science Core Collection**

The final search was conducted in **2026**, ensuring that all eligible publications from the review period were included.

Search Strategy

A structured search strategy was developed using Boolean operators (AND, OR), quotation marks, truncation symbols where appropriate, and controlled vocabulary derived from database indexing systems. The strategy combined descriptors related to generative artificial intelligence, conversational AI, chatbots, English language teaching, and educational contexts.

The principal search equation was formulated as follows:

Recepción: 01-04-2026 / Revisión: 03-04-2026 / Aprobación: 01-07-2026 / Publicación: 27-07-2026

("Generative Artificial Intelligence" OR "Generative AI" OR ChatGPT OR chatbot* OR
"AI chatbot*" OR "conversational AI" OR "large language model*" OR LLM*)
AND

("ESL" OR "EFL" OR "English language learning" OR "English language teaching" OR
"second language learning" OR "foreign language learning")
AND

(education OR teaching OR learning OR classroom)

Filters were applied to limit the search to:

- Publication years: **2021–2026**
- Journal articles
- English language
- Scopus and Web of Science indexed publications

Study Selection

The study selection process was initiated following Manchado Garabito et al. (2009). References were first exported to a reference management database and duplicates were removed. Titles and abstracts were then screened according to the criteria. Studies that may be relevant are then evaluated. The full articles were then evaluated for their methodological relevance and alignment to the objectives of the review. Publications that did not meet the inclusion criteria were excluded and the reasons for their exclusion were listed for transparency. The entire screening process was presented in a PRISMA-style flow diagram. The final sample consisted of n peer-reviewed papers that met all the eligibility criteria and formed the corpus investigated in this systematic exploratory review.

Data Extraction

Following the framework proposed by Manchado Garabito et al. (2009), standardized data extraction forms were developed to ensure consistency throughout the review.

For each included study, the following bibliometric variables were extracted:

- Author(s)
- Year of publication
- Journal
- Database
- Country
- Educational level
- Research design
- Participants

Additionally, content-related variables were extracted, including:

- Type of generative AI chatbot
- Pedagogical application
- Language skill(s) addressed
- Research objectives
- Methodology
- Educational opportunities
- Learning outcomes
- Benefits
- Implementation challenges
- Ethical considerations
- Technological limitations
- Future research recommendations

Data extraction was done separately by the reviewers and discrepancies were resolved by discussion until the group came to a consensus.

Data Analysis.

Similar to the exploratory review methodology of Manchado Garabito et al. (2009), data analysis was performed in two complementary stages. First, a bibliometric analysis was performed to indicate publication trends based on year, country, journal, educational level, research methodology, and database source. This analysis gave a summary of the development of scientific research on generative AI chatbots in ESL/EFL education. We then conducted a qualitative thematic synthesis of the included studies to highlight the common themes. We coded the extracted findings and categorized them into large-scale analytical categories like: Pedagogical applications of generative AI chatbots:

- Educational opportunities;
- Language learning outcomes;
- Learner and teacher perceptions;
- Benefits and instructional effectiveness;
- Ethical, technological, and pedagogical challenges;
- Research gaps;
- Emerging trends and future research priorities.

The synthesis relied on convergent evidence, methodological differences, recurring limitations, and unresolved issues rather than quantitative aggregation, and thus showed a comprehensive picture of the current state of knowledge.

Methodological rigor

All the steps of the review (search strategy development, study selection, data extraction, and thematic synthesis) were done in a structured way for the purpose of a systematic review. Review was conducted transparently, reproducibly and systematically based on Manchado Garabito et al. (2009). In addition, the reporting of study identification, screening, eligibility assessment, and inclusion was done using the PRISMA flow framework to ensure that data was documented and reliable in the review process and to enhance the reliability of the results.

