SCIENTIFIC CULTURE
www.sci-cult.com

SCIENTIFIC CULTURE, Vol. 12, No. 2.1, (2026), pp. 13124-13136
Open Access. Online & Print

DOI: 10.5281/zenodo.20385299

CHATBOT-ASSISTED EFL SPEAKING DEVELOPMENT: A SYSTEMATIC REVIEW OF PERFORMANCE, AFFECTIVE OUTCOMES, AND PEDAGOGICAL APPLICATIONS

Shevaani G[1*] and Calaivanane R[2]

[1]Research Scholar, Department of English, School of Social Sciences and Languages, Vellore Institute of Technology, Vellore, India. Email: shevaani.g2024@vitstudent.ac.in https://orcid.org/0009-0008-7746-8417
[2]Associate Professor, Department of Languages, School of Social Sciences and Languages, Vellore Institute of Technology, Vellore, India Email: rcalaivanane@vit.ac.in https://orcid.org/0000-0001-8590-4532

Received: 01/03/2026
Accepted: 26/04/2026

Corresponding Author: Dr. Calaivanane R (rcalaivanane@vit.ac.in)

ABSTRACT

This is a systematic review that focuses on the effect of artificial intelligence (AI)-based chatbots on the speaking skills of English as a Second Language (ESL) and English as a Foreign Language (EFL) learners. Based on the PRISMA procedures, 12 empirical studies published in 2020-2026 were identified in the Scopus database and analysed with the help of thematic analysis. The results indicate that AI chatbots have a significant positive impact on speaking performance, especially fluency, pronunciation, vocabulary, and communicative competence, which is mainly achieved by the means of long-term interaction and instant feedback. Moreover, chatbot-mediated conditions lead to positive affective results, such as decreased anxiety and greater readiness to communicate, which help learners to engage. Pedagogically, the most visible affordance is feedback, and personalised learning is not developed. Nevertheless, there are still issues associated with technological constraints, contextual sensitivity, and excessive dependence on AI. Combining performance, affective, pedagogical, and technological aspects, this review fills one of the gaps in the literature and offers a comprehensive picture of the development of speaking with the help of AI chatbots. The results have significant implications on teachers, developers, and researchers who aim to maximise the use of AI tools in language learning.

KEYWORDS: AI chatbots, ESL/EFL Speaking Skills, Thematic Analysis, Affective Factors, Language Learning.

1. INTRODUCTION

This systematic review gives a detailed synthesis of the effects of AI-based chatbots on the development of ESL/EFL students speaking by combining the results in terms of speaking performance, affective, pedagogical, and other related factors and concerns. Referring to 12 empirical studies, the review shows that AI chatbots can be viewed as versatile instruments that will greatly promote language learning and also have some significant limitations that should be carefully addressed.

The results suggest that AI chatbots have a significant impact on the performance improvement in speaking, especially fluency, pronunciation, vocabulary, and communicative competence. The main factor behind the improvement is the higher chances of prolonged interaction, instant feedback, and practice, in light-stress conditions (Grab, 2025; Kim et al., 2021; Waluyo and Pratiwi, 2024). Simultaneously, the review reveals the importance of affective factors as crucial mechanisms of speaking development since, in addition to a decrease in anxiety and stress, a greater desire to communicate are also needed (El Shazly, 2021; Waluyo and Pratiwi, 2024).

Pedagogically, feedback can be viewed as a core affordance of AI chatbots which helps learners to narrow the gap between their linguistic output in real time and facilitates autonomous learning. Nevertheless, the comparatively small focus on personalised learning implies that adaptive capabilities are not well-developed yet and are a significant way to participate in the future. Regardless of these benefits, the review also singles out some enduring difficulties: technological constraints, the inability to be contextually and culturally sensitive, and the possibility that AI tools may be overutilized. The above findings highlight the fact that chatbots technologies should be integrated into the language learning environment in a balanced and context-specific manner.

Notably, the reviewed study fulfils a considerable literature gap by going beyond a disjointed treatment of the reviewed studies and offering an integrative overview of the interrelated aspects of AI chatbot use during language learning. The study provides a more detailed insight into the role of AI-mediated interaction in the development of speaking by incorporating the performance, affective, pedagogical, and technological perspectives. It further supports some of the main theoretical views in the second language acquisition by showing that interaction, output and affective conditions all contribute to language learning processes in AI-enhanced situations.

The results have various implications to practice and research. In the case of education, AI chatbots should be introduced as the addition to the blended learning setting in order to help educators to practice speaking, without sacrificing the important role of human contact. In the case of developers, the design of chatbots needs to be improved in relation to the contextual awareness, accuracy of feedback, and personalization features. In the context of the scientific community, the future research should be dedicated to the longitudinal design and investigate the effects of using chatbots in the long term, and the creation of more sophisticated AI, which is able to facilitate genuine and socially sensitive communication.

Introduction of the artificial intelligence (AI) into the learning environment in the field of education has transformed the learning environment particularly in English as a Second Language (ESL) and English as a Foreign Language (EFL) teaching and learning environment radically. In the recent years, AI-powered technologies have become an increasingly popular subject because they increase the effectiveness of learning the language through the use of interactive, adaptive, and learner-driven systems (Grab, 2025; Waluyo and Pratiwi, 2024). Among these inventions is the chatbots based on AI, which have become one of the most used tools that can be used to communicate in real-time, still get immediate feedback, and be able to extend the language session outside of the classroom (Kim et al., 2021; El Shazly, 2021). Chatbots give learners an opportunity to train speaking in a low-stress and relaxed environment, which was previously a significant issue with limited exposure to natural communication, and as such can be utilized to fix the long-standing drawback of the latter.

