



Review Article

Generative AI-Mediated Speaking Practice and Intercultural Communicative Competence in EFL Contexts: A Scoping Review of the Literature (2022-2026)

Do Ngoc Lan¹ , Nguyen Xuan Hong^{2,*}

¹Phuong Nam Language Center, Ho Chi Minh City, Vietnam

²Faculty of Foreign Languages, Industrial University of Ho Chi Minh City, Ho Chi Minh City, Vietnam

Abstract

Speaking is widely regarded as the most challenging skill for English as a Foreign Language (EFL) learners, and generative artificial intelligence (GenAI) has rapidly become a prominent resource for oral practice, while a parallel but largely separate strand of research has examined how artificial intelligence (AI)-mediated tools contribute to intercultural communicative competence (ICC). This scoping review maps the evidence base at the intersection of GenAI-mediated speaking practice, oral fluency and ICC, with particular attention to the underrepresented Vietnamese EFL context. Following a PRISMA-ScR-informed approach, 17 empirical and review studies (2022-2026) were thematically synthesized into four clusters, namely GenAI and oral fluency, GenAI and affective-communicative outcomes, AI-mediated ICC and Vietnam-specific studies, alongside three foundational theoretical works. Findings show that GenAI-mediated speaking practice consistently improves short-term fluency, willingness to communicate and speaking anxiety; that AI-mediated interaction can support ICC development when paired with reflective, culturally responsive pedagogy, though this evidence remains preliminary; and that the Vietnamese literature, while fast-growing, is small and methodologically narrow, with no study yet examining ICC outcomes. Critically, no identified study combines a longitudinal design with simultaneous measurement of oral fluency and ICC within a single GenAI-mediated speaking intervention, opening a worthwhile direction for doctoral-level research. This review argues for such an integrated research agenda, situates it within Sociocultural Theory, the Interaction Hypothesis and Byram's ICC model and outlines its theoretical, methodological and pedagogical implications for the Vietnamese higher-education context.

Keywords

Generative AI, EFL Learners, Oral Fluency, Intercultural Communicative Competence, Scoping Review

1. Introduction

Speaking is consistently identified as the most anxiety-provoking and difficult-to-develop skill for EFL learners, particularly in foreign language contexts where opportunities for authentic oral interaction outside the classroom are limited [3].

Large classes, limited contact hours and learners' fear of negative evaluation combine to restrict the amount of meaningful spoken output most EFL students produce, a constraint that is especially acute in Vietnamese tertiary classrooms where

*Correspondence: Nguyen Xuan Hong (nguyenxuanhong@iuh.edu.vn)

Received: 27 August 2026; Accepted: 7 September 2026; Published: 30 September 2026



Copyright: © The Author(s), 2026. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

English is typically taught as one subject among many rather than as a medium of instruction. The emergence of generative AI (GenAI) tools capable of sustained, adaptive conversational interaction, including large language model-based chatbots and AI voice assistants, has introduced a qualitatively new practice resource because it is available on demand, low-stakes and capable of real-time linguistic feedback. A rapidly growing empirical literature has examined the effects of such tools on EFL learners' speaking outcomes, predominantly fluency, willingness to communicate (WTC) and speaking anxiety [12, 15, 18].

At the same time, English is increasingly used as a common language for communication among speakers of diverse linguistic and cultural backgrounds, placing growing importance on intercultural communicative competence (ICC) which is the ability to interact appropriately and effectively with people from other cultures [1]. Speaking proficiency and intercultural competence are, in principle, closely linked outcomes of the same communicative act. Particularly, a fluent utterance that is culturally inappropriate can still fail to achieve its communicative purpose, while an intercultural sensitive learner who lacks oral fluency may be unable to express that sensitivity in real time. A separate but related literature has begun to explore how AI-mediated environments including dialogue systems, virtual reality and conversational agents can support the development of ICC [5, 9, 19].

Despite the conceptual proximity of these two strands due to the fact that both concern AI-mediated oral or interactional practice and both are frequently framed around learners' preparation for real-world, cross-cultural communication, they have rarely been examined together, and the magnitude of this separation can be shown rather than merely asserted. Two of the most comprehensive syntheses available in this literature illustrate the scale of the gap directly. A meta-analysis pooling 21 primary studies on AI chatbots and EFL skill development [17] did not incorporate any measure of intercultural competence and a PRISMA-guided systematic review of 28 studies on AI dialogue systems and interactional competence [19] identified no study that combined fluency and ICC outcomes (see Section 4.1 and 4.3 for detail). This separation is not merely an artifact of disciplinary specialization because it has substantive consequences. Speaking-focused studies tend to rely on standardized fluency, accuracy or anxiety measures and short intervention windows while ICC-focused studies tend to rely on attitudinal surveys, reflective journals or case studies and rarely report any parallel measure of oral performance. As a result, it remains unknown whether the same GenAI-mediated speaking practice that reliably improves fluency can also support the attitudes, knowledge and interpretive skills that constitute ICC simultaneously, or whether the two kinds of gain require different pedagogical designs altogether. Addressing this gap, the present scoping review is guided by three review questions: (1) What outcomes has GenAI-mediated speaking practice been shown to produce for EFL learners' oral fluency and related affective variables? (2)

What outcomes has AI-mediated interaction been shown to produce for intercultural communicative competence? and (3) To what extent, and how, have these two lines of inquiry been integrated, particularly within the Vietnamese EFL context?