Results

Chart #1

The bibliometric distribution of the 15 selected studies indicates that the research on generative AI chatbots in ESL/EFL education rapidly grew worldwide in the years from 2023 to 2026. Most of the papers were published in 2025 (6 studies) and 2026 (5 studies), indicating the increasing interest in the field after the arrival of large language models. As we see, China (including Hong Kong) produced the largest number of studies (4), Spain (2), and Australia, the Czech Republic, Finland, India, Malaysia, Oman, Saudi Arabia, and the United States produced one paper each. Furthermore, Scopus has the largest number of publications (13), while Web of Science has only two papers, demonstrating that Scopus is currently the most important research source for generative AI chatbots in ESL/EFL education globally.

Characteristic of the selected articles by author, year, country and database.

No.	Author(s)	Year	Country	Database
1	Lukešová & Jennings	2026	Czech Republic	Scopus
2	Lázaro-Ibarrola, Luquin, & Roothoof	2025	Spain	Scopus
3	Kundu & Bej	2025	India	Web of Science
4	Algobaei & Alzain	2026	Saudi Arabia	Scopus
5	Hwang, Chang, & Sun	2025	China	Scopus
6	Steele	2023	Australia	Scopus
7	Guan, Li, & Gu	2024	China	Scopus
8	Law	2024	Hong Kong (China)	Scopus
9	Kohnke, Zou, & Su	2025	Hong Kong (China)	Scopus
10	Mont, Dooly, & Masats	2026	Spain	Scopus
11	Liu, Hashim, & Sulaiman	2026	Malaysia	Scopus
12	Al Harrasi, Al Subhi, Al Harthy, & Al Habsi	2025	Oman	Scopus
13	Torres & Kahveci	2025	United States	Scopus
14	Oyelere	2026	Finland	Scopus
15	Tan, Cheng, & Ling	2025	Hong Kong (China)	Scopus

Source: own elaboration

Chart #2

As shown in the table, most of these studies aim to assess the effectiveness of generative AI and chatbots in ESL/EFL learning in writing, grammar, feedback, learner engagement, and

personalized teaching. Most of the research aims to evaluate pedagogical effectiveness, identify learner perceptions, and develop evidence in favor of AI-supported language teaching. The corpus is diverse in terms of methods, including experimental studies, surveys, mixed-method research, conceptual analysis, systematic reviews, scoping reviews, and meta-analyses. This diversity adds to the review, as it reflects both empirical results and wider theoretical discussions, but also generates heterogeneity that can be difficult to compare with prior studies. As regards the demographics and sample, the table indicates a strong focus on university students, EAP learners, EFL learners, and published studies. However, sample sizes are not reported in the summary, which is a limitation for assessing methodological rigor. Full-text extraction is therefore necessary for the population to be found, as well as sample size and representativeness.

Characteristic of the selected articles by author, objective, methodology, population and sample.

No.	Author(s)	Objective	Methodology	Population	Sample
1	Lukešová & Jennings (2026)	Evaluate the impact of ChatGPT on mastering English tenses at the university level.	Experimental study	University students	Not reported
2	Lázaro-Ibarrola et al. (2025)	Examine reading comprehension in Spanish, Basque, and English among multilingual learners.	Quantitative comparative study	Primary multilingual learners	Not reported
3	Kundu & Bej (2025)	Synthesize empirical evidence on AI applications in school-based EFL teaching.	Systematic review	Empirical EFL studies	Not applicable
4	Algobaei & Alzain (2026)	Investigate the effectiveness of prompt engineering for ChatGPT-generated EFL writing feedback.	Mixed-method evaluation	EFL learners and experts	Not reported
5	Hwang et al. (2025)	Explore how and when L2 learners use ChatGPT during writing tasks.	Mixed-method study	L2 English learners	Not reported
6	Steele (2023)	Discuss the educational implications of ChatGPT in higher education.	Conceptual article	Not applicable	Not applicable
7	Guan et al. (2024)	Examine the effectiveness of GenAI in informal digital English learning.	Meta-analysis	Empirical studies	Number of studies not reported
8	Law (2024)	Review the application of GenAI in language teaching and learning.	Scoping review	Published literature	Number of studies not reported
9	Kohnke et al. (2025)	Explore learners' experiences with GenAI for personalized English learning.	Survey study	First-year EAP students	Not reported
10	Mont et al.	Investigate AI-driven telecollaboration	Project-based case	Young EFL	Not reported