An increasing amount of literature indicates that AI chatbots have a positive effect on different aspects of ESL/EFL speaking. The empirical studies show that there is a stronger improvement of fluency, pronunciation, and communicative competence which is frequently explained by the reason that there is always a practice, and the feedback is also immediate (Grab, 2025; Kim et al., 2021). Moreover, it has been reported to be linked to chatbot-mediated communication resulting in the enhanced communication between learners and the increased chances to generate language spontaneously (Waluyo and Pratiwi, 2024). These findings imply

that AI chatbots may be a helpful tool to facilitate active communication and promote learning oral communication.

Besides the language performance, the influence of AI chatbots on the emotional factor has been considered with a lot of attention. The results indicate that the chatbot-based platforms could reduce the state of language anxiety, enhance confidence, and willingness to communicate (WTC) of students (El Shazly, 2021; Waluyo and Pratiwi, 2024). The interaction with the chatbot is permissive and non-judgmental and this will allow the learners to talk without the fear of being judged negatively thereby providing a positive learning environment. Furthermore, the autonomy and flexibility of the chatbot use encourage self-motivation and self-directed learning which subsequently results in the long-term engagement in the language practice (Grab, 2025). AI chatbots offer various affordances that can be employed in pedagogical purposes to teach language. They provide interactive feedback on learning in real-time and in response, enable customized learning, and enable a learner to engage in an interactive dialogue resembling the real-world communication (Kim et al., 2021; El Shazly, 2021). The qualities can make chatbots an extra teaching tool that can be employed to complement the traditional instructional methods. At the same time, the aspect that AI systems can be tailored to the needs of specific learners can also display that they can be utilized to promote differentiated learning and promote learner autonomy.

Despite these benefits, AI chatbot application into ESL/ EFL contexts is not that sailing smoothly. Research has found shortcomings on technological restrictions such as problems of speech recognition precision and situational understanding (Grab, 2025). In addition, the lack of cultural responsiveness and the impossibility of chatbots to provide a more human-like experience has been raised (El Shazly, 2021). Further, some of the findings suggest that the affective advantages may not be universal, and that is why the effects of the AI-mediated environments on the experiences of learners need to be discussed in a more detailed manner.

The existing theoretical models in the second language acquisition can be used to explain the developments. One of the theories that explain the reduction of anxiety in AI-mediated contexts is Affective Filter Hypothesis (Krashen, 1982), according to which the lower the anxiety levels, the more linguistic input can be absorbed and thus language can be acquired. Interaction Hypothesis (Long, 1996) also emphasizes on the interactive exchange that is involved in the process of negotiating the meaning and feedback that culminates into language development. Similarly, the importance of language production in assisting learners to discover the gaps in their interlanguage and rectify their language accuracy is also emphasized on the same note in the Output Hypothesis (Swain, 1985). In addition to this, the Sociocultural Theory (Vygotsky, 1978) points to the importance of interaction and scaffolding in the learning process and states that the adaptive feedback offered by the AI chatbots can help learners to zone of proximal development. Despite the fact that the current literature offers useful information on the role of AI chatbots in language learning, it is still disjointed. Most of the studies tend to focus on the individual dimensions, such as speaking performance, affective factors, or pedagogical affordances without a clear picture of the whole picture. This does not provide a holistic view of the concomitant effect of AI chatbots on various dimensions of ESL/EFL speaking growth. As a result, a systematic review that would bring these dimensions together in a single analytical framework is evident.

The proposed study will address this gap, as it will provide the synthesis of the impact of AI-based chatbots on the ESL/EFL learners speaking skills. Specifically, the outcomes of speaking work, affective elements, pedagogical responsibilities, and the problems are included in the review.

In achieving this objective, the research questions that will be answered in the study will be as follows:

1. How can AI-based chatbots impact the speaking performance of ESL/EFL learners, including fluency, pronunciation, and communicative competence?
2. What is the impact of AI-based chatbots on affective aspects of learners such as anxiety, confidence, and willingness to interact?
3. What are the pedagogical roles of AI chatbots to help with ESL/EFL speaking instruction?
4. What are the limitations and difficulties of applying the AI chatbots in the development of speaking skills?

By providing an answer to these questions, this systematized review will contribute to the field by providing a synthesis of the fragmented research and a more detailed image of language learning with the use of AI chatbots. The study also relates the

theoretical experience and empirical data, and gives practical suggestions to the teachers, researchers, and developers interested in making the most of AI technologies in teaching speaking English as a second language.

2. METHODOLOGY

2.1. Research Design

The study will be systematic review design with the qualitative thematic synthesis approach to explore the ability of artificial intelligence (AI)-based chatbots to affect the speaking skills of English as a Foreign Language (EFL) and English as a Second Language (ESL) students. The systematic review is particularly appropriate in case the diffused empirical data is to be combined, and the unified cognition of the novel fields of investigation, such as AI-assisted language learning, is to be developed.

The review is founded on the guidelines of Preferred Reporting Items of Systematic Reviews and Meta-Analyses (PRISMA) to ensure the transparency of the methodology, rigor, and replicability. The systematic and progressive approach was taken, and it included database search, screening, eligibility assessment and final inclusion of the studies. The given steps were intended to minimize selection bias and enhance the review process.

In addition, thematic analysis was employed since it allows recognition, coding and explanation of common trends in heterogeneous studies in a logical way. The nature of AI chatbot research designs, contexts, and outcome measures is too varied, which means that a qualitative synthesis approach is more appropriate than a quantitative aggregation approach since the former enables a more profound conceptualization of the latter.

