

SSTSDS-Vol.4. N1. 095

Influence of artificial intelligence based applications on the development of english speaking skills among EFL learners

Influencia de las aplicaciones basadas en inteligencia artificial en el desarrollo de las habilidades de expresión oral en inglés en estudiantes de inglés como lengua extranjera

Autores:

Margarita Elizabeth Rocha Chasi
Unidad Educativa Sigchos
Cotopaxi- Ecuador

elizabeth.rocha@docentes.educacion.edu.ec
<https://orcid.org/0009-0008-1543-377X>

Johanna Carolina Pazmiño Freire
Universidad Técnica de Ambato
Tungurahua – Ecuador

jcpfcarolina@gmail.com
<https://orcid.org/0009-0007-5400-8285>

Wendy Ximena Ordóñez Pérez
Unidad Educativa Federico González Suárez
Chimborazo – Ecuador

bryakvaor@gmail.com
<https://orcid.org/0009-0009-3264-4955>

Mónica Katherine Guachi Caiza
Universidad de las Fuerzas Armadas ESPE
Tungurahua - Ecuador

monica.katerine.guachi@gmail.com
<https://orcid.org/0009-0006-2553-7754>

Jairo Mauricio Ases Villacís
Doctorando del programa doctorado en informática, UNADE - México
Cancún - México

jairo.ases@unade.edu.mx
<https://orcid.org/0009-0001-0053-450X>

Autor de correspondencia: Margarita Elizabeth Rocha Chasi,
elizabeth.rocha@docentes.educacion.edu.ec

Enviado: 09-05-2026
Aceptado: 26-07-2026

Revisado: 18-05-2026
Publicado: 29-07-2026



Cómo citar este artículo:

Rocha Chasi, M. E., Pazmiño Freire, J. C., Ordóñez Pérez, W. X., Guachi Caiza, M. K., & Ases Villacís, J. M. (2026). Influencia de las aplicaciones basadas en inteligencia artificial en el desarrollo de las habilidades de expresión oral en inglés en estudiantes de inglés como lengua extranjera. *Sage Sphere of Technology, Sciences, Discoveries And Society*, 4(1), 1-31. <https://doi.org/10.63688/gteyt983>

© 2026; Los autores. Este es un artículo en acceso abierto, distribuido bajo los términos de una licencia Creative Commons (<https://creativecommons.org/licenses/by/4.0>) que permite el uso, distribución y reproducción en cualquier medio, siempre que la obra original sea correctamente citada.



ABSTRACT

This systematic literature review analyzed the influence of artificial intelligence-based applications on the development of English-speaking skills among EFL learners. The study reviewed 19 scientific publications published between 2020 and 2026, selected from academic databases following the PRISMA 2020 guidelines. The findings were organized into five thematic categories: chatbots and conversational assistants; automatic speech recognition and pronunciation; fluency, accuracy, and oral performance; motivation, anxiety, autonomy, and willingness to communicate; and pedagogical integration, opportunities, and limitations.

The reviewed evidence suggests that AI-based applications have the potential to strengthen pronunciation, fluency, vocabulary, grammatical accuracy, confidence, autonomy, and willingness to communicate. Most studies reported generally favorable results and identified benefits such as immediate feedback, repeated speaking practice, and personalized learning opportunities. However, the findings were heterogeneous because the included sources differed in terms of AI tools, research designs, learner populations, proficiency levels, and intervention duration. Limitations related to inaccurate responses, unequal technological access, privacy concerns, and excessive dependence on automated systems were also identified. The review suggests that artificial intelligence may constitute a valuable complementary resource for English-speaking development when it is integrated with clear pedagogical objectives, teacher guidance, and authentic communicative activities.

Keywords: artificial intelligence, English speaking, EFL learners, educational technology.

RESUMEN

Esta revisión sistemática de la literatura analizó la influencia de las aplicaciones basadas en inteligencia artificial en el desarrollo de las habilidades de expresión oral en inglés en estudiantes de inglés como lengua extranjera. Se revisaron 19 publicaciones científicas publicadas entre 2020 y 2026, seleccionadas en bases de datos académicas mediante las directrices PRISMA 2020. Los hallazgos se organizaron en cinco categorías temáticas: chatbots y asistentes conversacionales; reconocimiento automático del habla y pronunciación; fluidez, precisión y desempeño oral; motivación, ansiedad, autonomía y disposición para comunicarse; e integración pedagógica, oportunidades y limitaciones.

La evidencia revisada sugiere que las aplicaciones basadas en inteligencia artificial tienen potencial para fortalecer la pronunciación, la fluidez, el vocabulario, la precisión gramatical, la confianza, la autonomía y la disposición para comunicarse. La mayoría de los estudios reportó resultados generalmente favorables e identificó beneficios como retroalimentación inmediata, práctica oral repetida y oportunidades de aprendizaje personalizado. Sin embargo, los hallazgos fueron heterogéneos debido a las diferencias entre las herramientas de inteligencia artificial, los diseños de investigación, las poblaciones estudiadas, los niveles de dominio del inglés y la duración de las intervenciones. También se identificaron limitaciones relacionadas con respuestas inexactas, desigualdad en el acceso tecnológico, riesgos de privacidad y dependencia excesiva de sistemas automatizados. La revisión sugiere que la inteligencia artificial puede constituir un recurso complementario valioso para el desarrollo de la expresión oral en inglés cuando se integra con objetivos pedagógicos claros, orientación docente y actividades comunicativas auténticas.

Palabras clave: inteligencia artificial, expresión oral, estudiantes EFL, tecnología educativa.



1. INTRODUCTION

English-speaking competence has become a fundamental objective in English as a Foreign Language (EFL) education because oral communication enables learners to participate effectively in academic, professional, social, and intercultural contexts. However, speaking is often considered one of the most challenging language skills, since learners must simultaneously manage vocabulary, grammatical accuracy, pronunciation, fluency, coherence, and interaction. In many EFL contexts, limited classroom time, fear of making mistakes, speaking anxiety, and insufficient opportunities to communicate in English restrict the development of oral proficiency. Consequently, technology-supported learning environments have gained importance as complementary spaces in which students can practice speaking more frequently and receive additional assistance beyond traditional classroom instruction (El Shazly, 2021; Yang et al., 2022).

