



Reframing GenAI-mediated EFL speaking: A systematic review of empirical studies through an Activity Theory lens

Tiantian Tang¹ and Supyan Hussin²

¹Faculty of Education, Universiti Kebangsaan Malaysia, Bangi, Malaysia (ORCID: 0009-0006-9535-4881)

²Faculty of Education, Universiti Kebangsaan Malaysia, Bangi, Malaysia (ORCID: 0000-0003-0544-0649)

Corresponding Author: Supyan Hussin, supyan@ukm.edu.my

Submitted: 17 August 2026 Revised: 7 October 2026 Accepted: 9 October 2026 Published online: 25 October 2026

Abstract

The rapid development of generative artificial intelligence (GenAI) has created more possibilities for English as a foreign language education. Through its generative capabilities, particularly in natural language generation and processing, GenAI has introduced new modes of interaction for learners' speaking development. However, empirical research on GenAI-mediated EFL speaking remains fragmented. Guided by PRISMA 2020, this systematic review synthesizes 43 empirical studies published between 2024 and 2026 to examine the trends, application methods, and challenges of GenAI-mediated EFL speaking activities. Findings indicate that research has focused primarily on mainland China and intermediate-level undergraduate learners, with ChatGPT as the most frequently used tool. GenAI was commonly used to provide personalized practice and immediate feedback in a comparatively low-pressure environment. Most studies reported positive outcomes, but findings varied. Behavioral outcomes received less attention than cognitive and affective aspects. Challenges mainly involved inauthentic interaction and limited feedback. Tensions are mainly from tools and subjects, as well as tools and objects from an Activity Theory lens. Based on these tensions, this study proposes a collaborative learning framework to provide complementary mediation in EFL speaking activities.

Keywords: Active Theory; Collaborative learning; English as a foreign language; Generative artificial intelligence; Speaking activities; Systematic literature review

1. Introduction

In some EFL contexts, learners have limited opportunities for authentic spoken interaction despite years of formal English instruction (Ma et al., 2025). Digital technologies have expanded opportunities for speaking practice beyond the classroom, although sustained engagement with these resources may require instructional support (Guo et al., 2026). More recently, GenAI has introduced new possibilities for responsive conversational practice and individualized feedback (Tai & Chen, 2025). These opportunities are particularly relevant given the persistent challenges associated with foreign language speaking anxiety (Kayhan, 2025). Consequently, growing research attention has focused on the potential of GenAI-supported interaction to enhance EFL speaking proficiency and learners' willingness to communicate (Tai & Chen, 2025; Zou et al., 2025).

Research on AI-assisted language learning has expanded rapidly, with existing reviews documenting the applications, outcomes, and challenges of conversational AI in English language teaching and EFL speaking practice (Lai & Lee, 2024). More recent reviews have also examined AI technologies specifically in ESL/EFL speaking development (Xing & Saeed, 2026). However, a more focused understanding is needed of how GenAI-mediated speaking activities are pedagogically organised and how their technological, social, and instructional components interact. Activity Theory [AT] offers a framework for examining these relationships by considering learners, mediating tools, learning objectives, rules, community, division of labour, and outcomes as interconnected elements (Yang & Kyun, 2022). Its relevance to AI-mediated EFL speaking activities has also been demonstrated in empirical research (Ebadi et al., 2025). Building on this literature, the present review synthesises 43 empirical studies of GenAI-mediated EFL speaking activities through an AT lens to examine research trends, pedagogical applications, reported outcomes, and implementation challenges in relation to the broader activity systems in which they occur. It is guided by the following research questions:

RQ 1) What are the research and publication trends of empirical studies on GenAI-mediated EFL speaking activities?

RQ 2) How is GenAI integrated and used to mediate EFL speaking activities?

RQ 3) What challenges have emerged in GenAI-mediated EFL speaking activities?

2. Literature Review

Artificial intelligence [AI] broadly refers to computational systems designed to perform tasks associated with human intelligence, including learning, reasoning, and language processing (Russell & Norvig, 2021). Its conceptual foundations can be traced to Turing's imitation game, proposed in 1950 and later known as the Turing Test (Gonçalves, 2023; Turing, 1950), while the Dartmouth Summer Research Project in 1956 marked the emergence of AI as a distinct research field (McCarthy et al., 2006). An early example of human-computer interaction was ELIZA, developed by Weizenbaum in 1966, which simulated natural language conversation through keyword-based decomposition and response-generation rules rather than semantic understanding (Weizenbaum, 1966). Subsequent developments expanded AI applications in language education to include natural language processing, automated feedback, and personalized learning (Doğan & Talan, 2025; Huang et al., 2023). More recently, GenAI has introduced new possibilities for natural language generation and interactive communication, creating opportunities for more responsive and personalized language-learning experiences (Kohnke et al., 2023).

Unlike many earlier AI systems designed to follow predefined rules or produce fixed responses, GenAI offers stronger natural language processing, contextual interaction, and content generation capabilities (Wang et al., 2024). For instance, ChatGPT, a representative GenAI application, has provided learners with more flexible and personalized opportunities for language interaction (Tai, 2026). However, early GenAI applications were predominantly text-based, which contributed to earlier and more extensive research on writing than on other language skills, such as speaking (Li et al., 2025). As GenAI technologies have evolved and incorporated complementary technologies such as automatic speech recognition, text-to-speech synthesis, and facial animation, they have increasingly supported voice-based interaction, facial expression, and affective responses (Wang et al., 2024). These developments have extended the potential of GenAI beyond text-based language learning and created new opportunities for EFL speaking development. Consequently, research attention has increasingly focused on GenAI-mediated EFL speaking activities.

Despite growing interest in GenAI applications, empirical evidence concerning their effectiveness in EFL speaking development remains mixed. Although several studies have reported improvements in speaking performance, fluency, and speaking anxiety following AI-supported practice (Ebadi et al., 2025; Zhou et al., 2025), others have found no significant advantage over conventional or peer-based speaking activities in overall speaking performance (Chen et al., 2025; Wang et al., 2024). Moreover, positive linguistic outcomes do not necessarily coincide with equally favorable learning experiences. Some learners have expressed concerns about excessive reliance on AI and reduced autonomy, while others have reported declining engagement and enjoyment during AI-mediated speaking activities (Ebadi et al., 2025; Lee, 2025). These mixed findings suggest that, although GenAI offers opportunities to address persistent challenges in EFL speaking instruction, its educational value may depend on how the technology is integrated into pedagogical activities, the outcomes assessed, and learners' experiences of AI-mediated interaction.

As Zhai et al. (2021) noted, computers were once regarded as potential game changers in education, yet their introduction did not necessarily result in substantial changes in core educational practices. A similar issue may arise with GenAI, as its educational value may depend substantially on how it is integrated into pedagogical activities and learning contexts (Guo et al., 2026). In particular, the design and implementation of GenAI-mediated speaking activities may be critical, as similar GenAI programs can produce various outcomes when embedded in different pedagogical interventions (Liu et al., 2025). This suggests that understanding GenAI-mediated EFL speaking requires attention not only to its technological affordances but also to the pedagogical

activities through which these affordances are enacted. Therefore, a systematic review is needed to synthesize how GenAI has been integrated into EFL speaking activities, and to examine the challenges and tensions that emerge from such integration. To provide a holistic perspective, Activity Theory offers a useful theoretical lens for examining how GenAI-mediated speaking activities are shaped by the intervention of multiple elements.

Building on Vygotsky's (1978) early conceptualization of activity, Engeström (2014) expanded AT by incorporating rules, community, and division of labour alongside the subject, object, and mediating tools. This expanded model conceptualizes learning as a dynamic, socially situated process shaped by interactions among the elements of an activity system within a socio-historical context (Koszalka & Wu, 2004). It is particularly relevant to the study of AI-human interaction because it enables researchers to examine how learners, mediating technologies, pedagogical practices, and social contexts interact to shape learning processes and outcomes (Fathi et al., 2024).