The review is intended to serve two objects. For researchers, it consolidates a literature that is currently scattered across applied-linguistics, educational-technology and communication journals, and it makes explicit a research gap that motivates a longitudinal, mixed-methods doctoral research agenda (Section 5). For practitioners and program designers, particularly in Vietnamese higher education institutions that are beginning to integrate GenAI tools into speaking curricula, it offers a thematically organized summary of what is currently known and not known about the likely benefits and limits of such tools. The remainder of this review is organized as follows: Section 2 outlines the theoretical framework; Section 3 details the review method; Section 4 presents the thematic synthesis including a tabulated summary of study characteristics; Section 5 discusses the synthesis and identifies the resulting research gap; and Sections 6 and 7 present the review's limitations and conclusion, respectively.

2. Theoretical Framework

This review is organized around three theoretical anchors, each of which addresses a different aspect of the GenAI-speaking-ICC relationship, particularly how learning happens through mediated interaction, why interaction drives acquisition and precisely what intercultural competence consists of.

2.1. Sociocultural Theory and Mediation

Sociocultural Theory of [11] considers GenAI tools as mediating ones within learners' Zone of Proximal Development (ZPD), supporting language production that might not yet be achievable unassisted. In a GenAI-mediated speaking task, the chatbot or voice assistant functions analogously to a more capable interlocutor because it can rephrase, simplify, model target-like pronunciation or supply vocabulary on demand, allowing the learners to perform slightly beyond their current independent levels. This analogy should nonetheless be drawn cautiously. Unlike the knowledgeable human mediator envisaged in Vygotsky's original account, a GenAI system has no genuine understanding of the learner, no intentional diagnostic model of what that learner currently can and cannot do, and generates its "scaffolding" through statistical pattern completion rather than the mutual, intersubjective attunement that characterizes human mediation. The GenAI-as-interlocutor analogy is therefore treated here as a heuristic that captures surface-level functional similarities in support behavior (rephrasing, modeling, on-demand vocabulary) rather than a claim that GenAI mediation is psychologically equivalent to human mediation within the ZPD; whether any of these specific functions actually operate as ZPD-consistent scaffolding for learners, as opposed to merely resembling it, remains an empirical question that the studies reviewed in Section 4 do not directly test. With that caveat in place, unlike

a human tutor, a GenAI interlocutor is available at any hour, does not fatigue and can repeat the same scaffolding indefinitely without apparent frustration. These are properties that plausibly widen access to ZPD-level practice for learners who otherwise have few opportunities for one-on-one oral interaction. This characteristic is directly relevant to the Vietnamese context, where large class sizes and limited speaking-focused contact hours constrain the amount of individualized scaffolding a teacher can realistically provide.

2.2. The Interaction Hypothesis

Interaction Hypothesis of [6] explains how negotiation of meaning including the clarification and repair sequences learners engage in with conversational AI can drive language acquisition. When a learner's utterance is misunderstood by an AI system and the system requests clarification, or when the learner must reformulate an utterance for the system to respond appropriately, the resulting negotiation sequence is held to make target-language input more comprehensible and to draw learners' attention to gaps in their own production. Several of the studies reviewed in Section 4.1 can be read as implicit tests of this mechanism. That is repeated chatbot interaction over several weeks is associated with fluency gains that the Interaction Hypothesis would predict from sustained negotiation-of-meaning practice even though few of the primary studies explicitly measure negotiation sequences as a mediating variable. This application of the Interaction Hypothesis to human-AI interaction should, however, be treated critically rather than assumed to hold in the same way it does between human speakers. The hypothesis was originally formulated for interlocutors who share genuine communicative intent, real-world conversational stakes and the capacity for authentic misunderstanding; a GenAI system, by contrast, does not misunderstand in that sense, and its clarification requests or reformulations may reflect model uncertainty, prompt design or training-data artifacts rather than a communicative breakdown analogous to those between human speakers. Whether AI-elicited negotiation sequences engage the same acquisition-relevant noticing and modified-output processes as human-human negotiation is therefore an open empirical question rather than a settled extension of the theory. Correspondingly, the fluency gains reported in Section 4.1 following chatbot interaction cannot, on their own, confirm that negotiation of meaning is the operative mechanism, since they are equally consistent with simpler accounts such as increased practice volume or a reduced affective filter. Future research applying the Interaction Hypothesis to GenAI contexts should therefore measure negotiation sequences directly, as noted in Section 5.3, rather than inferring their operation from fluency outcomes alone.