Within this technological transformation, artificial intelligence has introduced innovative applications that can support the teaching and learning of spoken English. AI-based tools include conversational chatbots, automatic speech recognition systems, intelligent virtual assistants, adaptive learning platforms, and generative artificial intelligence applications. These technologies can simulate conversations, recognize spoken language, generate context-related responses, identify linguistic errors, and provide immediate feedback according to learners' individual needs. Unlike conventional speaking activities that may depend on the availability of teachers or classmates, AI applications allow students to practice at different times, repeat activities, select topics of interest, and progress at their own pace. Therefore, these tools may create accessible and personalized environments in which learners can participate in meaningful oral interactions without the pressure commonly associated with speaking in front of other people (Du & Daniel, 2024; Kohnke et al., 2023). The influence of AI-based applications on English-speaking development extends to several components of oral production. Through continuous interaction with chatbots and speech-recognition systems, learners can practice pronunciation, fluency, grammatical structures, vocabulary, coherence, and conversational strategies. Recent empirical evidence suggests that AI-mediated speaking activities can positively influence learners' oral performance and willingness to communicate. For example, AI chatbots can function as conversation partners



that encourage students to produce spoken responses, negotiate meaning, and repeat expressions until they feel more confident. Studies have also reported improvements in fluency, pronunciation, content development, grammatical accuracy, and overall speaking performance after learners participated in structured AI-supported activities (Fathi et al., 2024; Tai & Chen, 2024; Yang et al., 2022).

In addition to linguistic development, artificial intelligence may influence emotional and motivational factors related to speaking. Many EFL learners avoid oral participation because they are concerned about negative evaluation, pronunciation mistakes, or their inability to respond immediately. AI applications can provide a less judgmental environment in which learners feel more comfortable experimenting with the language. Immediate and individualized feedback may help students identify their strengths and weaknesses, monitor their progress, increase their confidence, and develop greater responsibility for their learning. Research has shown that AI-supported practice can reduce speaking anxiety, strengthen motivation, promote learner autonomy, and increase willingness to communicate in English (El Shazly, 2021; Gortaire-Díaz et al., 2026; Sayed et al., 2024). These emotional benefits are particularly relevant because successful oral communication depends not only on linguistic knowledge but also on learners' confidence and willingness to use the language. Nevertheless, the incorporation of AI-based applications into EFL speaking instruction also presents important challenges. The feedback generated by AI systems may occasionally be inaccurate, incomplete, or inappropriate for the learner's proficiency level. Technical difficulties, limited access to digital devices, internet connectivity, data privacy, algorithmic bias, and excessive dependence on automated tools may also affect their educational effectiveness. Furthermore, interacting with an artificial system cannot entirely replace spontaneous human communication, since authentic conversations involve emotions, gestures, cultural knowledge, negotiation, and unpredictable responses. For this reason, AI applications should be considered complementary resources rather than substitutes for teachers and human interaction. Their implementation requires pedagogical planning, teacher supervision, clear learning objectives, digital literacy, and opportunities to apply the acquired skills in real communicative situations (Du & Daniel, 2024; Guillermo Morales & Carcausto Calla, 2025; Kohnke et al., 2023).



In this context, examining the influence of artificial intelligence-based applications on English-speaking skills is necessary to understand how these emerging technologies can contribute to contemporary EFL education. Although recent studies have identified positive results, the effectiveness of AI may vary according to the application employed, the learners' proficiency level, the duration of the intervention, the type of speaking activity, and the pedagogical guidance provided. Therefore, this research aims to analyze the impact of AI-based applications on the development of English-speaking skills among EFL learners, considering their contribution to oral performance, pronunciation, fluency, confidence, motivation, autonomy, and willingness to communicate. Accordingly, the study addresses the following research question: How do artificial intelligence-based applications influence the development of English-speaking skills among EFL learners?

2. THEORETICAL FRAMEWORK

Artificial Intelligence in Language Education

Artificial intelligence in language education includes computer systems capable of recognizing speech, generating responses, evaluating oral production, and adapting activities to learners' needs. Its main applications include conversational chatbots, automatic speech recognition systems, intelligent personal assistants, and automated speaking assessment platforms. These tools create interactive learning environments where students can receive immediate assistance and practice English beyond regular classroom hours. Huang et al. (2023) explain that AI has expanded language education by supporting personalization, automatic feedback, learner autonomy, and access to authentic communicative activities.

AI-Based Applications for Speaking Practice

AI-based applications can function as virtual conversation partners that allow learners to participate in simulated dialogues, answer questions, practice everyday situations, and repeat oral activities. Hsu et al. (2023) found that task-oriented chatbots can provide accessible speaking practice and help learners feel less afraid of communicating in English. Similarly, Chen et al. (2023) indicate that intelligent personal assistants such as Google Assistant encourage independent language practice because students can interact through questions, instructions, and voice commands. Nevertheless, these applications are most effective when their activities correspond to learners' proficiency levels and specific speaking objectives.



Pronunciation and Automatic Speech Recognition

Pronunciation is an essential component of speaking because it influences the intelligibility and comprehensibility of oral messages. Automatic speech recognition technology converts spoken language into digital information and compares learners' production with expected pronunciation patterns. Inceoglu et al. (2020) found that mobile ASR practice can support the production of specific English sounds, although its effectiveness may vary depending on the linguistic feature being practiced. Evers and Chen (2021) also demonstrated that sustained ASR practice contributes to pronunciation development among adult language learners.

The effectiveness of these systems depends considerably on the type of feedback and the duration of practice. Sun (2023) showed that combining ASR technology with peer correction improved pronunciation, comprehensibility, and overall speaking performance among EFL learners. Furthermore, the meta-analysis conducted by Ngo et al. (2024) identified a moderate positive effect of ASR on pronunciation, particularly when learners received explicit corrective feedback, practiced for sufficient periods, and participated in collaborative activities. Therefore, automatic recognition should not only indicate whether pronunciation is correct but also provide understandable guidance for improvement.

Fluency, Feedback, and Speaking Development

Speaking proficiency involves more than pronouncing individual words correctly. It also includes fluency, grammatical accuracy, vocabulary, coherence, rhythm, intonation, and the ability to respond appropriately during interaction. Zou et al. (2023) found that AI speech-evaluation systems contributed to improvements in pronunciation, fluency, grammar, vocabulary, reading aloud, and spontaneous speaking. Their findings also indicate that learners value clear scores, visual indicators, specific corrections, and practical suggestions. However, inaccurate recognition or unclear evaluations may confuse students, demonstrating the importance of teacher supervision.