In this review, AT was adopted as the analytical framework for examining the activity systems represented in the eligible studies. The key elements of the activity system are operationalized as follows. The subject refers to the individual or group engaged in activity; the object refers to the goal, motive, or problem toward which the activity is directed; the tools refer to psychological or material artefacts that mediate the relationship between the subject and the object; rules refer to the explicit or implicit norms, expectations, and conventions that guide participants' behavior and regulate interaction within the activity; Community comprises the individuals or groups who share or are involved in the same object of activity; While division of labour refers to the distribution of roles, responsibilities, and tasks among participants within the activity system (Zheng et al., 2020). Together, these interconnected elements provide an analytical framework for examining how learners, GenAI technologies, speaking-related practices, and social interactions shape the learning process and outcomes.

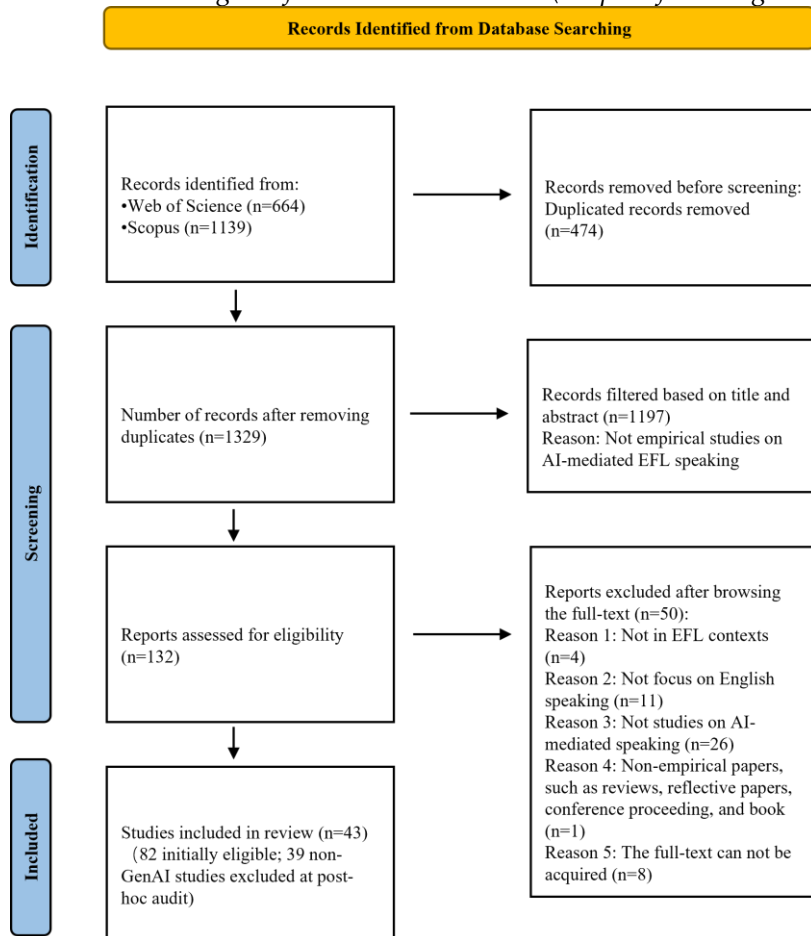
3. Methodology

This study conducted a systematic literature review of empirical studies on GenAI-mediated EFL speaking. To ensure methodological transparency, consistency, and reproducibility, the review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) guidelines (Page et al., 2021). The review process comprised four interconnected stages: (1) search strategy and literature retrieval, including database selection and the development and application of database-specific search strings; (2) eligibility assessment based on predefined inclusion and exclusion criteria; (3) methodological quality appraisal of the eligible studies using the Mixed Methods Appraisal Tool [MMAT] (Hong et al., 2018); and (4) systematic coding, synthesis, and analysis of the included studies. These procedures were intended to provide a transparent and reproducible basis for identifying eligible empirical evidence and synthesizing patterns in trends, application, and challenges in research on GenAI-mediated EFL speaking activities. Figure 1 presents the overall literature retrieval and screening process in the PRISMA flow diagram.

3.1. Search Strategy and Research Retrieval

A systematic literature search was conducted in two major academic databases, Web of Science [WOS] and Scopus, to identify empirical studies on AI-mediated EFL speaking activities. The selection of these databases was informed by Gusenbauer and Haddaway's (2020) evaluation of academic search databases. They evaluated 28 databases based on criteria including search efficiency, precision, recall, coverage, and reproducibility. Among the databases considered suitable for systematic literature searching, WOS and Scopus were selected for this review because of their broad multidisciplinary coverage and suitability for systematic literature searching. In addition, these two databases also provided a clearly defined and reproducible search scope, while keeping the retrieval and screening processes feasible and manageable.

Figure 1
The PRISMA diagram for literature retrieval (adapted from Page et al., 2021)



The search strategy comprised three conceptual domains: (a) AI technologies (“artificial intelligence”, “AI”, “generative artificial intelligence”, “GenAI”, “GAI”, “large language model”, “conversational agent”, or “chatbot”); (b) speaking-focused activities (“speaking”, “oral communication”, “speaking proficiency”, “speaking skill”, or “interaction”); (c) EFL contexts (“English as a foreign language learning”, “EFL”, “foreign language”, “English”, or “L2”).

Table 1 presents the complete search strings used in WOS and Scopus, including database-specific adaptations. Because the existing literature does not consistently distinguish between AI and GenAI, a broad AI-focused search strategy was adopted at the initial retrieval stage to reduce the risk of excluding potentially relevant studies. Thus, the initial search incorporated both general AI-related and GenAI-specific terms. After title and abstract screening and full-text eligibility, a post-hoc eligibility re-audit was conducted to identify studies that explicitly involved GenAI in EFL speaking activities.

In addition, three eligibility restrictions were applied during database retrieval in WOS and Scopus. First, the search was limited to studies published between 2021 and 2026, with the final search conducted on September 15, 2026. Second, only studies published in English were included. Third, the review included only peer-reviewed journal articles.

The initial searches retrieved 1,803 records from WOS ($n = 664$) and Scopus ($n = 1139$) (see Table 1). After the records were exported and consolidated, 474 duplicate records were identified and removed using the filtering function in Microsoft Excel. The deduplication results were also manually checked by two reviewers to verify the removal of duplicate records. Following deduplication, 1,329 records remained for the subsequent screening process.

3.2. Inclusion and Exclusion Criteria

After removing duplicate records, the remaining 1,329 records underwent title and abstract screening against the predefined inclusion and exclusion criteria. To provide a consistent basis for determining whether a study qualified as an empirical study on AI-mediated EFL speaking activities, we established an operational definition before screening. First, an empirical study was defined as a study that collected and analyzed original data or evidence from participants, learning activities, or observed interactions. Second, AI-mediated EFL speaking activities were defined as learning activities in which EFL learners use AI technologies as an integral part of practicing, performing, interacting, or receiving feedback on English speaking. In these activities, AI technologies had to play an active or mediating role in the interaction process, rather than being used for other language skills, administrative purposes, or as the object of technological evaluation. Thus, a study was considered relevant when the title and abstract met all the following criteria: (1) Participants (EFL learners); (2) Target language (English); (3) Speaking focus (rather than other language skills such as reading, writing, listening); (4) AI-mediated activities (AI as an integral part of the interaction process for providing practice, feedback, assessment, or speaking support); (5) empirical study (reports original empirical data or evidence). Overall, this offered a consistent basis for distinguishing empirical studies on AI-mediated EFL speaking from broader AI-related research. Table 2 outlines the inclusion and exclusion criteria for the following title and abstract screening and full-text eligibility.

At the title and abstract screening stage, 1,197 records were excluded because their titles or abstracts did not meet the criteria in Table 2. The screening was independently conducted by two reviewers. Any disagreements were discussed and resolved through consensus. When consensus could not be reached, the relevant sections were reexamined, and a third reviewer was consulted when necessary to make the final decision.