2.2. Database And Search Strategy

The Scopus database was chosen as the source of a systematic literature search due to its wide range of high-quality peer-reviewed articles in the field of applied linguistics and educational technology. The database was considered suitable in terms of capturing the relevant and recent studies in the interdisciplinary field of AI and language learning.

The search plan was created to find the studies that concentrate on AI chatbot applications to language learning, especially speaking skills. The search query was designed according to three main areas of conceptualization, including language learning, AI chatbots, and speaking skills. Relevant synonyms were added to make the search more comprehensive and Boolean operators (AND, OR) were used to organize and narrow down the search.

The search query was formulated as follows:

TITLE-ABS-KEY (
("language learning" OR "EFL" OR "ESL")
AND ("chatbot" OR "AI chatbot" OR "conversational agent")
AND ("speaking skills" OR "oral communication")
)
AND PUBYEAR > 2020 AND PUBYEAR < 2027
AND (LIMIT-TO (SRCTYPE, "j"))
AND (LIMIT-TO (PUBSTAGE, "final"))
AND (LIMIT-TO (DOCTYPE, "ar"))
AND (LIMIT-TO (LANGUAGE, "English"))

The search was restricted to articles published in the past 2020-2026, restricted to peer-reviewed journal articles in English and in final publication stage. This period was chosen to reflect the latest trends in AI chatbots technologies. The first search resulted in 45 records.

2.3. Study Selection Process

The process of study selection was based on PRISMA and included several steps to make the process transparent and replicable. First, 45 records were found by searching Scopus databases, and no duplicates were found.

Title screening was conducted in the first stage and in this case, 27 studies were excluded because they were not relevant to the research focus. The second stage involved abstract screening in which the rest of the 18 studies were screened, and four more studies were excluded due to the failure to pass the inclusion criteria. Then, 14 full-text articles were evaluated concerning eligibility. Among them, two studies were eliminated because of inaccessibility even after reasonable attempts to access them.

Finally, 12 studies were incorporated in the review. A PRISMA flow diagram is used to depict the entire process of the selection in Figure 1.

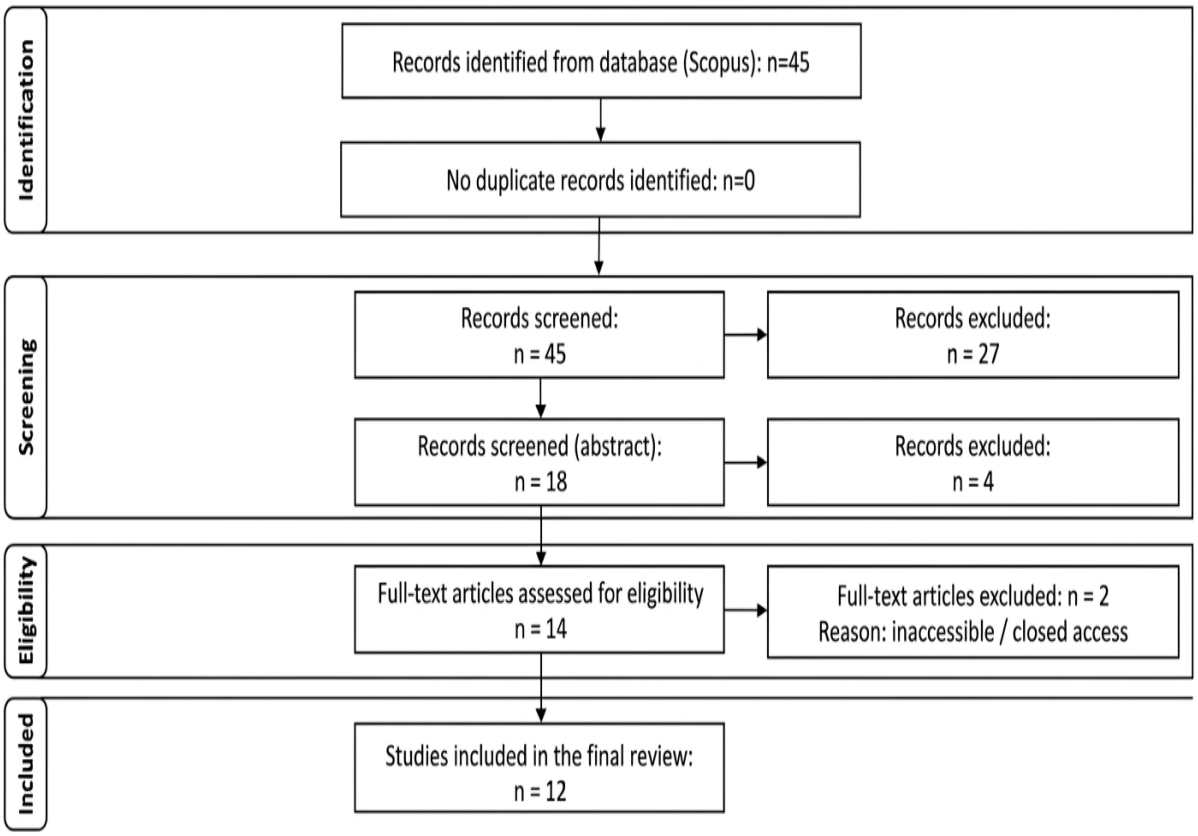

Figure 1: PRISMA Flow Diagram Representing the Data Selection Process.

The screening process was done in a systematic manner according to the set predefined criteria in order to increase the methodological rigor and minimize the possible bias. To increase the reliability, a sample of the screening decisions was checked on consistency, and any doubt was overcome by closely re-examining the inclusion criterion. This exercise provided a uniform and clear-cut process of study selection.