2.3. Byram's Model of Intercultural Communicative Competence

Model of Intercultural Communicative Competence of [1]

provides the framework used across the reviewed ICC-focused studies to operationalize and measure intercultural competence. The model comprises five interrelated components such as: (a) attitudes (curiosity and openness, readiness to suspend disbelief about other cultures and belief about one's own); (b) knowledge (of social groups, their products and practices and of the general processes of societal and individual interaction); (c) skills of interpreting and relating (the ability to interpret a document or event from another culture and relate it to one's own); (d) skills of discovery and interaction (the ability to acquire new knowledge of a culture and cultural practices and to operate that knowledge under real-time communication constraints); and (e) critical cultural awareness (the ability to evaluate critically and on the basis of explicit criteria, perspectives, practices and products in one's own and other cultures). This componential structure matters for the present review because it clarifies why ICC is difficult to capture with a single fluency-style outcome measure due to the fact that a GenAI intervention that builds vocabulary and repair strategies is not automatically also building critical cultural awareness, and reviewed studies vary considerably in which of the five components they actually measure.

In sum, taken together, the three frameworks mentioned above suggest a plausible but empirically untested causal chain. That is GenAI tools act as mediating artifacts [11] that afford repeated negotiation-of-meaning practice [6], and this practice is hypothesized, though rarely directly tested, to be one of the mechanisms through which AI-mediated interaction could simultaneously cultivate the skills and awareness components of ICC [1]. The theoretical logic of this chain is precisely what makes its absence from the empirical literature, documented in Sections 4 and 5, a substantive rather than a merely administrative gap.

3. Method

3.1. Review Design

This review follows a scoping review approach, informed by the PRISMA extension for Scoping Reviews (PRISMA-ScR), which is appropriate for mapping a heterogeneous and rapidly evolving evidence base rather than aggregating effect sizes from methodologically homogeneous studies. A scoping design was chosen over a full systematic review for three reasons. First, the GenAI-speaking-ICC literature spans widely different study designs (true experiments, quasi-experiments, surveys, case studies and conceptual papers) which are not commensurable in a meta-analytic sense. Second, the review's purpose is explicitly to identify a research gap and inform the design of a subsequent primary study rather than to produce a definitive effect estimate. Finally, the field is young enough (the earliest eligible empirical study dates to 2022) that a scoping map of 'what has been studied and how' is of more immediate practical value than a narrower effect-size synthesis would be.

3.2. Search Strategy and Eligibility Criteria

Sources were identified through iterative keyword searches (e.g., "generative AI speaking EFL," "AI chatbot fluency," "AI intercultural communicative competence," "Vietnamese EFL AI speaking") covering publications from 2022 to 2026. Eligibility was determined against the explicit inclusion and exclusion criteria below.

Inclusion criteria: (a) peer-reviewed journal articles, systematic reviews or conference proceedings published in English between 2022 and 2026; (b) addressing AI or GenAI-mediated oral/speaking practice in EFL/ESL (English as a Second Language) contexts, or AI-mediated development of intercultural communicative competence in language education; and (c) foundational theoretical works predating the 2020 window were additionally retained, on the grounds that a scoping review of an applied literature still requires an explicit theoretical anchor even when that anchor was established before the review's empirical window begins.

Exclusion criteria: (a) sources unrelated to language education; (b) sources addressing only written/reading skills with no spoken or interactional component; (c) sources focused exclusively on writing-support tools (e.g., automated writing evaluation) with no speaking or interactional component; and (d) non-English-language publications and sources for which neither the full text nor a sufficiently detailed abstract could be obtained.

3.3. Screening and Data Extraction

Titles and abstracts were screened for topical relevance. Full texts were then reviewed to confirm eligibility and to extract study design, sample, AI modality and reported outcomes. For a small number of sources where the full text could not be obtained despite institutional and interlibrary access attempts, eligibility and extraction relied instead on a detailed abstract. In these cases, the extracted characteristics were additionally cross-checked, where possible, against secondary descriptions of the same study in other reviews or meta-analyses citing it, and any characteristic that could not be corroborated in this way was recorded as "not reported" rather than inferred. This limited use of abstract-only extraction is acknowledged as a source of potential incompleteness and is revisited as an explicit limitation in Section 6. For each eligible study, the following categories of information were extracted where reported such as publication year and outlet; study design (experimental, quasi-experimental, survey, systematic review/meta-analysis, case study or conceptual); sample size, learner population and setting; the specific AI tool or modality used (text-based chatbot, voice-based intelligent personal assistant, dialogue system or virtual-reality environment); the outcome construct measured (fluency, WTC, speaking anxiety, self-perceived competence or a component of ICC); the direction and statistical significance of the reported effect. Seventeen empirical or review studies (published 2022-2026) met

the inclusion criteria and form the basis of the thematic synthesis in Section 4. Three additional foundational theoretical works [1, 6, 11], pre-dating the eligibility window, were retained to ground the theoretical framework in Section 2. The extracted characteristics of the 17 empirical/review studies, including sample size and country/context, are summarized in Table 1 (Section 4) to allow direct cross-study comparison of design, sample and AI modality alongside the narrative synthesis.