Subsequently, the remaining 132 records were independently assessed in full text by the two reviewers according to the criteria in Table 2. Of these studies, 50 studies were excluded at this stage because they were not in EFL contexts ($n = 4$), did not focus on English speaking ($n = 11$), were not studies on AI-mediated speaking activities ($n = 26$), or were not empirical articles ($n = 1$). Studies for which the full text could not be obtained despite reasonable retrieval attempts were also excluded ($n = 8$). After full-text eligibility, 82 primary empirical studies were included.

In addition, because the initial search strategy adopted a broad AI-related scope to minimize the risk of overlooking relevant studies, a post-hoc audit was conducted by two reviewers to identify studies that specifically involved GenAI in EFL speaking activities. GenAI applications in the speaking studies mainly included GenAI programs or chatbots (e.g., ChatGPT) and GenAI-enabled educational applications or agents (e.g., Call Annie) (Liu et al., 2026). Studies involving conventional rule-based AI systems that provided only preset responses or automated pronunciation assessment without generative capabilities were excluded. Following this post-hoc audit, 39 studies were excluded. Finally, 43 empirical studies on GenAI-mediated EFL speaking were retained for the final synthesis.

3.3. Quality Assessment

After the whole eligibility process, the methodological quality of the 43 included studies was appraised to contextualize the interpretation of the synthesized evidence. Because the included studies comprised qualitative, quantitative, and mixed-methods designs, the Mixed Methods Appraisal Tool (MMAT) (Hong et al., 2018) was employed to assess methodological quality. Each study was evaluated against the MMAT criteria for its research design. For each criterion, a judgement of "Yes", "No", or "Cannot Tell" was assigned, accompanied by a study-specific justification. The judgements were not combined into a single quality score, as the MMAT is intended to identify methodological strengths and limitations rather than to produce a numerical ranking of study quality.

Table 1
Database-specific search strings and initial search yields

Database	Search String	Filters Applied	Yield
WOS (Social Sciences Citation Index (SSCI), Arts & Humanities Citation Index (A&HCI))	TS = (("artificial intelligence" OR "AI" OR "generative artificial intelligence" OR "GenAI" OR "GAI" OR "large language model" OR "conversational agent" OR "chatbot") AND ("speaking" OR "oral communication" OR "speaking proficiency" OR "speaking skill" OR "interaction") AND ("English as a foreign language learning" OR "EFL" OR "foreign language" OR "English" OR "L2"))	Search within Abstract; Article; English; 2021 - 2026; Education/Linguistics/Social Science/Psychology	664
Scopus	TITLE-ABS-KEY (("artificial intelligence" OR "AI" OR "generative artificial intelligence" OR "GenAI" OR "GAI" OR "large language model" OR "conversational agent" OR "chatbot") AND ("speaking" OR "oral communication" OR "speaking proficiency" OR "speaking skill" OR "interaction") AND ("English as a foreign language learning" OR "EFL" OR "foreign language" OR "English" OR "L2"))	Search within Title, Abstract, and Keywords; Article; English; 2021 - 2026; Social Science/Psychology	1139

Table 2

Inclusion and Exclusion Criteria

<i>Inclusion criteria</i>	<i>Exclusion criteria</i>
Published 2021 - 2026	Published before 2021
Published in English	Not in English
Peer-reviewed journal article	Non-peer-reviewed publications, such as dissertations or theses
EFL contexts	ESL contexts
English as the target language	Other language as the target language
Pure speaking learning	Other language skills, such as writing, reading, listening
AI-mediated activity (AI as an integral part in interaction process for providing practice, feedback, assessment, or speaking support)	Studies focus on technological evaluation, quality evaluation of technology assessment, or cross-sectional study on relationship of related variables
Empirical studies reported original evidence or data	Non-empirical studies such as literature reviews, conceptual papers, or theoretical discussions

In the MMAT assessment, two reviewers independently appraised all included studies to enhance consistency and reduce subjective judgement. Their assessments were then compared, and disagreements were resolved through discussion and consensus. When methodological information was insufficient or unclear, the full text was reexamined. The quality appraisal informed the synthesis and interpretation of the findings, particularly the strength and consistency of evidence across study designs.

3.4. Coding and Analysis

The coding framework (Table 3) was developed in alignment with the three research questions. The coding categories and procedures used to address each research question are described below.

1. RQ1 (Publication and research trends): The number of studies was recorded to identify publication trends by year and geographical location, as well as research trends, including study design, intervention duration, and group structure.

2. RQ2 (GenAI-mediated EFL speaking activity): GenAI-mediated EFL speaking activities in the included studies were coded using the seven key elements of Engeström's (2014) AT as a guiding analytical framework: subject, tool, object, outcome, community, rules, and division of labour. Specific sub-codes for each element were developed inductively from the included studies, with reference to relevant literature reviews where appropriate. Each sub-code was assigned an alphabetical label (e.g., a, b, c, d...) to facilitate systematic coding and analysis.

3. RQ3 (Existing challenges): Challenges associated with GenAI-mediated EFL speaking activities were coded from two perspectives: (A) subjective and (B) objective aspects at three stages, including (1) preparation, (2) interaction, and (3) feedback. For instance, "A1" refers to a subjective challenge at the preparation stage of a GenAI-mediated EFL speaking activity. Specific challenges identified within each category were assigned alphabetical codes (e.g., a, b, c, d...) to facilitate systematic coding and analysis.

Thus, guided by the AT framework, the coding process was organized around its seven key elements, while the specific coding categories and definitions were informed by relevant literature reviews. Two coders independently extracted relevant information from the included studies, organized the reported evidence into the analytical categories, and developed sub-codes. To enhance coding reliability, their coding results were independently reviewed, and discrepancies were discussed and resolved through consensus (Nguyen et al., 2026). When clarification was needed, the original studies were reexamined to verify relevant information and confirm coding decisions, thereby maintaining consistency throughout the coding and analysis process.

4. Findings

The findings are organized around the three research questions and their corresponding coding results. They are presented in three sections: (1) Publication and research trends, (2) GenAI-mediated EFL speaking activities, and (3) emerging challenges. The findings are interpreted based on the coding results (see Supplementary material) from the included studies.

4.1. Publication and Research Trends

First, publication trends were examined by publication year and geographical distribution. Although the search covered publications from 2021 to 2026, all 43 eligible studies were published between 2024 and 2026. The number of publications increased from 5 in 2024 to 20 in 2025, with 18 studies by September 15, 2026. This pattern indicates rapidly growing research attention to GenAI application in EFL speaking activities from 2024 to 2025. The 2026 figure, although incomplete at the time of the research, suggests that the yields remain at a high level. Geographically, the 43 studies were conducted across 14 countries/regions (Table 4). Mainland China accounted for the largest share (18 studies), followed by Taiwan (5 studies). The concentration of studies in mainland China indicates that the empirical evidence is geographically uneven, which should be considered when interpreting the transferability of findings to other EFL contexts.