No.	Author(s)	Objective	Methodology	Population	Sample
	(2026)	in foreign language learning.	study	learners	
11	Liu et al. (2026)	Synthesize evidence on emerging technologies in TEFL across educational levels.	Systematic review	Published studies	Number of studies not reported
12	Al Harrasi et al. (2025)	Examine learners' perceptions of AI teacher effectiveness.	Quantitative survey	Learners using AI teachers	Not reported
13	Torres & Kahveci (2025)	Determine the effectiveness of AI in language teaching.	Meta-analysis	Published empirical studies	Number of studies not reported
14	Oyelere (2026)	Review hybrid intelligence systems for human-centered AI in education.	Scoping review	Published studies	Number of studies not reported
15	Tan et al. (2025)	Review AI applications in teaching and teacher professional development.	Systematic review	Published studies	Number of studies not reported

Source: own elaboration

Discussion

The overall results of this systematic review further demonstrate that generative AI chatbots are being used as new pedagogical tools to improve the quality and scope of English language teaching and learning. As they have improved grammatical accuracy, writing, learner engagement, autonomy, and personalized feedback in all studies reviewed in this review, AI-enabled conversational systems have also shown that they can make students better at language teaching and learning in ESL/EFL education, and this shows that these benefits are neither automatic nor independent of pedagogical design. Instead of being substitutes for teachers, generative AI chatbots provide value to the classroom when it comes to embedding them in educational environments with a teacher-driven learning context in which they are well integrated with students and are able to facilitate meaningful teacher mediation, learner reflection, and structured learning. This is in line with the current thinking of technology-based language learning which says that technology-enhanced teaching is a tool to develop students' learning and not just technology, and is not the only thing that is needed to achieve that.

Among the key findings is the role of teaching mediation in enhancing the effectiveness of generative AI in education. Lukešová and Jennings (2026) found that students with guided ChatGPT learning did significantly better with English tense learning than students learning

ChatGPT independently or with typical teaching methods. This suggests that generative AI is not so much about the technology as about teaching students how to engage with AI-generated explanations. The same conclusion is made in Kundu and Bej (2025), who suggest that AI-supported learning environments become pedagogically meaningful only when teachers integrate adaptive teaching, formative assessment, and learner-centered practices. Taken together these studies question the very notion that artificial intelligence can enhance language learning on its own and that teachers must be key mediators to turn AI information into meaningful learning experiences. This interpretation is in line with sociocultural theories of learning, in which knowledge construction is mediated through guided interaction rather than isolated technological exposure.

This synthesis also suggests that generative AI particularly develops productive language skills, especially when it comes to writing. Hwang et al. (2025) found that students used ChatGPT more throughout planning and idea generation (e.g., when they were writing the plans) while grammar and text editing were less common during the later drafting stages. Similarly, Algobaei and Alzain (2026) found prompt engineering significantly improved the teaching quality of ChatGPT-generated feedback and prompts that were specifically designed for EFL learners resulted in a clearer, more relevant, and more motivating response than generic prompts. All these findings contribute to our understanding of AI-assisted writing and may indicate that the effectiveness of AI-assisted writing is not just about the availability of chatbots but also about the quality of human-AI interactions. So prompt literacy is an emerging pedagogical competency in education that is comparable to traditional writing strategies and therefore the feedback which is more supportive for revision, language awareness, and self-regulated learning.

The reviewed evidence also cites personalization as one of the key pedagogical benefits of generative AI chatbots. Unlike traditional teaching tools which provide a set of explanations, examples and responses, conversational AI systems dynamically respond to learner input. Kundu & Bej (2025) describe personalization, feedback, learner autonomy, and adaptive learning pathways as core mechanisms that enable AI to enhance the learning of EFL in reading, writing, listening, speaking, vocabulary and grammar. These insights are also

supported by Algobaei and Alzain (2026) who show that feedback is much more valuable when prompts are tailored to the language proficiency and the cultural context. Personalization has to be seen not just as a technological trait but as a pedagogical principle that allows for differentiated instruction and ensures that learners are more confident, motivated, and remain engaged.