3.4. Quality Considerations

Because this is a scoping rather than a systematic review, no formal risk-of-bias scoring was applied to individual studies, consistent with PRISMA-ScR guidance that quality appraisal is optional and should be reported transparently either way. Nonetheless, three recurring design features were noted qualitatively during extraction, as they bear directly on how much weight the synthesis in Sections 4 and 5 places on any single finding. They are intervention duration (most reviewed interventions run four to fourteen weeks), the presence or absence of a comparison/control group (present in the true- and quasi-experimental studies in Cluster 1, absent in most Cluster 3 and Cluster 4 studies) and whether outcome measurement relied on standardized instruments versus researcher-developed surveys. These features are revisited as explicit limitations of the underlying evidence base in Section 5.

4. Results

The 17 eligible studies were grouped into four thematic clusters, presented in order of their relevance to the review's three guiding questions such as studies addressing oral fluency (Section 4.1), studies addressing affective-communicative outcomes (Section 4.2), studies addressing intercultural communicative competence (Section 4.3) and studies situated specifically within the Vietnamese EFL context (Section 4.4). Table 1 summarizes the design, sample and AI modality of each study to complement the narrative synthesis below.

4.1. GenAI Chatbots and Oral Fluency/Proficiency

Multiple studies report significant gains in speaking fluency and oral proficiency following GenAI-mediated practice. [3] designed a true experimental study, with random assignments to experimental and control groups and found that AI-mediated interaction improved EFL learners' speaking skills and willingness to communicate relative to conventional instruction. The randomized design gives this study comparatively strong internal validity within the reviewed set. [4] similarly reported that sustained practice with an intelligent personal assistant (Google Assistant) improved fluency and comprehensibility while reducing accentedness, extending the chatbot-focused literature to voice-based, embodied conversational

agents rather than text interfaces. [18] demonstrated that an AI chatbot functioning as a conversation partner supported measurable gains in EFL speaking classes, and [15] showed that embedding AI chatbots within think-pair-share activities improved speaking performance alongside reduced anxiety and greater language enjoyment, suggesting that GenAI tools may be most effective when combined with, rather than substituted for, collaborative classroom tasks.

At a broader scale, a meta-analysis by [17], synthesizing 21 primary studies published between 2008 and 2023, corroborated these single-study findings, confirming a positive overall effect of AI chatbots across multiple EFL skill domains including speaking, and noting that effect sizes vary with intervention duration and chatbot modality (text- versus voice-based) which is a moderating factor with direct relevance to the longitudinal design proposed in Section 5. Because the meta-analysis pools studies spanning fifteen years and a wide range of chatbot technologies, from rule-based systems to contemporary large-language-model agents, its aggregate effect size should be read as an approximate lower bound on what current-generation GenAI tools, which is markedly more linguistically flexible than the chatbots available for most of that window, are likely to achieve. However, none of the 21 pooled studies incorporated any measure of intercultural competence, reinforcing the separation between the fluency and ICC literatures noted in Section 1.

4.2. GenAI and Affective-communicative Outcomes

Beyond fluency, a related cluster of studies emphasizes affective and communicative-readiness outcomes that are theorized to precede or accompany fluency gains. [12] compared different conversational GenAI chatbots and identified differential effects on willingness to communicate, foreign language speaking anxiety and self-perceived communicative competence, indicating that not all GenAI tools produce equivalent affective benefits and that tool-selection decisions in curriculum design are not interchangeable. [13] turned attention to teachers rather than learners, examining EFL instructors' affective and cognitive responses to ChatGPT and underscoring that successful classroom integration depends as much on instructor readiness, confidence and attitudes as on learner engagement. This is a dimension almost entirely absent from the learner-centered studies in Cluster 1. [14] further discussed GenAI's potential to bridge participation divides and foster inclusivity in language education, arguing that low-stakes AI practice may particularly benefit learners who are reticent in teacher-fronted classrooms, though this discussion remains largely conceptual rather than empirical and would benefit from direct testing against the kind of experimental designs used in Cluster 1.