2.4. Inclusion And Exclusion Criteria

In order to be consistent and methodologically rigorous, before the screening process, clear inclusion and exclusion criteria were set. These criteria were used to select the relevant studies and made sure that they were in line with the objectives of the research. Table 1 indicates the inclusion and exclusion criteria.

Table 1: Inclusion And Exclusion Criteria.

Criteria	Included	Excluded
Publication Year	2020–2026	Before 2020
Language	English	Non-English
Source Type	Peer-reviewed journal articles	Conference papers, books, theses
Publication Stage	Final publication	In press / unpublished
Topic	AI-based chatbots in ESL/EFL speaking contexts	Not related to AI chatbots or speaking
Skill Focus	Speaking skills / oral communication	Only reading, writing, or listening
Participants	ESL/EFL learners	Non-language learners
Study Type	Empirical studies	Reviews, conceptual papers
Accessibility	Full-text available	Full-text not accessible despite reasonable efforts

2.5. Data Extraction

After the selection process, data were systematically retrieved in the 12 studies included in the study through a structured data extraction framework. This strategy provided uniformity, openness, and comparability of studies.

The information extracted consisted of bibliographic (author(s) and year of publication),

research context (country and educational level), and the characteristics of participants (sample size and learner type). The technological features were also recorded such as the type and capabilities of AI chatbots.

Moreover, such pedagogical factors like the instructional design, the nature of intervention, and the focus on the specific speaking skills were noted. Such methodological characteristics as research design (quantitative, qualitative, or mixed methods) and data collection instruments (e.g., tests, surveys, interviews) were extracted.

Data related to the outcome were also recorded, especially the variables that were related to speaking performance and the main findings that were reported in the studies. All the data extracted were systematised in order to carry out a thematic analysis later. The information that was not clear or was missing was noted to ensure transparency and analytical rigor.

2.6. Data Analysis and Coding Procedures

The identified literature was analysed with the help of thematic analysis with the aid of NVivo software. The inductive strategy was used in which the themes were formed out of the data as opposed to the predetermined ones.

The six-phase framework by Braun and Clarke (2006) was used to guide the coding process, which guaranteed a rigorous and systematic analytical process. The analysis was done in numerous steps.

In the first step, open coding was conducted to determine the common concepts and patterns in the studies. Axial coding was then used to cluster similar codes into general categories and identify connections among ideas. Selective coding was applied in the last phase to narrow down and combine these categories into broad themes.

This multi-stage and iterative coding procedure increased the level of analysis, conceptual clarity as well as consistency throughout the results.

2.7. Development Of Thematic Categories

The thematic analysis led to the discovery of a few broad categories that describe the complex effect of AI chatbots on language learning.

The first one, speaking performance, incorporates the accuracy, fluency, pronunciation, vocabulary and intonation dimensions. The second type is affective factors and includes variables like anxiety, stress and willingness to communicate. The third category is known as pedagogical roles and it encompasses feedback and personalised learning. Lastly, the category of challenges and limitations was introduced, and it included the technological and contextual constraints that were reported in the studies.

In order to boost analytical transparency, the frequency of each theme in all studies is provided in the Results section.

2.8. Reliability And Validity

The coding process was done in a systematic and iterative manner in order to guarantee the reliability and validity of the analysis. Codes and categories were constantly checked and modified to be more consistent and less ambiguous.

Transparency and replicability were improved through the use of a formal coding scheme and NVivo software. Moreover, the screening and coding processes were also subjected to consistency checks to reduce subjectivity and make them consistent with the previously established criteria. Moreover, the patterns that were discovered in several studies were cross-validated to guarantee the strength of the results, which enhanced the validity of the thematic interpretations.

2.9. Data Synthesis

The extracted and coded data were synthesised using a thematic synthesis approach. The findings were organised into major thematic categories, enabling the identification of patterns, comparison across studies, and interpretation of the effectiveness of AI chatbots in enhancing speaking skills.

This integrative approach provides a comprehensive understanding of the role of AI chatbots in ESL/EFL contexts by combining qualitative insights with frequency-based evidence.

3. RESULTS

3.1. Study Characteristics

Table 2 displays the nature of the studies included. Twelve studies were incorporated, and they represented different geographical settings, population of the participants, and research designs. The studies included experimental, quasi-experimental, qualitative, survey-based, and systematic review designs, including participants of primary school learners and university students and adult learners. The AI applications used were GPT-based conversational agents, speech-recognition chatbots, as well as custom-built AI platforms. Together, these works can be considered the widening of the range of AI application in the language learning setting and its variety.

Table 2: Characteristics Of Included Studies.