Despite these positive outcomes, the synthesis shows some important pedagogical flaws that temper an overly positive perception of AI integration. Hwang et al. (2025) reported that learners expressed concerns about inaccurate information, repetitive explanations, and excessive reliance on AI-generated content, and Lukešová and Jennings (2026) showed that unguided chatbot-based learning outcomes were significantly worse than teacher-based implementation. These findings indicate that engaging in generative AI for an unlimited amount of time might take away opportunities for critical thinking, independent problem solving, and metalinguistic reflection if students simply accept chatbot responses without reflection. Rather than serving to contradict the positive educational outcome findings, these limitations raise questions of how generative AI can be educationally useful. They argue that the successful implementation will require AI literacy, prompt literacy, critical thinking skills, and reflective learning strategies to transform chatbot interactions into learning experiences rather than passive information consumption.

Another important contribution of this synthesis is the role of cognitive and contextual factors in language learning. Lázaro-Ibarrola et al. (2025) did not directly study generative AI, but their multilingual reading development findings suggest that language learning outcomes are still heavily influenced by cognitive characteristics such as non-verbal intelligence and, to a lesser extent, socioeconomic conditions. These findings are complementary to AI-oriented research and indicate that technological solutions may not be able to address the wider cognitive and contextual factors that impact language learning. Rather, generative AI should be viewed as a supportive tool to help learners learn differently from one another and not an endgame for learning. This interpretation supports the argument that successful AI integration demands pedagogical differentiation which is determined by a learner's cognitive level, educational background, and sociocultural context.

All in all, the first group of studies suggests that generative AI chatbots are useful pedagogical resources and they may have an impact on language learning, especially when the design is guided by teachers' mediation, adaptive feedback, prompt engineering, learner autonomy and reflective engagement. Instead of viewing AI as a replacement for human instruction, the evidence has consistently identified AI as a partner in education that is not replacing the language teacher. This is why we have seen a growing literature to support a human-centered model of AI integration, where technological innovation and pedagogical expertise are combined to create a more personalized, engaging and effective ESL/EFL learning environment.

Conclusion

This systematic review shows that generative AI chatbots are now very useful tools in ESL/EFL education because they enable personal learning, immediate feedback, writing development, grammar practice, speaking opportunities, learner autonomy, and engagement. They have the most pedagogical value when integrated into teacher-led, ethical, and instructionally planned practice. But the results also highlight a number of challenges; overreliance, inaccurate information, academic integrity issues, data privacy concerns, unequal access, lack of AI literacy, and insufficient teacher training all must be taken into account. Generative AI chatbots, therefore, should be seen not as substitutes for teachers, but as complementary pedagogical mediators that need to be critically and sensibly applied and contextually integrated.

To answer this question in depth, the review of evidence has shown that generative AI chatbots are mainly applied in ESL/EFL education for writing support, grammar explanation, feedback generation, vocabulary development, speaking practice, reading support, telecollaboration, and autonomous learning. These tools are a great educational tool in that they provide students with adaptive inputs, individualized feedback, interactive dialogue, and flexible learning experiences beyond the classroom. They are generally good for writing, grammar, speaking confidence, self-regulation, and learner engagement. However, the effectiveness of these tools depends on the educational design, teacher mediation, prompt literacy, learner readiness, and support from the school system. Implementation challenges are still there. Ethical and technological issues such as hallucinated responses, bias,

plagiarism, assessment validity, data protection, and unequal access limit the uncritical adoption of generative AI chatbots in language education. Moreover, most studies concentrate on higher education and writing skills, leaving a gap in primary education, listening development, intercultural communication, long-term learning effects, teachers' professional development, and cross-cultural contexts. Future research should therefore focus on longitudinal, comparative, and classroom-based studies to find out how generative AI chatbots affect different language skills, learner profiles, educational levels, and sociocultural contexts.

In summary, this review concludes that generative AI chatbots represent a promising but complex innovation for ESL/EFL education. Their integration can strengthen evidence-based English language teaching when supported by pedagogical planning, ethical guidelines, teacher training, learner AI literacy, and continuous evaluation. Future directions should move beyond technological enthusiasm toward human-centered, inclusive, and sustainable models of AI-assisted language learning.