4.3. AI-mediated Intercultural Communicative Competence

A thematically distinct but conceptually adjacent literature addresses AI and ICC directly. Systematic review of 11 empirical studies by [5] concluded that AI tools such as chatbots and virtual reality can support ICC development when integrated with reflection and culturally responsive pedagogy, while cautioning against risks of cultural bias. AI systems trained predominantly on data from dominant linguistic and cultural contexts may model or reinforce a narrow range of communicative norms and unequal technology access among learners. [9] assessed AI's impact on intercultural communication among 115 postgraduate students from nine different countries in a multicultural university setting, using a purpose-built survey instrument. The majority of participants reported that AI already reduced perceived language and cultural barriers in their daily interactions, a finding that speaks to attitudes toward AI-mediated intercultural contact in general rather than to any classroom speaking intervention specifically. In a PRISMA-guided systematic review, [19] identified 28 eligible studies on AI dialogue systems for EFL interactional competence and derived six influencing dimensions spanning linguistic, affective and sociocultural factors. [20] extended this line of work through a case study of an AI dialogue system explicitly designed to model intercultural, humorous and empathetic dimensions of communication, illustrating that such dimensions can in principle be deliberately engineered into a system's design rather than emerging incidentally. [16] proposed a Cross-Cultural Intelligent Language Learning System (CILS) that integrates AI-driven personalization to improve both linguistic proficiency and cross-cultural communication, offering one of the few system-level attempts in the reviewed literature to target fluency and intercultural outcomes within a single tool. However, this is system-design and evaluation study, not a classroom-based intervention with pre-registered outcome measures. Notably, none of these ICC-focused studies incorporated a standardized measure of oral fluency, nor did any adopt a longitudinal design extending beyond a single semester.

4.4. Vietnam-specific Empirical Studies

A small but growing number of studies situate GenAI-mediated speaking practice within the Vietnamese EFL context specifically. [2] conducted an eight-week quasi-experiment with 30 Vietnamese undergraduate students using the Andy AI voice chatbot and reported statistically significant improvements ($p < .05$) in speaking accuracy across grammar, vocabulary and pronunciation, alongside students' own perceptions of improved speaking ability. It becomes one of the studies reviewed here which is the most methodologically comparable Vietnamese counterpart to the international quasi-experimental studies in Cluster 1. [8], based at the Industrial University of Ho Chi Minh City (IUH), found that regular use

of the Call Annie AI conversational agent as a homework activity improved students' emotional and behavioral engagement in speaking lessons, extending the evidence base beyond linguistic outcome measures to classroom engagement. [7] examined how Vietnamese EFL learners draw on translanguaging strategies, which is moving fluidly between Vietnamese and English when prompting GenAI tools, revealing learner-related factors such as proficiency level and confidence that shape these practices and that plausibly also mediate speaking-practice outcomes in Cluster 1-type interventions, though this mediating link has not yet been tested directly. [10] explored Vietnamese EFL students' technology-acceptance perceptions of ChatGPT for language learning, situating adoption within a technology-acceptance framework in which perceived usefulness and ease of use predict learners' willingness to engage with GenAI tools in the first place. This is a precondition for any of the speaking or ICC benefits identified elsewhere in this review to materialize at scale. Across these four studies, sample sizes are generally modest, interventions are short-term (a single semester or less) and none measure intercultural communicative competence alongside speaking outcomes. It is the specific gap that this review's proposed research agenda is designed to fill.

4.5. Summary of Study Characteristics

Table 1 consolidates the design, sample, AI modality and key finding of each of the 17 empirical/review studies discussed above, organized by the four thematic clusters. Read across rows, three cross-cutting patterns are visible. First, experimental and quasi-experimental designs with a comparison group are concentrated almost entirely in Cluster 1 (fluency-focused studies). Clusters 3 and 4 rely predominantly on surveys, systematic reviews, case studies or single-group pre-post designs. Second, sample sizes for primary (non-review) studies are uniformly modest, ranging from roughly 30 to 115 participants, which limits the statistical power available to detect any but moderate-to-large effects. Third, AI modality is unevenly distributed. Particularly, text- and voice-based chatbots dominate Cluster 1 and the Vietnam-specific literature, whereas dialogue systems and virtual-reality environments feature more prominently in the ICC-focused Cluster 3, a pattern that may itself partly explain why the two outcome domains have rarely been measured together because the tools most commonly used to study fluency are not always the same tools used to study ICC.

Table 1. Summary of reviewed empirical/review studies (2022-2026), organized by thematic cluster.