Table 3
Coding framework for the included studies

RQs	Themes	Sub-themes	Details	References
1	Publication trends	Year	The time of studies published on a yearly basis	/
	Research trends	Geographic location	The frequency of countries or regions of the studies conducted	(Chen et al., 2021)
		Study design	The methods of research design are clarified in the MMAT list (1, 2, 3, 4, 5)	(Hong et al., 2018)
		Intervention duration	The time on a weekly basis predefined for an intervention	/
		Group structure	The number of group divisions in a study	/
2	Subject	Education level	The education stage of the participants, including (a) primary education, (b) secondary education, (c) undergraduate, (d) postgraduate	(Lai & Lee, 2024)
	Object Tool	English proficiency	The English level of participants reported in studies, including (a) elementary, (b) intermediate, (c) high	/
		Name	The aims and motive of studies	/
			The name of the GenAI tool applied in the studies	/
			Two main types: (a) GenAI programs (chatbots), and (b) GenAI-enabled language learning apps (agents)	(Liu et al., 2026)
	Rule	Role	The roles of GenAI in speaking activity, including (a) teacher/tutor, (b) partner/peer, (c) learning tool/speaking assistant	(Hwang & Chen, 2023)
		Community	The norms, conventions, expectations, or regulations followed by learners to participate in a GenAI-integrated EFL speaking activity (a, b, c, d...)	(Engeström, 2014)
			The participants in the speaking activity, including the learner, GenAI, peer, and teacher/researcher	/
			The place of GenAI interaction, including (a) in class/program, (b) out of class, or (c) blended	/
	Division of labour	GenAI Teacher/Researcher Peer	The responsibilities and contributions of all participants for learners in speaking activity (a, b, c, d, e, f, g...)	/
3	Outcome	Cognitive	Related to knowledge comprehension and speaking-related ability in the studies (e.g., speaking skills)	(Krathwohl, 2002)
	Challenges	Affective	Related to emotional influences (e.g., willingness to communicate)	(Zhai & Wibowo, 2023)
		Behavioral	Related to actual actions (e.g., engagement, speaking frequency)	(Snowman et al., 2011)
			The limitations of GenAI application in speaking activity caused by (A) subjective and (B) objective factors in three stages of speaking activity, including (1) preparation, (2) interaction, and (3) feedback (e.g., "A1a" means the first kind of subjective challenge in the preparation stage)	(Zhou et al., 2025)

Table 4
Geographical distribution of the included studies

Country/Region	2024	2025	2026	Total
China	2	9	7	18
Taiwan	1	2	2	5
Indonesia	0	2	2	4
Saudi Arabia	1	1	1	3
South Korea	0	2	0	2
Oman	0	1	1	2
Peru	0	0	2	2
Ethiopia	1	0	0	1
Georgia	0	0	1	1
Iran	0	0	1	1
Iraq	0	1	0	1
Morocco	0	1	0	1
Turkey	0	1	0	1
Vietnam	0	0	1	1
Total	5	20	18	43

Second, research trends are examined in terms of three characteristics: study design, intervention duration, and group structure. Regarding study design, the 43 included studies comprised 1 qualitative, 1 randomized, 13 non-randomized, 2 quantitative descriptive, and 26 mixed-methods studies, corresponding to the five MMAT design categories. For intervention duration, 8-week interventions were most common (8 studies), followed by 6-week interventions (6 studies). Regarding group structure, two-group designs were the most frequently reported (27 studies). These two-group comparisons involved different instructional conditions, learner proficiency levels, or AI tools, with GenAI-mediated instruction versus traditional instruction being the most common comparison.

4.2. GenAI-mediated EFL Speaking Activity

4.2.1. Subject

The subject refers to the EFL learners participating in GenAI-mediated speaking activities. Their characteristics were analyzed regarding education level and English proficiency. As shown in Table 5, the reviewed studies predominantly focused on undergraduate learners, with 35 of the 43 studies involving this population. In comparison, relatively few studies involved primary ($n = 1$), secondary ($n = 4$), or postgraduate learners ($n = 2$). Thus, the participant profile was predominantly undergraduate.

Table 5
Education level of the participants in the reviewed studies

Education level	Number
Primary education	1
Secondary education	4
Undergraduate	35
Postgraduate	2
Unspecified	1

In terms of English proficiency (Table 6), 18 studies focused exclusively on intermediate-level learners, while 4 focused exclusively on elementary-level learners. 10 studies included learners across multiple proficiency levels, whereas 11 did not specify participants' proficiency levels. Thus, 22 studies focused on a single proficiency level, while others included learners across multiple levels. Overall, intermediate-level learners appeared most frequently ($n = 27$), followed by elementary-level learners ($n = 13$), whereas high-proficiency learners appeared in only 4 studies and received relatively limited attention.

Table 6
English proficiency of the participants in the reviewed studies

English proficiency	Number
a. elementary	4
b. intermediate	18
a. elementary, b. intermediate	6
a. elementary, b. intermediate, c. high	2
a. elementary, c. high	1
b. intermediate, c. high	1
unspecified	11
Sum of a, b, and c	a=13, b=27, c=4

4.2.2. Tool

The mediating tool in GenAI-mediated EFL speaking activity is GenAI, which can be conceptualized at two levels. First, GenAI refers to technologies embedded in computer systems; Second, it refers to software applications that incorporate such technologies. Combined with the information of the included studies, GenAI in EFL speaking activities can be broadly grouped into two types: (a) GenAI programs, primarily represented by GenAI chatbots, and (b) GenAI-enabled language learning applications, which incorporate GenAI technologies into language learning programs (Liu et al., 2026). The former typically provides learners with direct and flexible interaction, whereas the latter incorporates generative capabilities and human-like features into language learning systems with functions and tasks specifically designed for speaking practice and assessment.

This review analyzed the characteristics of GenAI tools used in speaking activities by tool name, type, and pedagogical role. In terms of tool names, ChatGPT was the most frequently reported (17 studies). Regarding tool type, GenAI programs, primarily represented by chatbots, were the most commonly used (36 studies), while the remaining studies used GenAI-enabled language learning applications. The included studies identified three main pedagogical roles of GenAI in speaking activities: (a) GenAI served as a teacher or tutor (also as the interviewer, trainer, examiner, or assessor) to provide learners with personalized assessment and immediate feedback; (b) GenAI functioned as a conversational partner or peer to provide personalized practice opportunities or supporting learners in completing speaking-related tasks, such as role playing; (c) GenAI served as a learning tool to provided scaffolding and assistant in a relatively low-risk learning environment, as well as supporting in knowledge resources and developing ideas for speaking activities. In terms of frequency, the conversational partner role was most common ($n = 39$), followed by the teacher or tutor role ($n = 31$) and the learning tool role ($n = 30$). 34 studies reported GenAI serving multiple roles within the same speaking activity. Of these, 20 studies used GenAI in all three roles: teacher, conversational partner, and speaking learning tool (see Table 7).

Table 7
GenAI roles in speaking activity

GenAI roles	Number
a	3
b	5
c	1
a, b	5
a, c	3
b, c	6
a, b, c	20
Sum of a, b, and c	a=31, b=39, c=30

4.2.3 Objects

Across the 43 included studies, the objects of GenAI-integrated EFL speaking activities can be grouped into four aspects: (a) Developing oral proficiency, (b) Addressing affective and

motivational barriers, (c) Understanding learners' experience and perceptions, and (d) Improving the pedagogical integration of GenAI into speaking activities. As shown in Table 8, the most frequently reported object was developing learners' oral proficiency (12 studies). These studies addressed difficulties in overall speaking performance as well as specific dimensions such as accuracy, fluency, pronunciation, complexity, and coherence. This indicates that improving learners' ability to start and sustain English conversations was the most prominent goal of GenAI-mediated EFL speaking activity.

In addition, 4 studies aimed to address affective barriers that may constrain participation in L2 speaking. These studies focused on problems such as speaking anxiety, limited willingness to communicate [WTC], low motivation, insufficient self-efficacy, and limited confidence. Rather than treating speaking development as a solely linguistic problem, these studies positioned GenAI as a tool that may create conditions to facilitate learners' willingness and readiness to participate in oral communication. Moreover, 3 studies focused on learners' experiences and perceptions of GenAI-integrated speaking activities. These examined how learners experienced and participated in these activities. Furthermore, 4 studies focused on the pedagogical integration of GenAI through role-playing, peer interaction, blended learning, and established instructional frameworks such as Think-Pair-Share. These studies focused on how GenAI could be incorporated into pedagogical arrangements to support speaking activities.