AI applications also allow learners to repeat speaking tasks without the time limitations commonly found in traditional classrooms. Through repeated practice, students can identify errors, adjust their speech, and monitor their progress. However, repetition must include meaningful communicative tasks rather than isolated mechanical exercises. AI-based speaking practice should consequently incorporate situations such as interviews, role plays, oral presentations, problem-solving activities, and everyday conversations.



Speaking Anxiety and Willingness to Communicate

Emotional factors directly influence learners' willingness to use English. Fear of making mistakes, negative evaluation, and limited confidence may prevent students from participating, even when they possess sufficient linguistic knowledge. Bashori et al. (2022) found that web-based speaking practice supported by automatic speech recognition can help reduce foreign-language speaking anxiety by providing learners with a private and less judgmental environment. This encourages students to experiment with pronunciation and repeat oral responses without immediate social pressure.

In addition, Zhang et al. (2024) demonstrated that interaction with an AI-speaking assistant increased learners' enjoyment and willingness to communicate while reducing speaking anxiety. These results indicate that AI applications can support both linguistic and emotional development. Nevertheless, AI interaction should complement communication with teachers and classmates because human conversations involve cultural interpretation, emotional expression, gestures, spontaneous reactions, and negotiation of meaning that automated systems cannot fully reproduce.

Pedagogical Integration of AI Applications

The educational value of AI depends on how teachers integrate these tools into the learning process. AI applications should be selected according to learners' proficiency, learning objectives, technological access, and the speaking component that requires improvement. Teachers remain responsible for designing activities, evaluating the accuracy of automated feedback, explaining errors, and creating opportunities to transfer AI-supported practice into authentic communication.

In summary, the theoretical evidence suggests that AI-based applications can contribute to EFL speaking development by expanding practice opportunities, providing immediate feedback, supporting pronunciation and fluency, reducing speaking anxiety, and strengthening learners' willingness to communicate. However, their effectiveness depends on sustained practice, reliable feedback, meaningful oral activities, and appropriate pedagogical guidance.

3. METHODOLOGY

Research Design



The present study was conducted as an integrative literature review combining primary and secondary evidence to analyze the influence of artificial intelligence-based applications on the development of English-speaking skills among EFL learners. The review focused on the contribution of conversational chatbots, automatic speech recognition systems, generative AI tools, virtual assistants, and automated feedback platforms to pronunciation, fluency, vocabulary, grammatical accuracy, confidence, motivation, and willingness to communicate. A qualitative descriptive-analytical approach was adopted to organize, compare, and interpret the findings reported in previous studies. Systematic reviews provide a structured method for identifying relevant evidence, evaluating studies according to previously established criteria, and synthesizing their principal findings in a transparent manner (Varsha et al., 2024). In educational research, this process also facilitates the identification of recurring results, methodological limitations, and research gaps within a specific field of study (Ghamrawi et al., 2025).

The evidence corpus included primary studies, such as empirical, experimental, case, mixed-methods, and design-based research, as well as secondary and conceptual sources, including systematic reviews, a meta-analysis, a literature review, and a conceptual article. These two types of evidence were analyzed separately to reduce the risk of double counting findings reported in both primary studies and previous reviews. Primary studies were used as the main basis for identifying effects on English-speaking development, whereas secondary and conceptual sources were used to contextualize, compare, and interpret the findings.

Information Sources and Search Strategy

The bibliographic search was conducted in academic databases related to education, applied linguistics, language learning, and educational technology. The databases consulted were:

- Scopus
- Web of Science
- ScienceDirect
- Taylor & Francis Online
- SpringerLink
- SAGE Journals
- ERIC



- SciELO
- Redalyc
- Google Scholar

These databases were selected because they provide access to peer-reviewed publications concerning English language teaching, artificial intelligence in education, computer-assisted language learning, automatic speech recognition, chatbot-assisted learning, and oral communication development.

Search Fields and Applied Filters

The searches were conducted in the title, abstract, and keyword fields whenever these options were available in the selected databases. In databases with more limited search functions, the search was performed using the available general search field. The following filters were applied: publication years from 2020 to 2026, English or Spanish language, peer-reviewed journal articles, and documents with accessible full text. Editorials, opinion articles, essays, and publications without a defined methodology were excluded.

The search was conducted between March 10 and March 15, 2026. During this period, search equations were applied to the selected databases, the identified records were organized, and duplicate studies were removed. A final verification of authorship, publication details, journal information, and DOI data was conducted on March 20, 2026.

Reference Management and Duplicate Removal

All identified references were organized in a Microsoft Excel spreadsheet. Duplicate records were identified through a combination of automatic sorting and manual comparison using the authors, title, publication year, journal, and DOI as comparison fields. Records with incomplete or inconsistent information were reviewed manually before removal.

Table 1

Records Identified by Database

Database	Records identified
Scopus	14
Web of Science	12
ScienceDirect	13



Database	Records identified
Taylor & Francis Online	11
SpringerLink	10
SAGE Journals	12
ERIC	13
SciELO	8
Redalyc	9
Google Scholar	10
Total	112

Note. A total of 112 records were identified through the selected databases. After removing 27 duplicate records, 85 studies remained for the screening stage. During the review of titles and abstracts, 52 studies were excluded because they did not correspond directly to the research objectives. Consequently, 33 articles were assessed through full-text evaluation. Of these, 14 studies were excluded because they did not specifically examine artificial intelligence, English-speaking development, EFL learners, oral communication, or related affective factors. Finally, 19 scientific studies were included in the systematic review.

Search Strategy

The search strategy was developed using keywords in English combined with the Boolean operators AND and OR. The main search terms were:

“artificial intelligence”, “AI-based applications”, “English-speaking skills”, “EFL learners”, “AI chatbots”, “automatic speech recognition”, “generative artificial intelligence”, “speaking performance”, “pronunciation”, “speaking fluency”, “speaking anxiety”, “willingness to communicate”, “automated feedback”, “English language learning”

The search equations were adapted to the characteristics and advanced search options of each database.