Study	Design	Sample	Sample size	Country	AI modality	Key finding
<i>Cluster 1 – GenAI chatbots and oral fluency/proficiency (Section 4.1)</i>						
[3]	True experiment (randomized experimental/control groups)	EFL learners, single institution	65 total (33 experimental, 32 control)	Iran	AI-mediated conversational interaction	AI-mediated practice group outperformed the control group on speaking skill and willingness to communicate (WTC) measures.
[4]	Quasi-experimental, pre-post	EFL learners using an intelligent personal assistant	20	Iran	Google Assistant (voice-based IPA)	Gains in fluency and comprehensibility and a reduction in perceived accentness after sustained IPA practice.
[18]	Quasi-experimental classroom intervention	EFL speaking-class learners	314	South Korea	Text-based AI chatbot as conversation partner	Chatbot-mediated conversation practice produced measurable classroom speaking gains.
[15]	Quasi-experimental, think-pair-share design	EFL learners in Indonesia	75	Indonesia	AI chatbot embedded in collaborative tasks	Combining chatbot practice with think-pair-share activities reduced speaking anxiety and raised enjoyment and performance.
[17]	Meta-analysis (21 primary studies, 2008-2023)	Aggregated EFL samples across studies	Not reported in aggregate (21 studies pooled)	Multiple (meta-analysis; not country-specific)	Various AI chatbots	Overall positive, moderate effect of AI chatbots on EFL language-skill development, moderated by intervention duration and modality (text vs. voice).
<i>Cluster 2 – GenAI and affective-communicative outcomes (Section 4.2)</i>						

Study	Design	Sample	Sample size	Country	AI modality	Key finding
[12]	Comparative quantitative study	Chinese EFL learners using different chatbots	99 total (control N=33; two experimental groups, N=33 each)	China	Multiple conversational GenAI chatbots	Chatbot type differentially affected WTC, speaking anxiety and self-perceived communicative competence.
[13]	Quantitative survey	EFL teachers (higher education)	187	China	ChatGPT (teacher-facing use)	Teachers' affective and cognitive responses to ChatGPT shape classroom integration as much as learner-side factors do.
[14]	Conceptual/discussion paper	Not applicable	Not applicable	Not applicable	GenAI tools (general)	Argues GenAI can bridge participation divides and foster inclusivity, but offers little direct empirical evidence.
<i>Cluster 3 – AI-mediated intercultural communicative competence (Section 4.3)</i>						
[5]	Systematic review (11 empirical studies)	University students, multiple countries	Not aggregated (11 studies reviewed)	Multiple (systematic review)	Chatbots and virtual-reality simulations	AI can support ICC development but raises cultural-bias and unequal-access risks; calls for human oversight.
[9]	Cross-sectional survey	postgraduate students from nine countries	115	Malaysia (study site; students from 9 countries)	AI in daily communication (general, not speaking-specific)	Most participants viewed AI as reducing language and cultural barriers and easing cross-cultural interaction.
[19]	PRISMA-guided systematic review (28 studies)	EFL learners, multiple contexts	Not aggregated (28 studies reviewed)	Multiple (systematic review)	AI dialogue systems	Identifies six dimensions shaping AI's contribution to interactional competence; no study combined fluency and ICC outcomes.
[20]	Case study	EFL learners interacting with one system	37	Not stated in source	AI dialogue system modeling humor/empathy	Illustrates how affective and intercultural dimensions can be deliberately engineered into a dialogue system.
[16]	System design and evaluation study	EFL learners	Not specified	Not stated in source	Cross-Cultural Intelligent Language Learning System (CILS)	An AI-personalized system can jointly target linguistic proficiency and cross-cultural communication skills.
<i>Cluster 4 – Vietnam-specific empirical studies (Section 4.4)</i>						
[2]	8-week quasi-experiment	Vietnamese undergraduates	30	Vietnam	AI voice chatbot (Andy)	Statistically significant gains ($p < .05$) in speaking accuracy across grammar, vocabulary and pronunciation.
[8]	Classroom-based study, IUH	Vietnamese EFL students	77	Vietnam	Call Annie AI conversational agent (homework)	Regular homework use improved students' emotional and behavioral engagement in speaking lessons.
[7]	Exploratory mixed-methods study	Vietnamese EFL learners	68	Vietnam	GenAI prompting (text-based)	Learners draw on translanguaging strategies when prompting GenAI, shaped by proficiency and

Study	Design	Sample	Sample size	Country	AI modality	Key finding
[10]	Survey (technology-acceptance framework)	Vietnamese English-major students	369	Vietnam	ChatGPT/ GenAI tools (general)	confidence. Perceived usefulness and ease of use shape students' acceptance of GenAI for language learning.

5. Discussion: Synthesis, Identified Gaps and a Future Research Direction

5.1. Synthesis

Taken together, the reviewed literature supports two conclusions with reasonable consistency: (a) GenAI-mediated speaking practice is associated with improved EFL learners' oral fluency and reduced speaking-related anxiety over short intervention periods and (b) AI-mediated interaction can, under appropriate pedagogical conditions, support the development of intercultural communicative competence. Through the theoretical lens of Section 2, finding (a) is broadly consistent with the Interaction Hypothesis - repeated negotiation-of-meaning episodes with a responsive AI interlocutor plausibly drive the fluency gains reported across Cluster 1 while finding (b) is consistent with Sociocultural Theory's account of AI as a mediating artifact that can scaffold not only linguistic but also interpretive and attitudinal development, provided the mediation is paired with human-led reflection, as [5] emphasize. However, three specific gaps emerge from this synthesis:

No identified study measures oral fluency and intercultural communicative competence as joint outcomes of the same GenAI-mediated speaking intervention.