References

- Al Harrasi, N., Al Subhi, N., Al Harthy, S., & Al Habsi, J. (2025). *Learner perspectives on AI teacher effectiveness: The role of engagement, motivation, efficiency, and educational experience*. *Social Sciences & Humanities Open*, 12, 101950. <https://doi.org/10.1016/j.ssaho.2025.101950>
- Algobaei, F., & Alzain, E. (2026). *Prompt engineering for non-native English learners: A generative AI approach to personalised language feedback*. *Social Sciences & Humanities Open*, 13, 102341. <https://doi.org/10.1016/j.ssaho.2025.102341>
- Guan, L., Li, S., & Gu, M. M. (2024). *AI in informal digital English learning: A meta-analysis of its effectiveness on proficiency, motivation, and self-regulation*. *Computers and Education: Artificial Intelligence*, 7, 100323. <https://doi.org/10.1016/j.caeai.2024.100323>
- Hwang, H., Chang, X., & Sun, J. (2025). *Generative AI is useful for second language writing, but when, why, and for how long do learners use it?* *Journal of Second Language Writing*, 69, 101230. <https://doi.org/10.1016/j.jslw.2025.101230>
- Lázaro-Ibarrola, A., Luquin, M., & Roothoof, H. (2025). *The reading rainbow of young multilingual learners: Reading comprehension in the majority (Spanish), regional*

(Basque) and foreign (English) language. System, 131, 103665.
<https://doi.org/10.1016/j.system.2025.103665>

Law, L. (2024). *Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping literature review. Computers and Education Open, 6, 100174.*
<https://doi.org/10.1016/j.caeo.2024.100174>

Liu, M., Hashim, H., & Sulaiman, N. A. (2026). *A systematic review of emerging technology applications for teaching English as a foreign language across different educational levels. Educational Research Review, 52, 100795.*
<https://doi.org/10.1016/j.edurev.2026.100795>

Lukešová, A., & Jennings, P. J. (2026). *AI-assisted language learning: ChatGPT's impact on mastering English tenses at the university level. Social Sciences & Humanities Open, 13, 102532.*
<https://doi.org/10.1016/j.ssaho.2026.102532>

Mont, M., Dooly, M., & Masats, D. (2026). *Enhancing foreign language learning through AI-driven telecollaboration. Journal of Responsible Technology, 25, 100153.*
<https://doi.org/10.1016/j.jrt.2026.100153>

Oyelere, S. S. (2026). *A scoping review of hybrid intelligence systems for human-centred AI in education. Computers in Human Behavior: Artificial Humans, 7, 100273.*
<https://doi.org/10.1016/j.chbah.2026.100273>

Kundu, A., & Bej, T. (2025). *Transforming EFL teaching with AI: A systematic review of empirical studies. International Journal of Artificial Intelligence in Education, 35, 2281–2314.*
<https://doi.org/10.1007/s40593-025-00470-0>

Kohnke, L., Zou, D., & Su, F. (2025). *Exploring the potential of GenAI for personalised English teaching: Learners' experiences and perceptions. Computers and Education: Artificial Intelligence, 8, 100371.*
<https://doi.org/10.1016/j.caeai.2025.100371>

Steele, J. L. (2023). *To GPT or not GPT? Empowering our students to learn with AI. Computers and Education: Artificial Intelligence, 5, 100160.*
<https://doi.org/10.1016/j.caeai.2023.100160>

Tan, X., Cheng, G., & Ling, M. H. (2025). *Artificial intelligence in teaching and teacher professional development: A systematic review. Computers and Education: Artificial Intelligence, 8, 100355.*
<https://doi.org/10.1016/j.caeai.2024.100355>

Torres, P. J., & Kahveci, Y. E. (2025). *Effectiveness of artificial intelligence (AI) in language teaching. Computers and Education: Artificial Intelligence, 9, 100522.*
<https://doi.org/10.1016/j.caeai.2025.100522>