Search Equations

Scopus and Web of Science

("artificial intelligence" OR "AI-based applications" OR "AI chatbots")

AND



("English speaking" OR "speaking skills" OR "oral communication")

AND

("EFL learners" OR "English language learners")

ScienceDirect and SpringerLink

("artificial intelligence" OR "generative AI" OR "automatic speech recognition")

AND

("English language learning")

AND

("pronunciation" OR "fluency" OR "speaking performance")

Taylor & Francis Online and SAGE Journals

("AI-assisted language learning" OR "AI chatbot")

AND

("EFL speaking")

AND

("motivation" OR "speaking anxiety" OR "willingness to communicate")

ERIC, SciELO, and Redalyc

("artificial intelligence" AND "English speaking")

OR

("chatbots" AND "EFL learners")

Google Scholar

"artificial intelligence" "English-speaking skills" "EFL learners"

Additional searches included the following combinations:

"automatic speech recognition" AND "English pronunciation"

"generative AI" AND "EFL speaking performance"

"AI chatbot" AND "speaking anxiety"

"automated feedback" AND "English fluency"

"artificial intelligence" AND "willingness to communicate"

Eligibility Criteria

Inclusion Criteria



Because Google Scholar may retrieve a large number of records, the screening was limited to the first 100 results, ordered by relevance. This limit and the ordering criterion were established before the screening process to ensure consistency and reproducibility.

The following inclusion criteria were established:

- Scientific articles published in academic or indexed journals.
- Peer-reviewed empirical studies or systematic reviews.
- Research related to artificial intelligence and English language learning.
- Studies examining English-speaking skills, pronunciation, fluency, vocabulary, grammatical accuracy, confidence, motivation, or willingness to communicate.
- Studies focused on EFL or ESL learners.
- Publications written in English or Spanish.
- Studies published between 2020 and 2026.
- Articles with accessible full text.
- Publications with a DOI or verifiable bibliographic information.

Exclusion Criteria

The following exclusion criteria were applied:

- Duplicate documents.
- Opinion articles, editorials, essays, or publications without a defined methodology.
- Studies unrelated to English language learning.
- Research focused exclusively on reading, writing, or listening without addressing speaking.
- Studies concerning artificial intelligence in educational areas unrelated to language learning.
- Publications outside the period from 2020 to 2026.
- Documents without accessible full text.
- Publications without reliable bibliographic information.
- Studies that did not clearly report participants, methods, or results.

Study Selection Process

The study selection process followed the identification, screening, eligibility, and inclusion stages established in the PRISMA 2020 statement. This reporting framework promotes



transparency by requiring researchers to explain how records were found, screened, excluded, and included in a systematic review (Page et al., 2021).

The process consisted of four stages:

1. Identification

A total of 112 records were identified through the selected academic databases.

2. Screening

After removing 27 duplicate records, 85 studies remained. Their titles and abstracts were reviewed, and 52 studies were excluded because they did not meet the research objectives.

3. Eligibility

A total of 33 studies were evaluated through full-text analysis. During this stage, 14 articles were excluded for the following reasons:

- They did not directly examine artificial intelligence-based applications.
- They did not focus on English-speaking skills.
- They addressed educational technology without analyzing EFL or ESL learning.
- They did not present sufficient methodological or empirical evidence.
- Their publication dates were outside the established period.

4. Inclusion

Finally, 19 scientific studies met all inclusion criteria and were incorporated into the systematic review.

Data Extraction and Analysis

A data extraction matrix was created to organize the information from the 19 selected studies. The following elements were recorded:

- Authors and year of publication.
- Country or educational context.
- Research design.
- Participant characteristics.
- Type of AI application.
- Speaking skill or affective variable analyzed.
- Principal findings.
- Reported limitations.



- DOI or bibliographic identifier.

Data Extraction and Verification Protocol

Data were extracted using a standardized matrix developed for this review. The extraction process included authorship, publication year, country or educational context, research design, participant characteristics, type of AI application, speaking skill or affective variable analyzed, principal findings, limitations, DOI, educational level, and overall outcome. Each record was extracted by one reviewer and verified through repeated cross-checking against the full text of the original publication. Any inconsistency was resolved by returning to the source document and comparing the extracted information with the complete article. Although this procedure helped reduce transcription and classification errors, the absence of independent verification by a second reviewer is recognized as a methodological limitation.

The information was analyzed through thematic categorization.

Before thematic categorization, the sources were divided into two evidence groups: primary evidence and secondary or conceptual evidence. Primary evidence included empirical, experimental, case, mixed-methods, and design-based studies. Secondary or conceptual evidence included systematic reviews, the meta-analysis, the literature review, and the conceptual article. Findings from secondary sources were not counted as additional independent empirical effects when they summarized studies already represented in the primary evidence corpus.

The findings were grouped into the following five thematic categories:

Chatbots and conversational assistants.

Automatic speech recognition and pronunciation.

Fluency, accuracy, and oral performance.

Motivation, anxiety, autonomy, and willingness to communicate.

Pedagogical integration, opportunities, and limitations.

This procedure made it possible to compare the results of the selected studies and identify the main contributions of AI-based applications to English-speaking development.

Due to the heterogeneity of the included sources, a common eight-criterion appraisal framework was used to assess methodological elements applicable across document types, including clarity of objectives, theoretical foundation, methodological transparency, appropriateness of the design, description of the evidence or participants, clarity of



the analysis, correspondence between results and conclusions, and acknowledgment of limitations. This common framework facilitated comparison across sources; however, it did not replace design-specific critical appraisal instruments. Therefore, the scores were used to guide the interpretation and weighting of the evidence rather than to establish direct methodological equivalence among the different study designs.

Each criterion was scored as 0 when it was not met, 1 when it was partially met, and 2 when it was fully met. The maximum total score was 16 points. The percentage was calculated by dividing the score obtained by 16 and multiplying the result by 100. Sources scoring 80% or higher were classified as high quality, those scoring between 60% and 79% as moderate quality, and those scoring below 60% as low quality.

The review was not prospectively registered, and no public protocol was published before the search and screening process. The search strategy, eligibility criteria, extraction variables, and analytical procedures are therefore reported in detail in this article to improve transparency. The absence of prior registration is recognized as a limitation of the review.