However, the four dimensions of objects were not mutually exclusive, with some studies addressing multiple objects simultaneously. This overlap was observed in 18 studies. Among these, oral development was frequently examined alongside affective concerns (9 studies). Overall, oral development was the most frequently identified object ($n = 29$) across the included studies. Meanwhile, affective factors also received considerable attention ($n = 13$). Another emerging object concerned the pedagogical integration of GenAI into EFL speaking instruction ($n = 12$). By contrast, learners' experiences and perceptions of GenAI were identified less frequently as an object ($n = 9$).

Table 8

Objects of GenAI application in EFL speaking activity

<i>Objects</i>	<i>Number</i>
a	12
b	4
c	3
d	4
a, b	9
a, c	3
a, d	3
c, d	1
a, c, d	2
unspecified	2
Sum of a, b, c, and d	a=29, b=13, c=9, and d=12

4.2.4 Outcomes

The outcomes identified across the included studies covered cognitive, affective, and behavioral dimensions. Cognitive outcomes mainly concerned learners' oral language development. Across 43 studies, these outcomes were further classified into overall speaking skills, fluency, vocabulary, grammar, pronunciation, colloquial language, correctness, complexity, discourse management, gesture, volume, relatedness, and adaptability. Overall, the findings were mixed but predominantly positive. 27 studies reported significant improvements in one or multiple aspects of speaking performance following GenAI-integrated speaking practice. However, 12 studies reported limited or non-significant improvements in cognitive outcomes, with pronunciation being the most frequently reported area of limited improvement (4 studies). 4 studies did not report related cognitive outcomes. Overall, these findings indicate that reported outcomes varied across

different dimensions of speaking ability, rather than showing consistent improvement across all cognitive outcomes.

In addition, the included studies examined a range of affective outcomes related to GenAI-mediated speaking activities, including interest, anxiety, confidence, enjoyment, motivation, self-efficacy, WTC, grit, satisfaction, linguistic risk-taking, academic buoyancy, and hesitation. Among these outcomes, speaking anxiety was the most frequently examined, appearing in 20 of the 43 studies. Overall, 24 studies reported positive affective outcomes, whereas 5 reported no significant effects. Of these 5 studies, 4 studies found that GenAI-integrated speaking practice did not significantly reduce learners' speaking anxiety, while 1 study showed lower enjoyment with GenAI than with peer interaction. The remaining 14 studies did not report affective outcomes.

Moreover, 15 of the 43 studies reported behavior outcomes related to learners' participation and behavioral engagement in GenAI-integrated speaking activities. These outcomes included engagement, expression performance, social and ego resilience, task response frequency, meaningful language learning, learner autonomy, social dynamics, shared learning experience, and mutual resolution of language problems. Engagement was the most frequently examined outcome (9 studies). Overall, the findings on behavioral outcomes were predominantly positive. However, 3 studies reported non-significant or less favorable effects. Of these 3 studies, 1 study found no significant difference in emotional engagement, 1 study reported lower engagement and meaningful language learning compared with peer interaction, and another found no significant effect on ego resilience.

4.2.5. *Community*

The community dimension was examined in terms of participating parties and intervention settings. Regarding participating parties, learners, GenAI, peers, teachers, and researchers were the primary participants involved in GenAI-integrated speaking activities. Across the 43 included studies, four main participant configurations were identified: (a) learner-GenAI (27 studies), (b) learner-GenAI-researcher/teacher (7 studies), (c) learner-GenAI-peer (2 studies), and (d) learner-GenAI-peer-teacher (5 studies). The remaining 2 studies adopted a comparative configuration of participants. One compared learner-GenAI with learner-GenAI-teacher/researcher to examine the role of teacher intervention or feedback in GenAI-mediated speaking activities. The other compared learner-GenAI with learner-GenAI-peer to examine the role of peer involvement by contrasting individual GenAI interaction and GenAI-integrated peer interaction. Overall, learner-GenAI interaction was the dominant participant configuration, whereas direct involvement of peers and teachers in GenAI-integrated speaking activities was less common.

Regarding intervention settings, 14 studies implemented GenAI-integrated speaking activities in class, 12 conducted them after class, and 10 adopted blended settings that combine in-class and after-class activities. The remaining 7 studies did not specify the context of GenAI-mediated EFL speaking activities. In-class activities were mainly integrated into teachers' instructional practices, whereas after-class activities were mainly used as supplementary speaking practice beyond formal instruction. Blended settings typically introduced GenAI and its use during class and extended speaking practice to autonomous after-class activities. Overall, the included studies used a range of intervention settings, with in-class implementation most frequently reported, followed by after-class and blended settings.

4.2.6. *Rules*

Among the 43 studies, 38 specified one or more rules governing GenAI-integrated speaking activities. Eight primary types of rules were identified: (a) Providing well-designed prompts or training prompts; (b) Following structured interaction steps based on pedagogical frameworks; (c) Using predefined topics aligned with the coursebooks; (d) English-only interaction; (e) Predefined time allocations and setting; (f) Specified interaction modes, such as speaking or typing; (g) Device requirements; and (h) Interaction principle, such as meaning-focused or form-focused practice. Of the 38 studies, structured interaction steps based on pedagogical frameworks were the most

frequently reported (16 studies), followed by the rule of prompts (13 studies), predefined topics for speaking practice (11 studies), and predefined time allocations and settings (8 studies). However, only 1 study reported English-only interaction, while 3 studies explicitly encouraged free conversation.

4.2.7. Division of labour

In GenAI-mediated speaking activities, the division of labour across the included studies showed a differentiated allocation of tasks among GenAI, teachers or researchers, and peers. Each participant may play multiple roles in these speaking activities. For example, GenAI was primarily involved in tasks related to providing speaking practice opportunities (35 studies), generating feedback (29 studies), supporting idea or topic development (11 studies), and providing platforms for speaking activities (8 studies). Teachers or researchers were more frequently responsible for guiding learners (15 studies), organizing activities (14 studies), designing activities (11 studies), supervising the learning process (11 studies), and developing or training prompts (9 studies). Notably, although GenAI can provide instant feedback, 5 studies reported that teachers' feedback remained present alongside GenAI feedback.

In comparison, peer involvement was relatively limited, appearing in 9 studies. Peers were primarily assigned tasks related to collaboration in completing group tasks (5 studies) and participating in speaking practice with GenAI support (2 studies). Overall, the division of labour across the included studies showed a differentiated allocation of tasks among participants.

4.3. Emerging Challenges

Across the 43 included studies, challenges associated with GenAI-mediated EFL speaking activities were identified from both subjective and objective perspectives. These challenges were mainly reported across three stages of the speaking activities: preparation, interaction, and feedback. Notably, one study often reported multiple challenges (see Table 9).

Table 9

Challenges of GenAI application in EFL speaking activity

1. Preparation	2. Interaction	3. Feedback
A. Subjective		
a. insufficient AI literacy for designing appropriate prompts (3 studies)	a. insufficient English proficiency in pronunciation to be recognized by machines (1 study)	a. over-reliance and complete trust in AI feedback (1 study)
b. unequal labour division of pair tasks (1 study)	b. insufficient AI literacy for multi-turn interaction (3 studies)	/
B. Objective		
a. geographical constraints (3 studies)	a. unnatural and inauthentic (21 studies)	a. topic deviation and irrelevant responses (4 studies)
b. subscription requirements (1 study)	b. limitations in speech recognition and meaning comprehension (6 studies)	b. response delays (3 studies)
/	c. device and internet errors (1 study)	c. undetailed feedback (4 studies)
/	/	d. limitations on speech, pronunciation and fluency feedback (6 studies)
/	/	e. inaccurate feedback (6 studies)

Subjective challenges were relatively limited across the 43 studies. At the preparation stage, two kinds of challenges were reported, including insufficient AI literacy for designing appropriate

prompts (3 studies) and unequal opportunities for oral practice in inadequate labour division of pair tasks (1 study). During interaction, subjective barriers involved insufficient pronunciation for machine recognition (1 study) and insufficient AI literacy to maintain multi-turn dialogue (3 studies). 1 study reported subjective limitations related to over-reliance and complete trust in AI feedback at the feedback stage.