Existing intervention studies are predominantly short-term (4-14 weeks). Longitudinal evidence on the durability of fluency or ICC gains is absent.

Empirical evidence from the Vietnamese EFL context remains limited in scale and scope, with no study yet examining ICC outcomes specifically.

5.2. Research Gaps

These three gaps are not independent of one another. The absence of joint fluency-ICC measurement (gap 1) is partly a consequence of the modality split noted in Section 4.5. Particularly, studies using chatbots to build fluency rarely also apply five-component ICC framework of [1], while studies applying that framework rarely use the standardized fluency or WTC instruments common in Cluster 1. The short-term nature of existing designs (gap 2) further limits what can be said about either outcome because a four- to fourteen-week intervention

may be sufficient to detect a fluency gain but is unlikely to capture the slower-forming attitudinal and critical-awareness components of ICC which plausibly require sustained, reflective engagement over a full academic year or longer. The Vietnamese-context gap (gap 3) compounds both of the preceding gaps due to the fact that the four Vietnam-specific studies identified in Section 4.4 are themselves short-term and fluency- or engagement-focused, so even the modest evidence base that exists internationally for AI-mediated ICC development has, as yet, no local counterpart.

A further point of difference across clusters concerns whose readiness is measured. Cluster 1 and most of Cluster 4 measure learner-side outcomes exclusively, whereas [13] is the only reviewed study to measure teacher-side affective and cognitive responses to GenAI tools. Given that Vietnamese higher-education institutions including IUH, where the [8] study was conducted, are actively deciding how and how much to integrate GenAI speaking tools into existing curricula, the near-total absence of Vietnamese teacher-perspective research represents a practical, as well as theoretical, blind spot that a future study could usefully address alongside the learner-focused fluency-ICC gap identified above.

5.3. A Future Research Direction

Because the three gaps arise from different, if related, causes, they call for partly distinct methodological responses rather than a single generic recommendation to "do more research." The directions below are organized against each gap in turn before being integrated into a single proposed design.

Addressing gap 1 (no joint fluency-ICC measurement): The most direct response is instrument pairing rather than instrument invention. Future studies should administer, within the same sample and the same intervention, at least one standardized oral-fluency or WTC measure of the kind used across Cluster 1 (e.g., timed speech-sample coding, the WTC scales used by [12] alongside an ICC instrument explicitly mapped onto five components of [1] rather than a generic "intercultural sensitivity" scale. Where no such component-level instrument yet exists for the Vietnamese EFL population, an interim step, as itself a publishable contribution, would be to adapt and validate one, since measurement-instrument adaptation is a precondition for, rather than a distraction from, the joint-outcome study envisaged here. Analytically, the two outcome sets should be modeled together (e.g., as parallel or correlated

growth trajectories) rather than reported in separate tables, so that the study can speak directly to whether fluency and ICC gains covary, diverge or trade off against one another.

Addressing gap 2 (short intervention windows): Given that the attitudinal and critical-awareness components of ICC plausibly form more slowly than fluency gains, future designs should extend well beyond the four-to-fourteen-week norm identified in Section 4.5. Ideal spanning is at least one full academic year and including at least two or three measurement waves rather than a single pre-post comparison. A staggered or multiple-cohort design, introducing the GenAI intervention to successive cohorts a semester apart, would additionally let researchers separate maturation effects (learners simply gaining fluency and cultural exposure over time regardless of the intervention) from intervention-specific effects, a confound that none of the short-term studies in Clusters 1 through 4 were positioned to rule out.

Addressing gap 3 (limited Vietnamese-context evidence): Beyond simply replicating international designs locally, a Vietnamese-context research agenda should examine features of the local setting that plausibly moderate GenAI's effects such as large class sizes and limited speaking-focused contact hours (Section 2.1), the translanguaging strategies documented by [7] and the technology-acceptance factors documented by [10], any of which could strengthen or dampen the fluency and ICC gains observed elsewhere. Multi-site studies spanning more than one Vietnamese institution, rather than the single-institution designs typical of the existing four Vietnam-specific studies, would also allow the field to distinguish genuinely generalizable findings from features specific to one program or student population, and would support the kind of institutional comparison that a single-site doctoral study cannot provide alone.

A cross-cutting direction from the teacher's perspective: As Section 5.2 notes, [13] is the only reviewed study to examine instructors' own affective and cognitive responses to GenAI tools, and no equivalent study exists for Vietnamese EFL teachers specifically. Future research should therefore pair any learner-outcome study with a teacher-facing strand such as surveys, interviews or classroom observation to capture how Vietnamese instructors perceive, adopt and adapt GenAI speaking tools, since teacher buy-in plausibly moderates how faithfully any classroom-based intervention is implemented, and therefore how interpretable its learner-outcome findings ultimately are.