Author & Year	Country	Participants	Level	AI Tool Used	Research Design	Focus (Speaking Skill)	Key Outcome
Jeon, Lee & Choe (2023)	USA / South Korea	37 studies	Mixed	Speech-recognition chatbots	Systematic Review	Interaction, WTC, anxiety	Improved interaction, reduced anxiety
Wang et al. (2024)	China	99 students	University	Typebot, D-ID Agent	Experimental	WTC, anxiety, performance	Improved WTC, reduced anxiety
Jeon (2024)	South Korea	36 learners	Primary	Dialogflow chatbot	Qualitative	Interaction, motivation	Mixed affordances
Yuan (2023)	China	74 students	Primary	AI chatbot	Experimental	Oral proficiency, WTC	Improved speaking and WTC
Xing & Saeed (2025)	Malaysia	55 studies	Mixed	Various AI tools	Systematic Review	Fluency, pronunciation	Improved speaking, some challenges
Waluyo & Pratiwi (2025)	Asia	Multiple studies	Mixed	ChatGPT, Alexa	Meta-synthesis	WTC, fluency	Improved confidence and WTC
Lin (2025)	New Zealand	3 teachers	Teacher	ChatGPT	Case Study	Interaction	Supports communication
Kim et al. (2021)	South Korea	49 students	University	AI chatbot	Experimental	Fluency, intonation	Improved speaking skills
El Shazly (2021)	Egypt	48 students	University	AI chatbot	Quasi-experimental	Anxiety, performance	Improved performance, anxiety unchanged
Alenezi & Alenezi (2025)	Saudi Arabia	30 students	Secondary	GPT chatbot	Quasi-experimental	Conversation skills	Improved competence
Grab (2025)	Turkey	120 students	University	Multiple chatbots	Experimental	Fluency, pronunciation	Improved all speaking skills
Jiang et al. (2025)	Southeast Asia	737 learners	Mixed	Various chatbots	Survey (SEM)	Interaction, motivation	Motivation drives usage

Figure 2 illustrates the distribution of the types of AI chatbots utilized in the included studies. The commonly used chatbots were GPT-based and custom-developed, and less commonly used were speech-recognition-based and mixed AI systems. This trend is indicative of the modern trends in technology in language learning with the help of AI

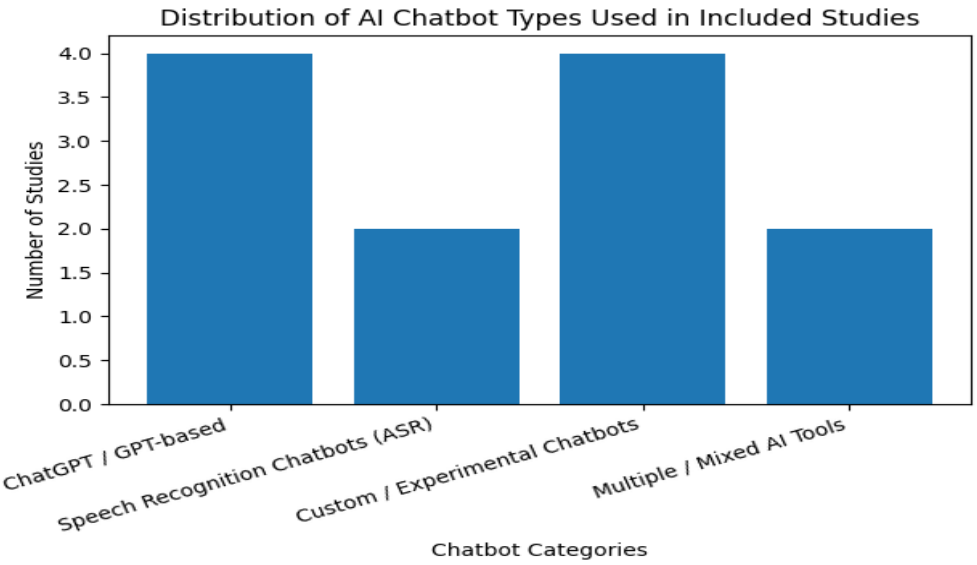

Figure 2: Distribution Of AI Chatbot Types Used in Included Studies.

3.2. Thematic Synthesis of Findings

The results were generalized into four broad themes, including affective factors, speaking performance, chatbot pedagogical role, and challenges and limitations. The combination of these themes reflects both the pedagogical opportunities and limitations of AI-assisted language learning that are highly complex in terms of the overall effect that the introduction of chatbots has on the speaking process of learners.

Table 3: Thematic Summary of Findings.

Theme	Subtheme	n (References)	No. of Studies (Files)
Affective Factors	Anxiety	96	12
	Stress	21	8
	Willingness to Communicate	74	6
Speaking Performance	Accuracy	42	11
	Fluency	66	10
	Intonation	25	5
	Pronunciation	82	11
	Vocabulary	73	12
Chatbot Pedagogical Role	Feedback	135	12
	Personalized Learning	5	3
Challenges and Limitations	Challenges	31	11
	Limitations	60	11

As shown in Table 3, feedback (n = 12 studies) and anxiety (n = 12 studies) turned out to be the most prevalent ones, and personalized learning (n = 3 studies) was given relatively little coverage. This trend highlights a high research focus on short-term pedagogical assistance and affective aspects, at the same time demonstrating a relative lack of adaptive and personalized learning system development.

3.3. Affective Factors: Anxiety, Stress, And Willingness to Communicate

The majority of studies revealed that AI-based tools were effective in the improvement of affective conditions of learners by lowering anxiety and stress and increasing the willingness to communicate. These affective variables were strongly connected with each other and were also at the centre of the development of the learner engagement in speaking activities.

One of the most common results was the decrease of speaking anxiety (n = 12). Traditional classroom setting learners tended to be anxious because of the fear of being negatively evaluated, peer pressure, and lack of practice. Conversely, AI-mediated settings offered a low-pressure and non-judgmental setting, which allowed learners to participate in speaking practice without any fear of making mistakes. This decrease was largely explained by the privateness of the interaction, the fact that most people had plenty of practice opportunities and that there was no social comparison in the present moment. Besides, the availability and responsiveness of AI tools at all times helped to build confidence and overcome emotional barriers.

Likewise, a number of studies also indicated that a stress-free learning environment was developed (n = 8). The versatility and availability of AI tools enabled learners to learn at their speed, thus eliminating stress related to performance. This helped in maintaining the attention and making the learners more comfortable when performing speaking activities.