Objective challenges were comparatively prevalent across all stages. At the preparation stage, two types of limitations were identified: restricted access to GenAI due to geographical constraints (3 studies) and subscription requirements (1 study). During interaction, the reported challenges included unnatural and inauthentic interaction experience (21 studies), limitations in speech recognition and meaning comprehension (6 studies), and device- or internet-related errors (1 study). At the feedback stage, five types of challenges were identified: topic deviation or irrelevant responses (4 studies), response delays (3 studies), undetailed feedback (4 studies), limitations in providing feedback on speech, pronunciation, and fluency (6 studies), and inaccurate feedback (6 studies).

Overall, the feedback stage showed the greatest diversity of objective challenges, with five categories reported across 23 studies. Across all stages, the most frequently reported challenge was the unnatural and inauthentic interaction experience (21 studies), mainly attributed to monotonous intonation, formulaic response patterns, artificial appearance, and lengthy responses containing advanced vocabulary.

5. Discussion

The discussion interprets the review findings in relation to the three research questions. It firstly discusses the publication and research trends identified in the included studies, then examines how GenAI-mediated speaking activities are structured through the seven elements of an activity system, followed by the emerging challenges and tensions. This part pays particular attention to the relationships among these elements rather than treating them as independent categories. From an Activity Theory perspective, these findings are combined together to identify tensions and develop an integrated conceptual framework for future research on GenAI-mediated EFL speaking learning.

5.1. Publication and Research Trends

The findings indicate that empirical research on GenAI-mediated EFL speaking has expanded rapidly from 2024 to 2025. Although the search covered publications from 2021 to 2026, all included studies were published between 2024 and 2026. This temporal pattern is consistent with the broader emergence of GenAI in language education (Liu et al., 2026). In addition, the included studies spanned 14 countries or regions, but a considerable portion were concentrated in mainland China. This distribution indicates that the current evidence base is geographically concentrated, highlighting the need for further research across a wider range of EFL contexts.

Regarding research trends, mixed-methods designs and two-group comparisons were frequently used, while intervention durations varied, with eight weeks being the most commonly reported period. This indicates that existing studies have primarily examined GenAI effects by comparing GenAI-mediated speaking practice and traditional instruction. Given this, future research could examine GenAI-mediated speaking activities across different EFL contexts, compare distinct pedagogical methods integrated by GenAI, or use longer intervention periods to assess whether observed effects are sustained or changed over time.

5.2. GenAI-mediated EFL Speaking Activity

The findings suggest that the activity system in the reviewed studies consists primarily of EFL learners as subjects engaging in GenAI-mediated speaking activities. Undergraduate and intermediate-level learners were particularly prominent. This concentration suggests that undergraduate students are the primary population, given the accessibility of mobile devices, self-directed abilities, and technology literacy, which is consistent with the review of Zhou et al. (2023).

In addition, the concentration of intermediate-level learners may reflect the view that English proficiency is a prerequisite for engaging in GenAI-mediated EFL speaking activities, which aligns with the literature review of Zhai and Wibowo (2023). However, in the reviewed studies, Wang et al. (2024) reported that even intermediate-level students prefer typing rather than speaking with GenAI chatbots. While typing diminishes the authentic sense of social interaction (Zhang et al., 2025), it may also undermine its effects on speaking performance (Kim et al., 2021). At the same time, the primary objects of GenAI-intermediate EFL speaking activities were reported to improve learners' speaking performance. Thus, these findings reveal the tension between the subject and the object in existing GenAI-mediated EFL speaking activities. Specifically, although GenAI provides conversational opportunities, learners may be unable or unwilling to initiate a meaningful oral interaction with GenAI independently. This has also been reflected in the emerging challenges of GenAI-mediated EFL speaking activities.

In addition, ChatGPT was the most frequently examined tool. Therefore, the empirical evidence was concentrated on ChatGPT as a kind of GenAI chatbots. However, in their review (2013-2023) of conversational AI tools, Lai and Lee (2024) found that Google Assistant was the most widely used tool for English conversation activities. This may reflect a shift in tools used for EFL speaking activities. In terms of its roles, GenAI served as a feedback provider, a conversational partner, and linguistic support. Because GenAI plays multiple roles traditionally performed by teachers and classmates, it may contribute to the redistribution of pedagogical tasks among GenAI, teachers, and peers. From an Activity Theory perspective, this redistribution matters because changes in the division of labour can alter how other components of the activity system operate. In particular, when GenAI assumes interaction or feedback functions, teachers may spend more time supporting GenAI-mediated speaking activities rather than directly participating in interaction activities.

However, the rules indicate that although teachers were not directly involved in most learner-GenAI conversations, they played an important role in shaping and regulating the overall activity process. In the reviewed studies, GenAI-integrated speaking activities were generally structured rather than left to unrestricted learner-GenAI interaction. This suggests that GenAI functions were increasingly embedded within existing instructional structures rather than operating as a standalone conversational tool. Such structured rules may provide necessary scaffolding for EFL learners, particularly those with insufficient linguistic proficiency and AI literacy, but they may also constrain spontaneous interaction and learner autonomy. Thus, this reconfiguration creates new requirements for teachers in facilitating the outcomes of GenAI-mediated speaking activities.

Moreover, although the community includes teachers, peers, learners, and the broader educational environment, the included studies emphasized learner-GenAI interaction in the activities. GenAI can provide personalized conversational practice and instant feedback in a comparatively low-pressure environment with less direct dependence on human interlocutors in individual interaction. However, the limited involvement of human interaction revealed a gap between outcomes from GenAI-mediated practice and actual performance in authentic communication. This also highlighted the tension between tool and object, which further reflects the contradictions between inauthentic interaction with GenAI and requirements in real communication. Notably, despite the less involvement of human interlocutors in learner-GenAI interaction, the included studies were still largely situated in classroom contexts, indicating continued dependence on contextual support with teachers and peers.

5.3. Tensions as Drive Force of the Collaborative Framework

These elements within the activity system further give rise to tensions that may become key sources of possible transformation. The tensions are mainly from tools and subjects, as well as tools and objects in the included studies.

The first tension concerns the tool's affordances and the subject's requirements. As reported in the included studies, GenAI can provide learners with personalized practice and instant feedback in a low-pressure context. Such conditions may be particularly valuable for learners to engage in speaking, which are the most anxiety-inducing language skill (Ebadi et al., 2025). Some studies

have similarly identified potential benefits of GenAI in speaking-related outcomes. However, the reviewed studies also raised concerns about an unnatural, inauthentic experience caused by formulaic responses, monotonous or unnatural intonation, artificial appearances, and limited pragmatic sensitivity (Wu et al., 2026; Zhang, 2026). These limitations suggest that learner-GenAI can provide individual practice opportunities without reproducing authentic social interaction.

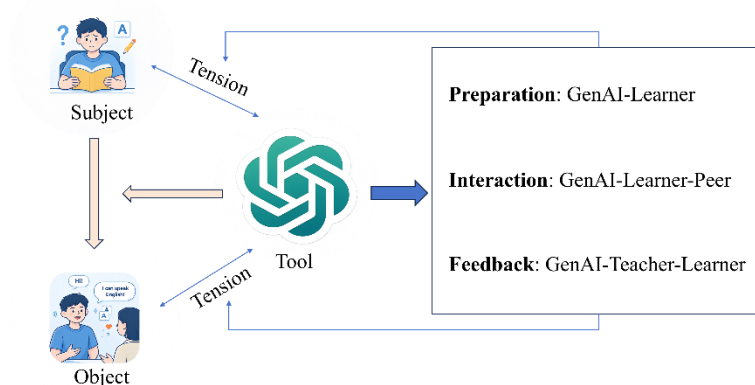
Another tension also concerns the relationship between the tool's affordances and the subjects' abilities. GenAI enables learners to practice oral English in a comparatively low-pressure environment (Cui & Zhang, 2026). However, the reviewed studies also identified insufficient English proficiency, limited AI literacy, and over-reliance as challenges for engaging in these activities. Thus, technological advancements put forward new requirements for learners' abilities. Effective autonomous use of GenAI requires learners to formulate meaningful prompts, initiate and sustain interaction, evaluate GenAI responses, recognize inappropriate feedback, and regulate their learning. Thus, access to GenAI does not automatically result in effective autonomous learning.