Table 2
Results of the Methodological Quality Assessment

No.	Author and year	Type of document	C1	C2	C3	C4	C5	C6	C7	C8	Total score	Percentage	Quality level
1	Du and Daniel (2024)	Systematic literature review	2	2	2	2	2	2	2	1	15/16	93.75%	High
2	El Shazly (2021)	Case study	2	2	1	2	2	1	2	1	13/16	81.25%	High
3	Fathi et al. (2024)	Empirical research article	2	2	2	2	2	2	2	1	15/16	93.75%	High
4	Gortaire- Díaz et al. (2026)	Empirical research article	2	2	1	1	2	1	2	1	12/16	75.00%	Moderate



No.	Author and year	Type of document	C1	C2	C3	C4	C5	C6	C7	C8	Total score	Percentage	Quality level
5	Guillermo Morales and Carcausto Calla (2025)	Systematic literature review	2	2	1	2	1	2	2	1	13/16	81.25%	High
6	Kohnke et al. (2023)	Conceptual research article	2	2	1	1	1	2	2	1	12/16	75.00%	Moderate
7	Sayed et al. (2024)	Empirical research article	2	2	2	2	2	2	2	1	15/16	93.75%	High
8	Tai and Chen (2024)	Experimental research article	2	2	2	2	2	2	2	2	16/16	100.00%	High
9	Yang et al. (2022)	Empirical research article	2	2	2	2	2	2	2	1	15/16	93.75%	High
10	Bashori et al. (2022)	Empirical research article	2	2	2	2	2	1	2	1	14/16	87.50%	High
11	Chen et al. (2023)	Empirical research article	2	2	1	2	2	1	2	1	13/16	81.25%	High



No.	Author and year	Type of document	C1	C2	C3	C4	C5	C6	C7	C8	Total score	Percentage	Quality level
12	Evers and Chen (2021)	Experimental research article	2	2	2	2	2	1	2	1	14/16	87.50%	High
13	Hsu et al. (2023)	Design-based empirical study	2	2	2	2	2	1	2	1	14/16	87.50%	High
14	Huang et al. (2023)	Literature review	2	2	2	2	2	2	2	1	15/16	93.75%	High
15	Inceoglu et al. (2020)	Empirical pronunciation study	2	2	1	2	2	1	2	1	13/16	81.25%	High
16	Ngo et al. (2024)	Meta-analysis	2	2	2	2	2	2	2	2	16/16	100.00%	High
17	Sun (2023)	Mixed- methods research article	2	2	2	2	2	2	2	1	15/16	93.75%	High
18	Zhang et al. (2024)	Empirical research article	2	2	1	1	2	1	2	1	12/16	75.00%	Moderate
19	Zou et al. (2023)	Empirical research article	2	2	1	1	2	1	2	1	12/16	75.00%	Moderate

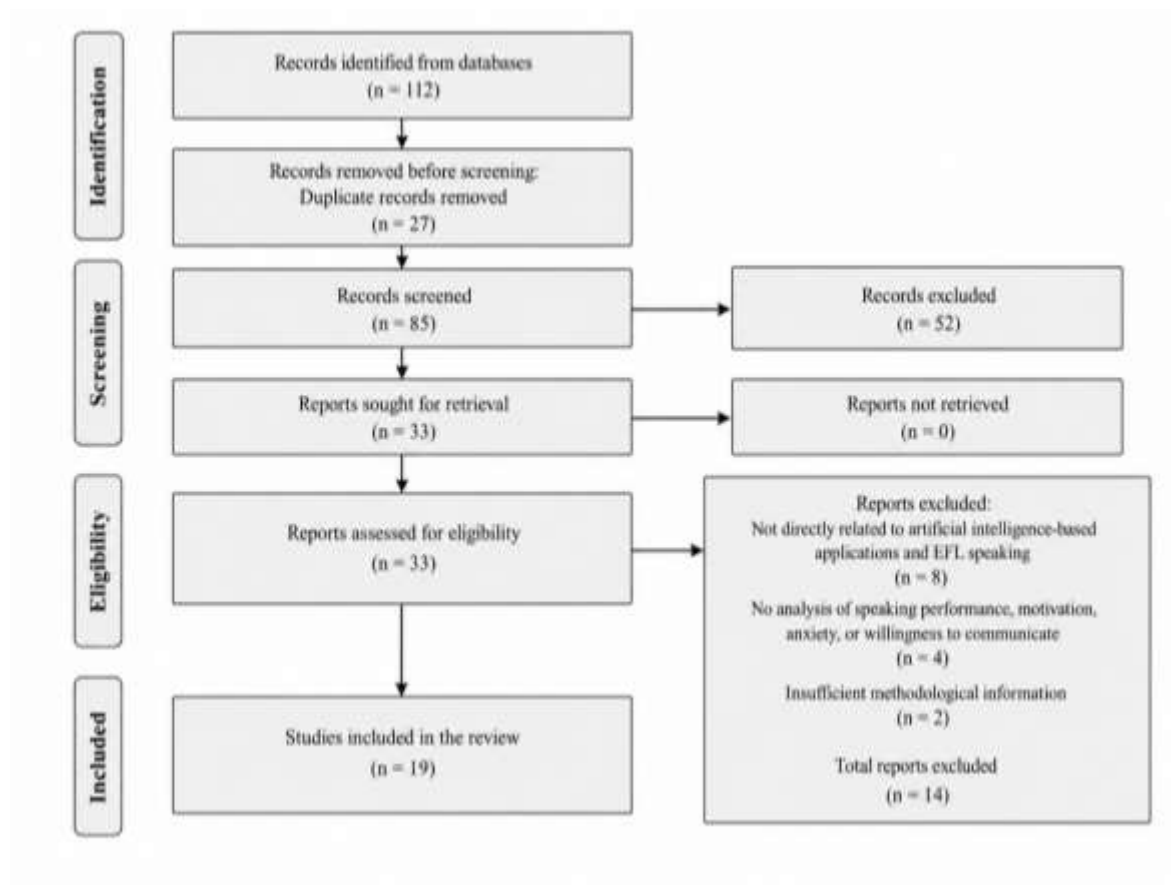
Note. C1 = clarity of objectives; C2 = adequacy of the theoretical foundation; C3 = methodological transparency; C4 = appropriateness of the research design; C5 = description of participants, corpus, or evidence; C6 = clarity of the analysis; C7 = correspondence between findings and conclusions; C8 = acknowledgment of limitations. Each criterion was



scored from 0 to 2, with a maximum total score of 16 points. NR means that the individual criterion score was not reported in the original assessment.

Figure 1

PRISMA 2020 Flow Diagram



Note. Flow diagram developed according to the PRISMA 2020 reporting guidelines proposed by Page et al. (2021). The systematic search identified 112 records related to artificial intelligence, English language learning, and speaking development. After duplicate removal, title and abstract screening, and full-text assessment, 19 scientific studies met the eligibility criteria and were included in the final review.

Reviewers and Disagreement Resolution

The study selection process was conducted by one reviewer. Titles and abstracts were screened first, followed by the full-text assessment of potentially eligible publications. Because only one reviewer participated, no inter-reviewer disagreement resolution procedure



or inter-rater agreement assessment was performed. This limitation was considered when interpreting the transparency and reproducibility of the selection process.