The other important conclusion was the willingness to communicate increased (n = 6). The lower the anxiety and the higher the confidence, the more willing the learners were to start communication and maintain it. The process was facilitated by AI tools through the creation of genuine interaction, instant feedback, and a positive communicative atmosphere.

The decrease in anxiety and stress seems to operate as one of the major processes that underlie the growth in the willingness to communicate, which suggests that affective gains are facilitators in the participation of learners. Nonetheless, few studies alluded that these results are dependent on the quality of interaction and system design. Besides, there is scanty evidence on long-term sustainability and transferability of such affective gains.

On the whole, these results prove that AI-mediated settings are a good way to decrease affective barriers and increase communicative preparedness. These affective improvements are closely linked with those of speaking as will be discussed in the next section.

3.4. Speaking Performance: Fluency, Pronunciation, Vocabulary, Accuracy, And Intonation

The majority of the studies showed that AI-based

tools were very helpful in improving speaking performance of learners in various aspects such as fluency, pronunciation, vocabulary, accuracy, and intonation.

One of the main results was the fluency (n = 10), which was mainly motivated by more chances to practice in a repetitive manner and engage in real-time communication. Students could speak in a more natural and effortless manner in low stress conditions.

Other common improvements included pronunciation (n = 11), which was complemented by the speech recognition technologies and instant corrective feedback. Nevertheless, the reliability of feedback was sometimes limited by the inaccuracy of recognition.

Likewise, the majority of studies indicated improvement in vocabulary development (n = 12), which was achieved by interaction in a contextualized manner and repetition of lexical items.

The other important area that was improved was accuracy (n = 11), especially grammatical structure and sentence construction. The presence of instant feedback allowed the learners to perfect their output and acquire more linguistic accuracy.

Regarding the intonation dimension (n = 5), a positive change was noted in case the AI tools were equipped with voice-based interaction, but this aspect was relatively under researched.

Although these results were positive, it was observed that there were differences depending on the level of proficiency of learners and technology limitations. There is limited evidence on the transferability of speaking gains in the long-term.

Altogether, these results suggest that positive changes in speaking performance are strongly associated with the possibility to practice, the system of feedback, and learning in interactive conditions. Chatbots, especially feedback and personalization, are strongly supported by the pedagogical roles of chatbots.

3.5. Chatbot Pedagogical Role: Feedback and Personalized Learning

The majority of the studies proved that AI-based chatbots play a core pedagogical role with the help of feedback and individualized learning processes.

One of the prevailing results was the significance of feedback (n = 12) as the main pedagogical force. Chatbots offered multimodal, continuous, and immediate feedback and allowed learners to recognize and rectify errors immediately. This helped in faster learning and better language accuracy.

The feedback was also improved with the assistance of repetitive practice opportunities, continuous interaction, and non-judgmental environment that facilitated active engagement in learners. Also, feedback helped in learner autonomy and self-regulated learning.

On the contrary, personalized learning (n = 3) was less commonly covered. Even though there were systems that were modified to suit the proficiency levels and learning pace of the learners, little was done on the implementation of personalization.

There were certain differences, especially in terms of feedback accuracy and contextual appropriateness. Unstable or too generic feedback decreased effectiveness in some instances.

On the whole, feedback as the main pedagogical process and individualized learning as the underdeveloped but potentially promising field can be identified. Nonetheless, there are a number of challenges and limitations that are associated with these pedagogical benefits and they are as follows.

3.6. Challenges And Limitations

The majority of the studies found a list of difficulties and constraints related to the use of AI tools in language learning.

Practical difficulties (n = 11) were the problems with adaptation, task design, and limitations on infrastructure. The technological constraints (n = 11) were mainly associated with failures in speech recognition and situational comprehension.

Emotional intelligence was also weak in AI systems, which do not offer empathetic and contextually relevant responses. Other issues were the overdependence on AI, the decrease in human contact, privacy, and access.

Even though there were learners who became accustomed to these limitations, more general methodological limitations, including small samples and limited intervention time, restrict the ability to generalize results.

In general, even though AI technologies have significant potential in improving language learning, their usefulness is limited by technical, pedagogical, and contextual considerations. These issues should be resolved in order to have a more efficient and sustainable integration.

Combined, the results indicate that AI-based chatbots have a complex effect on language learning as they positively affect the affective conditions, facilitate pedagogical mechanisms, and improve the performance of speaking, yet are limited by technological and contextual factors.

4. DISCUSSION

This systematic review summarises the empirical studies on the role of AI chatbots in the process of ESL/EFL speaking development by combining the results concerning the speaking performance, affective factors, pedagogical affordances, and technological constraints. Though in many cases previous research has focused on these aspects separately, the current review reveals that all these dimensions are connected and influence each other as a whole to determine the success of AI-assisted language learning.

One of the main contributions of this review is that it can reveal constant patterns in the studies, as well as outline significant limitations and gaps in the existing research environment. On the whole, the results indicate that AI chatbots do not only serve as technological resources, but as pedagogically significant spaces, which affect cognitive and affective aspects of language acquisition.

4.1. Influence On Speaking Performance

The results suggest that AI chatbots can lead to the quantifiable change in the speaking performance of learners, especially in the areas of fluency, pronunciation, grammatical accuracy, and confidence. These advances are explained by the Output Hypothesis, which states that language production allows identifying, testing hypotheses, and reorganizing linguistic knowledge (Swain, 2020).

Throughout the analyzed literature, a general trend can be observed where repetition and instantaneous feedback seem to be the most important processes that trigger progress. This indicates that the success of AI chatbots does not strongly hinge on the level of technological advancement but is more directly associated with the number and quality of interactions between the learner. In this regard, chatbot mediated practice generates conditions that are similar to sustained output and iterative refinement, which are the focus of language development.