The last tension concerns the relationship between the tool's affordance in feedback and the object for speaking development. GenAI can provide instant responses and feedback for improving speaking performance. However, challenges in the reviewed studies reported a relatively diverse range of feedback limitations. It indicates that although GenAI can provide instant feedback, the quality of GenAI feedback is still a question (Lee, 2025). In addition, deficiencies in evaluating pronunciation and fluency suggest that GenAI-generated feedback may be limited.

Based on these tensions, a collaborative learning framework (Figure 2) for GenAI-mediated EFL speaking was developed. The framework does not propose GenAI as a replacement for teachers or peers. Instead, it conceptualizes GenAI, peers, learners and teachers as complementary mediating resources within a socially situated EFL speaking activity system. Thus, the framework aims to provide a more supportive context for EFL learners engaging in an authentic and meaningful conversation.

Figure 2

Collaborative learning framework for GenAI-mediated EFL speaking learning



The framework is organized around three stages of speaking activity: preparation, interaction, and feedback. These stages come from patterns identified across the reviewed studies. At the preparation stage, GenAI primarily functions as a cognitive and linguistic scaffold. Learners may use it to generate ideas, explore vocabulary and expressions, organize content, rehearse possible responses, and clarify linguistic problems before engaging in oral interaction. This may provide more supports for EFL learners, especially for those with insufficient English proficiency to engage in oral practice. At the interaction stage, learners engage in speaking tasks with their peers, while GenAI serves as an auxiliary resource for clarification, rehearsal, vocabulary support, or on-demand assistance. This arrangement addresses a central tension reflected in the challenge of unnatural and inauthentic interaction in the review. At the feedback stage, teacher involvement becomes particularly important. GenAI can provide immediate linguistic feedback and identify

possible areas for revision, while teachers contribute contextualized pedagogical judgement, interpretive support, and feedback aligned with learners' communicative goals and instructional context. This combination of GenAI feedback and teacher feedback may address limitations in the accuracy and appropriateness of GenAI feedback.

Overall, the conceptual framework is derived from the reviewed studies and is to be further validated as an instructional model. Future studies should move beyond examining effects of individual GenAI interaction on learners' speaking performance. More attentions should be given to address the emerging tensions and challenges in GenAI-mediated speaking activities to develop learners' abilities for authentic conversation.

6. Conclusion and Limitations

This systematic review examined empirical research on GenAI-mediated EFL speaking through an AT lens. Based on 43 eligible studies published between 2024 and 2026, the findings indicate that GenAI has increasingly been integrated into EFL speaking activities as a conversational partner, a feedback provider, and a learning scaffold. However, GenAI can not replace humans in interaction activities. The mediation of GenAI in EFL speaking activities leads to a reconfiguration of participants. Teachers focuses more on activity design, organization, monitoring, and pedagogical judgement from the traditional knowledge imparter and feedback provider. While peer involvement remained comparatively limited.

This reconfiguration brings several tensions. Although GenAI can provide extra personalized practice and instant feedback in a comparatively low-pressure context for EFL learners, concerns about unnatural interaction, speech recognition, and feedback reliability may constrain its contribution to meaningful oral communication. Effective GenAI mediation requires learners to possess sufficient English proficiency and AI literacy to formulate prompts, start and sustain interaction, and critically evaluate GenAI-generated feedback. These findings suggest that the educational value of GenAI depends not only on its technological affordances but also on how tools, rules, community, and division of labour align with the speaking activity's objects.

Drawing on these findings, the review proposes a collaborative framework for complementary human-GenAI mediation in EFL speaking. The framework organizes GenAI-mediated speaking activity into three operational stages of preparation, interaction, and feedback, and conceptualizes GenAI, peers, and teachers as complementary mediating resources. GenAI primarily provides cognitive and linguistic scaffolding, peers provide authentic interaction for meaning negotiation, and teachers contribute contextualized pedagogical guidance and judgement. Notably, this framework is a conceptual synthesis derived from the reviewed evidence to be further validated as an instructional model.

This review has several limitations. First, it was restricted to English-language publications retrieved from two databases, which may have limited the breadth of the evidence base. Second, the relatively small number of eligible studies and their strong geographical concentration, particularly in mainland China, may constrain the generalizability of the findings across diverse EFL contexts. Third, excluding dissertations and conference proceedings may have omitted relevant emerging evidence. Future reviews should therefore expand database and language coverage, consider a broader range of publication types, and include more diverse geographical contexts to provide a more comprehensive account of GenAI-mediated EFL speaking research.

Artificial intelligence (AI) use disclosure: The authors only used ChatGPT and Grammarly for grammar correction and language refinement during manuscript preparation.

Author contributions: All authors contributed to all processes of the manuscript, including conceptualization, data collection, data analysis, and reporting.

Data availability: No new datasets were generated or analyzed in this study. All data underlying the review were obtained from publicly available published literature. The article reports the sources included in the review and the study selection process.

Declaration of interest: The authors declare no conflict of interest.

Ethical statement: Ethical approval was not required for this study, as it was a systematic review based exclusively on previously published literature and did not involve the collection of primary data from human participants or animals.

Funding: This research received no external funding.

References

- Chen, A., Jia, J., Li, Y., & Fu, L. (2025). Investigating the effect of role-play activity with GenAI agent on EFL students' speaking performance. *Journal of Educational Computing Research*, 63(1), 99-125. <https://doi.org/10.1177/07356331241299058>
- Chen, X., Zou, D., Cheng, G., & Xie, H. (2021, July). Artificial intelligence-assisted personalized language learning: systematic review and co-citation analysis. In M. Chang, N.-S. Chen, D. G. Sampson, & A. Tlili (Eds.), *2021 international conference on advanced learning technologies (ICALT)* (pp. 241-245). IEEE. <https://doi.org/10.1109/icalt52272.2021.00079>
- Cui, W., & Zhang, J. (2026). AI applications in language learning: improving speaking fluency and pronunciation using chatbots and speech feedback. *Interactive Learning Environments*, 34(4), 2531-2546. <https://doi.org/10.1080/10494820.2025.2546634>
- Doğan, Y., & Talan, T. (2025). Artificial intelligence in foreign language learning: A bibliometric analysis. *Journal of Pedagogical Research*, 9(2), 206-230. <https://doi.org/10.33902/JPR.202427734>
- Ebadi, S., Velayati, S., Ramezanzadeh, A., & Salman, A. R. (2025). Exploring the impact of AI-powered speaking tasks on EFL learners' speaking performance and anxiety: An activity theory study. *Acta Psychologica*, 259, 105391. <https://doi.org/10.1016/j.actpsy.2025.105391>
- Engeström, Y. (2014). *Learning by expanding: An activity-theoretical approach to developmental research*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139814744>
- 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>
- Gonçalves, B. (2023). Can machines think? The controversy that led to the Turing test. *AI & Society*, 38(6), 2499-2509. <https://doi.org/10.1007/s00146-021-01318-6>
- Guo, S., Wong, R., & Lee, J. S. (2026). From sporadic to sustained: Facilitating AI-mediated informal digital learning of English (AI-IDLE) speaking practice through a teacher support framework. *Innovation in Language Learning and Teaching*. Advance online publication. <https://doi.org/10.1080/17501229.2026.2642395>
- Gusenbauer, M., & Haddaway, N. R. (2020). Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. *Research synthesis methods*, 11(2), 181-217. <https://doi.org/10.1002/jrsm.1378>
- Hong, Q. N., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M. P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M. C., Vedel, I., & Pluye, P. (2018). The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. *Education for information*, 34(4), 285-291. <https://doi.org/10.3233/efi-180221>
- 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)
- Hwang, G. J., & Chen, N. S. (2023). Exploring the potential of generative artificial intelligence in education: applications, challenges, and future research directions. *Educational Technology & Society*, 26(2), 1-18. [https://doi.org/10.30191/ETS.202304_26\(2\).0014](https://doi.org/10.30191/ETS.202304_26(2).0014)
- Kayhan, S. (2025). The effect of intervention studies on foreign language speaking anxiety: A meta-analysis study. *Journal of Pedagogical Research*, 9(3), 62-83. <https://doi.org/10.33902/JPR.202533156>
- Kim, H. S., Kim, N. Y., & Cha, Y. J. (2021). Is it beneficial to use AI chatbots to improve learners' speaking performance?. *Journal of Asia TEFL*, 18(1), 161-178. <https://doi.org/10.18823/asiatefl.2021.18.1.10.161>
- Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELJ Journal*, 54(2), 537-550. <https://doi.org/10.1177/00336882231162868>
- Koszalka, T. A., & Wu, C. P. (2004). *A cultural historical activity theory [CHAT] analysis of technology integration: Case study of two teachers* (ED485000). ERIC.
- Krathwohl, D. R. (2002). A revision of bloom's taxonomy: An overview. *Theory Into Practice*, 41(4), 212-218.