4. RESULTS

The analysis of the 19 studies selected through the systematic literature review allowed the identification of the main trends, contributions, and challenges associated with the use of artificial intelligence-based applications to improve English-speaking skills among EFL learners. The included studies were organized into five thematic categories: chatbots and conversational assistants; automatic speech recognition and pronunciation; fluency, accuracy, and oral performance; motivation, anxiety, autonomy, and willingness to communicate; and pedagogical integration, opportunities, and limitations.

Separation of Primary and Secondary Evidence

Of the 19 included sources, 14 were classified as primary studies and 5 as secondary or conceptual sources. Primary studies provided direct evidence regarding changes in pronunciation, fluency, accuracy, oral performance, motivation, anxiety, autonomy, and willingness to communicate. Secondary and conceptual sources were used to support the interpretation of these findings, identify broader trends, and discuss pedagogical opportunities and limitations. To avoid duplication, conclusions about effectiveness were based primarily on the findings of the primary studies.

Distribution of Studies by Thematic Category

The 19 included sources were distributed across the five thematic categories established for the review. Five studies focused on chatbots and conversational assistants, four examined automatic speech recognition and pronunciation, three addressed fluency, accuracy, and oral performance, four analyzed motivation, anxiety, autonomy, and willingness to communicate, and three discussed pedagogical integration, opportunities, and limitations. This distribution shows that the largest proportion of the corpus concentrated on conversational AI tools and affective factors, whereas fewer studies focused specifically on oral performance and pedagogical integration.

The AI applications examined across the included studies comprised conversational chatbots, generative AI chatbots, ChatGPT, Google Assistant, task-oriented chatbots, automatic speech recognition software, mobile pronunciation applications, web-based speaking



platforms, and AI-supported speech-evaluation systems. These tools were used mainly for conversational practice, pronunciation training, automated feedback, self-directed learning, and the development of speaking confidence.

Most primary studies reported positive effects of AI-supported practice on pronunciation, fluency, speaking performance, motivation, autonomy, confidence, and willingness to communicate. However, several studies also reported mixed findings because the benefits varied according to the accuracy of automated feedback, the duration of practice, learners' proficiency levels, and the type of speaking task.

Pronunciation was examined mainly in the studies by Evers and Chen (2021), Inceoglu et al. (2020), Ngo et al. (2024), and Sun (2023). Fluency, grammatical accuracy, vocabulary, and broader oral performance were addressed in the studies by Fathi et al. (2024), Sayed et al. (2024), and Zou et al. (2023). Affective variables, including anxiety, autonomy, motivation, enjoyment, and willingness to communicate, were examined by El Shazly (2021), Gortaire-Díaz et al. (2026), Bashori et al. (2022), and Zhang et al. (2024).

Experimental and primary empirical evidence was obtained mainly from studies such as Tai and Chen (2024), Evers and Chen (2021), Fathi et al. (2024), Yang et al. (2022), Bashori et al. (2022), Inceoglu et al. (2020), Sun (2023), Zhang et al. (2024), and Zou et al. (2023). In contrast, broader conclusions regarding pedagogical integration, general trends, and limitations were supported by the systematic reviews of Du and Daniel (2024) and Guillermo Morales and Carcausto Calla (2025), the meta-analysis by Ngo et al. (2024), the literature review by Huang et al. (2023), and the conceptual contribution by Kohnke et al. (2023).

This classification made it possible to understand the influence of AI-based applications from technological, pedagogical, linguistic, and emotional perspectives, identifying their potential contribution to pronunciation, fluency, vocabulary, grammatical accuracy, learner autonomy, confidence, and willingness to communicate. Likewise, the analysis considered limitations reported in the literature, including inaccurate automated feedback, unequal access to technology, insufficient teacher guidance, privacy concerns, and excessive dependence on AI tools.

Results of the Methodological Quality Assessment



The methodological quality assessment of the 19 selected studies showed that 15 studies were classified as high quality and 4 as moderate quality. No studies were identified as having low methodological quality.

The studies classified as high quality presented clear research objectives, appropriate theoretical foundations, transparent methodological procedures, and a strong relationship between their findings and conclusions. These studies provided greater evidence regarding the contribution of chatbots, automatic speech recognition, generative AI, and automated feedback to the development of English-speaking skills.

The studies classified as moderate quality mainly presented limitations related to small samples, short intervention periods, reduced empirical analysis, or a stronger conceptual orientation. However, these studies were considered relevant because they contributed perspectives regarding motivation, learner autonomy, speaking anxiety, and the pedagogical integration of artificial intelligence.

The methodological quality assessment was not used as an automatic exclusion criterion; instead, it was applied to determine the level of confidence assigned to each source during the synthesis process. Studies with higher methodological quality received greater consideration when identifying the main trends and establishing conclusions.

Table 3

General Matrix of the Sources Included in the Systematic Review

No.	Author and year	Type of source	Main focus	Thematic category
1	Du and Daniel (2024)	Systematic literature review	AI-powered chatbots for EFL-speaking practice	Chatbots and conversational assistants
2	El Shazly (2021)	Case study	Speaking anxiety and speaking performance	Motivation, anxiety, autonomy, and willingness to communicate



No.	Author and year	Type of source	Main focus	Thematic category
3	Fathi et al. (2024)	Empirical research	Speaking skills and willingness to communicate	Fluency, accuracy, and oral performance
4	Gortaire-Díaz et al. (2026)	Empirical research	Self-directed EFL learning through AI tools	Motivation, anxiety, autonomy, and willingness to communicate
5	Guillermo Morales and Carcausto Calla (2025)	Systematic literature review	Chatbots in language learning	Pedagogical opportunities and limitations.
6	Kohnke et al. (2023)	Conceptual research	ChatGPT in language teaching and learning	Pedagogical opportunities and limitations.
7	Sayed et al. (2024)	Empirical research	Speaking skills, autonomy, and psychological well-being	Motivation, anxiety, autonomy, and willingness to communicate
8	Tai and Chen (2024)	Experimental research	Generative AI chatbots and speaking development	Chatbots and conversational assistants
9	Yang et al. (2022)	Empirical research	AI chatbot as an English conversation partner	Chatbots and conversational assistants
10	Bashori et al. (2022)	Empirical research	Web-based learning and speaking anxiety	Motivation, anxiety, autonomy, and willingness to communicate