Moreover, the results were in line with the Interaction Hypothesis (Long, 1996), since AI chatbots model conversational interactions that promote meaning negotiation. These interactional opportunities seem to facilitate the process of communicative competence development since they allow learners to have meaningful dialogue without time or social pressure of the classroom.

Nonetheless, it is also found that there is an unequal development in language skills. Although productive skills like speaking demonstrate steady improvement, receptive and pragmatic aspects are not thoroughly investigated and improved in a more consistent way. It implies that chatbot-based learning can be biased towards linguistic production, at the expense of socially situated communication, which can be seen as a weakness of existing applications.

4.2. Affective Factors and Learner Psychology

The review highlights an important role of affective factors in the mediation of the effectiveness of the use of AI chatbots. In line with Affective Filter Hypothesis (Krashen, 1982), the less anxiety and high confidence, the more participation and language production.

An evident trend in the literature suggests that AI chatbots establish a non-judgmental and low-pressure atmosphere, which motivates learners to be riskier and more active in speaking activities. This decrease in performance anxiety seems to be a decisive issue behind the gains in speaking proficiency. Moreover, a higher willingness to communicate implies that affective support offered by chatbots is not just one of the side effects, but a key mechanism that determines learning outcomes.

Simultaneously, the positive results of the affective outcome are not consistently favourable. Other studies indicate that anxiety is alleviated but not completely removed, and that in the long run, the engagement with AI systems could result in the lack of novelty and interest. Such results suggest that the affective benefits are relative and could have to be pedagogically designed to ensure that the motivation of the learners is maintained.

4.3. Pedagogical Affordances of AI Chatbots

The results emphasize the existence of a number of pedagogical affordances that make AI chatbots effective in language learning. It is important to note that one of the key features is the ability to provide immediate and personalised feedback. The feedback types assist in self-regulated learning and allow learners to participate in the iterative improvement which is usually hard to accomplish in the traditional classroom environment.

Moreover, chatbots enable adaptable and self-directed learning because a learner can practice at his/her pace. This is in line with learner-centred pedagogies and differentiated instructions. The communicative language teaching principles are also consistent with the possibility of chatbots to simulate conversational interaction since the learners are subjected to meaningful and contextually relevant language use.

Nonetheless, the review also indicates that the pedagogical usefulness of chatbots is contingent on

their implementation in the instructional settings. Chatbots can be applied as a supplementary tool, not as a standalone solution. Teacher mediation is also needed to deal with complicated linguistic and sociocultural elements of communication, and to keep the learners engaged and offer contextual advice.

4.4. Challenges And Limitations

The AI chatbots have various limitations that limit their performance in spite of their potential. A common problem with research is that there is little contextual and cultural sensitivity of chatbot response. Although chatbots may mimic interpersonal communication, they are usually not as rich and flexible as human interaction should be.

Other technical constraints such as inaccuracy in speech recognition and limited conversational range also impact the quality of interaction. Such restrictions can reduce the possibilities of the higher-order communicative competence development, especially in the real-life situations.

The other important issue is the possibility of overdependence on AI tools. Although chat bots offer a great chance at practice, overreliance can limit learners' exposure to natural human interactions which is crucial in achieving pragmatic and sociocultural competence. This shows the necessity of the balanced representation of AI tools as a part of the extended pedagogical approaches.

4.5. Addressing The Research Gap

The review is valuable to the literature since it offers a comprehensive view of the use of AI chatbots in language learning. This review shows that cognitive, affective, and pedagogical dimensions are interrelated as opposed to earlier research that examines only one of the variables of performance, motivation, etc.

The results indicate that the levels of speaking performance are highly interconnected with lower anxiety levels and more chances of interacting and producing. Meanwhile, the review indicates the paucity of studies that consider the interaction of these dimensions over time and in a natural classroom environment. It means that more holistic and context-sensitive methods should be used in further studies.

4.6. Implications

4.6.1. Pedagogical Implications

AI chatbots must be considered as the additional tools that are to be integrated in ESL/EFL classrooms. Their ability to offer low-pressure interactive practice is what makes them especially effective in the development of fluency and confidence. Learning methods where AI tools are used together with teachers will provide more balanced and sustainable learning results.

4.6.2. Theoretical Implications

The results support major second language acquisition theories. Output Hypothesis is used to describe the enhancement of speaking performance, Interaction Hypothesis is applied to explain the role of conversational engagement, and Affective Filter Hypothesis is used to indicate the significance of emotional factors in the process of language learning. More importantly, the review speculates that such theoretical frameworks do not work in isolation but do interact in the AI-mediated environments.

4.6.3. Practical Implications

In the case of the developers, enhancing the accuracy of chatbots, their contextual awareness, and adaptive abilities is vital. In the case of educators, the integration strategies and training are required to ensure that the pedagogical value of AI tools is maximised and the limitations of the latter are minimised.

4.7. Limitations And Future Research

There are a number of limitations to this review. The number of included studies is relatively small and the interventions chosen are mostly short-term, which limits the generalisability of the results. Moreover, differences in research design and outcome measures in different studies are a challenge to synthesis.

Future studies need to focus on longitudinal studies that could test the long-term effects of AI chatbots on language acquisition. It is also necessary to have the studies that investigate the real-world classroom integration and the formation of pragmatic and sociocultural competence. Additionally, new AIs with greater versatility and emotional intelligence should be researched more to get a clearer picture of their capabilities in language learning.