https://doi.org/10.1207/s15430421tip4104_2

- Lai, W. Y. W., & Lee, J. S. (2024). A systematic review of conversational AI tools in ELT: Publication trends, tools, research methods, learning outcomes, and antecedents. *Computers and Education: Artificial Intelligence*, 7, 100291. <https://doi.org/10.1016/j.caeai.2024.100291>
- Lee, J. (2025). Speaking English with AI or Humans: What Engages EFL Learners More?. *English Teaching*, 80(5), 41-66. <https://doi.org/10.15858/engtea.80.5.202512.41>
- Li, B., Tan, Y. L., Wang, C., & Lowell, V. (2025). Two years of innovation: A systematic review of empirical GenAI research in language learning and teaching. *Computers and Education: Artificial Intelligence*, 9, 100445. <https://doi.org/10.1016/j.caeai.2025.100445>
- Liu, H., Lv, C., & Chen, J. (2025). Willingness to communicate and anxiety in human-human and human-chatbot interaction contexts: An idiodynamic investigation. *Innovation in Language Learning and Teaching*. Advance online publication. <https://doi.org/10.1080/17501229.2025.2560101>
- Liu, X., Tang, Q., Wu, W., Chen, X., & Wang, C. (2026). GenAI-agents in education: A systematic review of macro-level trends, educational roles, theoretical foundations and future agendas. *Interactive Learning Environments*. Advance online publication. <https://doi.org/10.1080/10494820.2026.2672562>
- Ma, M., Noordin, N., & Razali, A. B. (2025). Improving EFL speaking performance among undergraduate students with an AI-powered mobile app in after-class assignments: An empirical investigation. *Humanities and Social Sciences Communications*, 12, Article 370. <https://doi.org/10.1057/s41599-025-04688-0>
- McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955. *AI magazine*, 27(4), 12-14. <https://doi.org/10.1609/aimag.v27i4.1904>
- Nguyen, D. T., Vu, P. X., & Loi, N. T. (2026). Understanding school-related well-being in primary education: A systematic review and integrated framework. *Journal of Pedagogical Sociology and Psychology*, 8(3), e47001. <https://doi.org/10.33902/JPSP.202647001>
- 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., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, Article n71. <https://doi.org/10.1136/bmj.n71>
- Russell, S. J., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson.
- Snowman, J., McCown, R., & Biehler, R. (2011). *Psychology applied to teaching* (13thed.). Wadsworth Cengage Learning.
- Tai, T.-Y. (2026). ChatGPT-assisted EFL learning beyond the classroom: Impact on oral proficiency, enjoyment, and grit. *Computer Assisted Language Learning*. Advance online publication. <https://doi.org/10.1080/09588221.2026.2640087>
- Tai, T.-Y., & Chen, H. H.-J. (2025). Impact of generative AI chatbots and interaction modes on the speaking proficiency of adolescent EFL learners. *Computer Assisted Language Learning*. Advance online publication. <https://doi.org/10.1080/09588221.2025.2572999>
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433-460. <https://doi.org/10.1093/mind/LIX.236.433>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- 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. <https://doi.org/10.1016/j.system.2024.103533>
- Weizenbaum, J. (1966). ELIZA—A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36-45. <https://doi.org/10.1145/365153.365168>
- Wu, T. T., Hapsari, I. P., & Huang, Y. M. (2026). Effects of incorporating AI chatbots into think-pair-share activities on EFL speaking anxiety, language enjoyment, and speaking performance. *Computer Assisted Language Learning*, 39(5-6), 1408-1446. <https://doi.org/10.1080/09588221.2025.2478271>
- Xing, C., & Saeed, M. A. (2026). A systematic review on artificial intelligence (AI) technologies in ESL/EFL speaking skills. *International Journal of TESOL Studies*, 8(2), 240-270. <https://doi.org/10.58304/ijts.250908>
- Yang, H., & Kyun, S. (2022). The current research trend of artificial intelligence in language learning: A systematic empirical literature review from an activity theory perspective. *Australasian Journal of Educational Technology*, 38(5), 180-210. <https://doi.org/10.14742/ajet.7492>

- Zhai, C., & Wibowo, S. (2023). A systematic review on artificial intelligence dialogue systems for enhancing English as foreign language students' interactional competence in the university. *Computers and Education: Artificial Intelligence*, 4, 100134. <https://doi.org/10.1016/j.caeai.2023.100134>
- Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., Liu, J. B., Yuan, J., & Li, Y. (2021). A Review of Artificial Intelligence (AI) in Education from 2010 to 2020. *Complexity*, 2021(1), 8812542. <https://doi.org/10.1155/2021/8812542>
- Zhang, H., Su, Y., Xu, X., Lei, V. N. L., & Hao, S. (2025). Beyond theory: A mixed-methods exploration of postgraduate engagement with ChatGPT for English speaking proficiency. *International Journal of Technology in Education*, 8(4), 1022-1047. <https://doi.org/10.46328/ijte.1221>
- Zhang, W. (2026). The impact of AI chatbots on EFL learners' oral proficiency and willingness to communicate. *System*, 136, 103919. <https://doi.org/10.1016/j.system.2025.103919>
- Zheng, X. L., Kim, H. S., Lai, W. H., & Hwang, G. J. (2020). Cognitive regulations in ICT-supported flipped classroom interactions: An activity theory perspective. *British Journal of Educational Technology*, 51(1), 103-130. <https://doi.org/10.1111/bjet.12763>
- Zhou, Q., Hashim, H., & Sulaiman, N. A. (2023). A Systemic Review (2014-2023) on the Mobile-Assisted Blended Learning for English as a Foreign Language Education with a Focus on Empirical Studies. *International Journal of Interactive Mobile Technologies*, 17(24), 153-172. <https://doi.org/10.3991/ijim.v17i24.42019>
- Zhou, Q., Hashim, H., & Sulaiman, N. A. (2025). Supporting English speaking practice in higher education: The impact of AI chatbot-integrated mobile-assisted blended learning framework. *Education and Information Technologies*, 30(10), 14629-14660. <https://doi.org/10.1007/s10639-025-13401-2>
- Zou, B., Xie, S., & Wang, C. (2025). Students' willingness to Communicate (WTC) in using artificial intelligence (AI) technology in English-speaking practice. *International Journal of Information and Communication Technology Education*, 21(1), 1-18. <https://doi.org/10.4018/IJICTE.375387>