No.	Author and year	Type of source	Main focus	Thematic category
11	Chen et al. (2023)	Empirical research	Google Assistant for foreign-language learning	Chatbots and conversational assistants
12	Evers and Chen (2021)	Experimental research	Speech-recognition software and pronunciation	Automatic speech recognition and pronunciation development
13	Hsu et al. (2023)	Design-based study	Task-oriented chatbot for speaking practice	Chatbots and conversational assistants
14	Huang et al. (2023)	Literature review	Applications of AI in language education	Pedagogical opportunities and limitations.
15	Inceoglu et al. (2020)	Empirical research	ASR-supported pronunciation practice	Automatic speech recognition and pronunciation.
16	Ngo et al. (2024)	Meta-analysis	Effectiveness of ASR in pronunciation development	Automatic speech recognition and pronunciation.
17	Sun (2023)	Mixed-methods research	ASR, pronunciation, and speaking skills	Automatic speech recognition and pronunciation.
18	Zhang et al. (2024)	Empirical research	Enjoyment, anxiety, and willingness to communicate	Motivation, anxiety, autonomy, and willingness to communicate
19	Zou et al. (2023)	Empirical research	AI speech evaluation and automated feedback	Fluency, accuracy, and oral performance



Note. The evidence corpus includes 14 primary studies and 5 secondary or conceptual sources. Primary studies were used to determine the direct effects of AI-based applications, whereas reviews, the meta-analysis, and the conceptual article were used for contextual and interpretive support. Their findings were not treated as additional independent empirical evidence when overlap with primary studies was possible.

5. DISCUSSION

The discussion distinguishes between findings derived from primary studies and contributions obtained from secondary and conceptual sources. Primary studies provided direct evidence regarding the influence of AI-based applications on English-speaking development, whereas systematic reviews, the meta-analysis, the literature review, and the conceptual article offered broader interpretations, theoretical support, and pedagogical considerations. This distinction reduced the risk of double counting evidence and strengthened the transparency of the synthesis.

The findings indicate that artificial intelligence-based applications can support the development of English-speaking skills among EFL learners by providing interactive practice, immediate feedback, repeated speaking opportunities, and personalized learning experiences. Chatbots, automatic speech recognition systems, virtual assistants, generative AI tools, and speech-evaluation platforms enabled learners to practise pronunciation, fluency, vocabulary, grammatical accuracy, and oral interaction beyond the limitations of traditional classroom time.

The positive effects reported across the primary studies were not uniform and appeared to depend on several methodological and pedagogical conditions. Longer interventions and repeated practice tended to provide greater opportunities for pronunciation adjustment, fluency development, and confidence building, whereas short interventions offered more limited evidence of sustained improvement. Learners' English proficiency also appeared to influence the outcomes because automated systems may respond differently to beginners, intermediate learners, and students with non-standard accents.

The type of AI tool also shaped the findings. Chatbots and conversational assistants were more closely associated with interaction, confidence, autonomy, motivation, and willingness to communicate, whereas automatic speech recognition tools were more directly related to



pronunciation, intelligibility, and corrective feedback. AI speech-evaluation systems contributed to fluency, grammar, vocabulary, and broader oral performance, but their usefulness depended on the accuracy, clarity, and appropriateness of the feedback provided. Another important contribution was the expansion of speaking practice. AI applications allowed learners to repeat activities, simulate conversations, experiment with language, and receive feedback without depending exclusively on teachers or classmates. This flexibility may strengthen oral confidence and encourage more active participation, particularly among learners who experience fear of making mistakes or speaking anxiety.

The reviewed studies also identified positive effects on motivation, autonomy, enjoyment, and willingness to communicate. Because AI tools provide flexible and comparatively less judgmental environments, learners may feel more comfortable practising at their own pace. However, these affective benefits should not be interpreted as automatic outcomes, since they depend on the quality of the activities, the reliability of the system, the learners' digital skills, and the pedagogical support available.

Pedagogical design was another determining factor. Studies that integrated AI tools with structured tasks, teacher guidance, peer interaction, corrective feedback, and authentic speaking activities reported more meaningful outcomes than approaches based only on isolated repetition or unsupervised use. These findings suggest that the effectiveness of artificial intelligence does not depend exclusively on the technology itself, but also on the duration of its use, learners' proficiency level, the type of application selected, and the way the tool is incorporated into the learning process.

Automatic speech recognition was particularly useful for pronunciation development. Nevertheless, some systems may incorrectly interpret regional accents, non-standard pronunciation, background noise, or incomplete utterances. Inaccurate recognition can confuse learners and reduce the educational value of automated feedback. Therefore, pronunciation practice supported by AI should be complemented by teacher evaluation, peer interaction, and authentic communicative activities.

Several limitations were also identified, including inaccurate or incomplete responses, privacy concerns, unequal access to devices and internet connectivity, excessive dependence on automated systems, algorithmic bias, and limited cultural or emotional understanding.



These limitations reinforce the need to use artificial intelligence as a complementary resource rather than as a replacement for teachers or human interaction.

Overall, AI-based applications can contribute to English-speaking development when they are integrated with clear learning objectives, meaningful oral tasks, reliable feedback, sustained practice, and appropriate pedagogical supervision. The evidence suggests that the strongest outcomes emerge when technological tools are combined with teacher guidance and authentic opportunities for communication.

6. CONCLUSIONS

The reviewed evidence suggests that artificial intelligence-based applications have the potential to support English-speaking development among EFL learners. Most of the included studies reported generally favorable outcomes related to opportunities for oral practice, immediate feedback, pronunciation, fluency, confidence, autonomy, and willingness to communicate.

However, the findings were heterogeneous and should be interpreted with caution. The included sources differed in terms of AI applications, methodological designs, educational populations, English proficiency levels, and intervention duration. In addition, the review combined primary studies with systematic reviews, a meta-analysis, a literature review, and a conceptual article, and it did not conduct an independent meta-analysis. These characteristics limit the generalizability of the conclusions and prevent the establishment of uniform effects across all contexts.

The evidence also suggests that the effectiveness of AI-supported speaking practice depends on the reliability of the technological tool, the quality of automated feedback, the duration and frequency of practice, and the pedagogical guidance provided. Therefore, artificial intelligence should be considered a complementary resource rather than a substitute for teachers, peer interaction, or authentic human communication.

In conclusion, AI-based applications show potential to strengthen English-speaking learning when they are used responsibly, integrated into structured communicative activities, and adapted to learners' needs and proficiency levels. Further experimental and longitudinal research is required to determine the consistency, magnitude, and durability of these effects.