5. CONCLUSION

The systematic review offers a synthesis of the effects of AI-based chatbots on the speaking development of ESL/EFL learners through the incorporation of results related to speaking performance, affective aspects, pedagogical affordances, and the related issues. Referring to 12 empirical studies, the review proves that AI chatbots

are complex instruments that may greatly improve language acquisition and also pose significant drawbacks that are to be considered attentively.

The results have shown that AI chatbots make significant contributions to the enhancement of speaking performance, especially fluency, pronunciation, vocabulary, and communicative competence. The primary causes of these improvements are the enhanced chances of constant interaction, instant feedback, and drilling in low-pressure settings (Grab, 2025; Kim et al., 2021; Waluyo and Pratiwi, 2024). Simultaneously, the review also emphasizes the importance of affective factors, demonstrating that the decreases in anxiety and stress levels as well as the more willingness to communicate are the most significant processes that enable speaking to develop (El Shazly, 2021; Waluyo and Pratiwi, 2024).

Pedagogically, feedback is the key affordance of AI chatbots, which helps learners to correct their linguistic output in real-time and facilitates self-directed learning. The comparatively low focus on personalised learning, however, implies that adaptive capabilities are not developed and that it is a promising focus of innovation in the future. Along with these benefits, the review also states that there are still unresolved challenges, such as limitations in the use of technology, insensitivity to the context and culture, and the possible danger of excessive dependence on AI tools. These results highlight the importance of the balanced and context-sensitive combination of chatbot technologies in the language learning setting.

Notably, the reviewed paper fills an important knowledge gap in the literature as it goes beyond the disjointed research efforts and offers a composite overview of the inter-relationship between the various dimensions of AI chatbot as a language learning tool. The combination of performance, affective, pedagogical, and technological views provide the study with a more in-depth insight into the ways in which AI-mediated interaction facilitates the development of speaking. This way, it also supports the major theoretical views in second language acquisition, proving that the combination of interaction, output, and affective conditions can influence language learning processes in AI-mediated situations.

The research findings have various implications on practice and research. To teachers, AI chatbots must be planned as an addition to other learning resources in a blended learning setting, which can facilitate speaking skills without ruining the critical element of human communication. Among developers, the development of chatbots should be improved in terms of contextual awareness, accuracy of feedback, and personalization. As a researcher, the next step in the study should be related to longitudinal designs to investigate the effect of using chatbots in the long term and the creation of more sophisticated AI systems that can facilitate genuine and socially rich communication.

To sum up, it is possible to state that AI-based chatbots are a prospective but developing resource in the teaching of ESL/EFL speaking. Although they can have great pedagogic and affective advantages, their success will be determined by careful incorporation, technological excellence, and further empirical research. This study offers a more holistic synthesis of existing studies and, thus, helps to obtain a more in-depth and sophisticated insight into the current research, as well as provides a reference point to further improvements in the domain.

REFERENCES

Alenezi, A., & Alenezi, A. (2025). Evaluating the effectiveness of chatbot-assisted learning in enhancing English conversational skills among secondary school students. *Education Sciences*, *15*(9), 1136.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

El Shazly, R. (2021). Effects of artificial intelligence on English speaking anxiety and speaking performance: A case study. *Expert Systems*, *38*(3), e12667.

Grab, M. Ö. (2025). Integrated AI chatbot practice: A pathway to improved ESL speaking skills. *Social Sciences & Humanities Open*, *12*, 101933.

Hinkel, E. (Ed.). (2011). *Handbook of Research in Second Language Teaching and Learning: Volume 2*. Routledge.

Jeon, J. (2024). Exploring AI chatbot affordances in the EFL classroom: Young learners' experiences and perspectives. *Computer Assisted Language Learning*, *37*(1-2), 1-26.

Jeon, J., Lee, S., & Choe, H. (2023). Beyond ChatGPT: A conceptual framework and systematic review of speech-recognition chatbots for language learning. *Computers & Education*, *206*, 104898.

Jiang, S., Numtong, K., & Asavaratana, W. (2025, November). Toward sustained interaction: investigating the drivers of continued use of AI chatbots for language learning in Confucius Institutes. In *Frontiers in Education* (Vol. 10, p. 1656204). Frontiers Media SA.

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

Krashen, S. (1982). Principles and practice in second language acquisition.

Lin, J. (2025). Diffusion of Chatbots as Conversational Partners in Chinese Language Teaching: Insights From a Case Study. *Journal of Technology & Chinese Language Teaching*, *16*(1).

Long, M. (1996). The role of the linguistic environment in second language acquisition. *Handbook of second language acquisition*.

Sales, A. (2020). The Output Hypothesis and its influence in the second language learning/teaching: An interview with Merrill Swain. *Interfaces Brasil/Canadá*.

Waluyo, B., & Pratiwi, D. I. (2025). AI chatbot-assisted English learning and willingness to communicate: A narrative meta-synthesis of evidence from Asian English as a foreign language contexts. *The JALT CALL Journal*, *21*(3), 102884-102884.

Wang, C., Zou, B., Du, Y., & Wang, Z. (2024). The impact of different conversational generative AI chatbots on EFL learners: An analysis of willingness to communicate, foreign language speaking anxiety, and self-perceived communicative competence. *System*, *127*, 103533.

Xing, C., & Saeed, M. A. (2025). A Systematic review on artificial intelligence (AI) technologies in ESL/EFL speaking skills. *International Journal of TESOL Studies*, 1-31.

Yuan, Y. (2024). An empirical study of the efficacy of AI chatbots for English as a foreign language learning in primary education. *Interactive Learning Environments*, *32*(10), 6774-6789.