REFERENCIAS BIBLIOGRÁFICAS

- Bashori, M., van Hout, R., Strik, H., & Cucchiarini, C. (2022). Web-based language learning and speaking anxiety. *Computer Assisted Language Learning*, 35(5–6), 1058–1089. <https://doi.org/10.1080/09588221.2020.1770293>
- Chen, H. H.-J., Yang, C. T.-Y., & Lai, K. K.-W. (2023). Investigating college EFL learners' perceptions toward the use of Google Assistant for foreign language learning. *Interactive Learning Environments*, 31(3), 1335–1350. <https://doi.org/10.1080/10494820.2020.1833043>
- Du, J., & Daniel, B. K. (2024). Transforming language education: A systematic review of AI-powered chatbots for English as a foreign language speaking practice. *Computers and Education: Artificial Intelligence*, 6, 100230. <https://doi.org/10.1016/j.caeai.2024.100230>
- El Shazly, R. (2021). Effects of artificial intelligence on English speaking anxiety and speaking performance: A case study. *Expert Systems*, 38(3), e12667. <https://doi.org/10.1111/exsy.12667>
- Evers, K., & Chen, S. (2021). Effects of automatic speech recognition software on pronunciation for adults with different learning styles. *Journal of Educational Computing Research*, 59(4), 669–685. <https://doi.org/10.1177/0735633120972011>
- Fathi, J., Rahimi, M., & Derakhshan, A. (2024). Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions. *System*, 121, 103254. <https://doi.org/10.1016/j.system.2024.103254>
- Ghamrawi, N., Shal, T., Ghamrawi, N. A. R., Abu-Tineh, A., Alshaboul, Y., & Alazaizeh, M. A. (2025). A step-by-step approach to systematic reviews in educational research. *European Journal of Educational Research*, 14(2), 549–566. <https://doi.org/10.12973/eu-jer.14.2.549>
- Gortaire-Díaz, D., Almache-Granda, G., Real-Roby, R., & Mora-Herrera, E. (2026). Bridging the gap: How AI tools support self-directed EFL skill development inside the classroom. *Revista InveCom*, 6(2), e602012. <https://doi.org/10.5281/zenodo.15788013>



- Guillermo Morales, L. E., & Carcausto Calla, W. H. (2025). El uso de chatbots en el aprendizaje de idiomas: Una revisión sistemática. *Revista InveCom*, 5(3), e050372. <https://doi.org/10.5281/zenodo.14532857>
- Hsu, M.-H., Chen, P.-S., & Yu, C.-S. (2023). Proposing a task-oriented chatbot system for EFL learners' speaking practice. *Interactive Learning Environments*, 31(7), 4297–4308. <https://doi.org/10.1080/10494820.2021.1960864>
- Huang, X., Zou, D., Cheng, G., Chen, X., & Xie, H. (2023). Trends, research issues and applications of artificial intelligence in language education. *Educational Technology & Society*, 26(1), 112–131. [https://doi.org/10.30191/ETS.202301_26\(1\).0009](https://doi.org/10.30191/ETS.202301_26(1).0009)
- Inceoglu, S., Lim, H., & Chen, W. H. (2020). ASR for EFL pronunciation practice: Segmental development and learners' beliefs. *Journal of Asia TEFL*, 17(3), 824–840. <https://doi.org/10.18823/asiatefl.2020.17.3.5.824>
- Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, 54(2), 537–550. <https://doi.org/10.1177/00336882231162868>
- Ngo, T. T.-N., Chen, H. H.-J., & Lai, K. K.-W. (2024). The effectiveness of automatic speech recognition in ESL/EFL pronunciation: A meta-analysis. *ReCALL*, 36(1), 4–21. <https://doi.org/10.1017/S0958344023000113>
- 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., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Sayed, B. T., Bani Younes, Z. B., Alkhayyat, A., Adhamova, I., & Teferi, H. (2024). To be with artificial intelligence in oral test or not to be: A probe into the traces of success in speaking skill, psychological well-being, autonomy, and academic buoyancy. *Language Testing in Asia*, 14, Article 49. <https://doi.org/10.1186/s40468-024-00321-0>
- Sun, W. (2023). The impact of automatic speech recognition technology on second language pronunciation and speaking skills of EFL learners: A mixed methods



investigation. *Frontiers in Psychology*, 14, 1210187.
<https://doi.org/10.3389/fpsyg.2023.1210187>

Tai, T.-Y., & Chen, H. H.-J. (2024). Improving elementary EFL speaking skills with generative AI chatbots: Exploring individual and paired interactions. *Computers & Education*, 220, 105112. <https://doi.org/10.1016/j.compedu.2024.105112>

Varsha, P. S., Chakraborty, A., & Kar, A. K. (2024). How to undertake an impactful literature review: Understanding review approaches and guidelines for high-impact systematic literature reviews. *South Asian Journal of Business and Management Cases*, 13(1), 18–35. <https://doi.org/10.1177/22779779241227654>

Yang, H., Kim, H., Lee, J. H., & 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>

Zhang, C., Meng, Y., & Ma, X. (2024). Artificial intelligence in EFL speaking: Impact on enjoyment, anxiety, and willingness to communicate. *System*, 121, 103259. <https://doi.org/10.1016/j.system.2024.103259>

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

CONFLICTS OF INTEREST

None.

FUNDING

Not applicable.

AUTHOR CONTRIBUTIONS

Conceptualization: Margarita Elizabeth Rocha Chasi, Johanna Carolina Pazmiño Freire, Wendy Ximena Ordóñez Pérez, Mónica Katherine Guachi Caiza, and Jairo Mauricio Ases Villacís.

Data curation: Margarita Elizabeth Rocha Chasi, Johanna Carolina Pazmiño Freire, Wendy Ximena Ordóñez Pérez, Mónica Katherine Guachi Caiza, and Jairo Mauricio Ases Villacís.

Formal analysis: Margarita Elizabeth Rocha Chasi, Johanna Carolina Pazmiño Freire, Wendy Ximena Ordóñez Pérez, Mónica Katherine Guachi Caiza, and Jairo Mauricio Ases Villacís.



Writing – original draft: Margarita Elizabeth Rocha Chasi, Johanna Carolina Pazmiño Freire, Wendy Ximena Ordóñez Pérez, Mónica Katherine Guachi Caiza, and Jairo Mauricio Ases Villacís.

Writing – review and editing: Margarita Elizabeth Rocha Chasi, Johanna Carolina Pazmiño Freire, Wendy Ximena Ordóñez Pérez, Mónica Katherine Guachi Caiza, and Jairo Mauricio Ases Villacís.