Integrating these directions, the specific study proposed here would examine GenAI-mediated speaking practice as a joint driver of oral fluency and intercultural communicative competence among Vietnamese EFL learners across at least one full academic year. Concretely, such a study would need to (a) extend the intervention window well beyond the four-to-fourteen-week norm identified in Cluster 1, with multiple measurement waves rather than a single pre-post comparison, so the more slowly forming ICC components have an opportunity to develop and be observed; (b) administer, at minimum,

one standardized fluency or WTC instrument alongside one validated instrument mapped explicitly onto five components [1], so that any correlation or dissociation between the two outcome domains can be directly tested rather than inferred across separate studies; (c) include a qualitative or mixed-methods strand such as reflective journals, stimulated-recall interviews or similar which are capable of capturing the attitudinal and critical-awareness dimensions of ICC that purely quantitative fluency measures cannot reach; and (d) incorporate a teacher-facing strand and, where feasible, more than one Vietnamese institution, so that findings are not confounded with the practices of a single program or the attitudes of a single teaching team.

6. Limitations of This Review

This review was conducted through iterative keyword-based web and database-indexed searches rather than a full, deduplicated export from a single bibliographic database (e.g., Scopus, Web of Science or ERIC) with independent dual-reviewer screening, as would be expected for a full systematic review submitted to a peer-reviewed journal. Screening and data extraction were additionally carried out by a single reviewer, which precludes the calculation of an inter-rater reliability statistic that a two-reviewer systematic review would normally report. Accordingly, the search cannot be considered exhaustive, and some relevant studies, particularly recent conference proceedings or non-English-language Vietnamese publications not indexed in the databases searched, may have been missed. For the small number of studies where eligibility and extraction relied on a detailed abstract rather than the full text (Section 3.3), the extracted characteristics are correspondingly less certain and were cross-checked against secondary sources only where such sources existed; this is a further, related limit on completeness. Similarly, the country/context values reported for a small number of studies in Table 1 could not be confirmed independently of the information stated in the source itself or the reviewing team's knowledge of the authors' institutional affiliations, and are flagged accordingly for verification against the original full texts. The 17-study evidence base is also small enough that the thematic clusters in Section 4, while internally consistent, should be read as an indicative map of the field rather than an exhaustive census of it.

7. Conclusion

This scoping review maps a rapidly growing but thematically fragmented literature on GenAI-mediated EFL speaking practice and AI-mediated intercultural communicative competence. Seventeen empirical and review studies published between 2022 and 2026, synthesized here alongside three foundational theoretical works and summarized in a comparative table of study characteristics, show that each strand inde-

pendently demonstrates promising learner outcomes. Particularly, GenAI-mediated speaking practice reliably improves fluency and reduces speaking anxiety over short interventions, while AI-mediated interaction shows preliminary but genuine promise for supporting intercultural communicative competence when paired with reflective, culturally responsive pedagogy. However, their integration, particularly within a longitudinal design and within underrepresented contexts such as Vietnam, remains an open and worthwhile direction for doctoral-level research. Grounding that future research explicitly in Sociocultural Theory, the Interaction Hypothesis and model of intercultural communicative competence of [1], as outlined in Section 2, offers one principled way to design a study capable of finally testing, rather than merely inferring, the connection between GenAI-mediated oral practice and intercultural communicative development.

Abbreviations

AI	Artificial Intelligence
EFL	English as a Foreign Language
ICC	Intercultural Communicative Competence
GenAI	Generative AI
WTC	Willingness to Communicate
ZPD	Zone of Proximal Development
ESL	English as Second Language
CILS	Cross-Cultural Intelligent Language Learning System
IUH	Industrial University of Ho Chi Minh City

Author Contributions

Do Ngoc Lan: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology

Nguyen Xuan Hong: Resources, Supervision, Validation, Writing – original draft, Writing – review & editing

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] Byram, M. (1997). Teaching and assessing intercultural communicative competence. *Multilingual Matters*.
- [2] Duong, T. T. V. & Suppasetsee, S. (2024). The effects of an artificial intelligence voice chatbot on improving Vietnamese undergraduate students' English-speaking skills. *International Journal of Learning, Teaching and Educational Research*, 23(3), 293–321. <https://doi.org/10.26803/ijlter.23.3.15>
- [3] Fathi, J., Rahimi, M., & Derakhshan, A. (2024). Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions. *System*, 121, Article 103254. <https://doi.org/10.1016/j.system.2024.103254>
- [4] Fathi, J., Rahimi, M., & Teo, T. (2025). Applying intelligent personal assistants to develop fluency and comprehensibility, and reduce accentedness in EFL learners: An empirical study of Google Assistant. *Language Teaching Research*. Advance online publication. <https://doi.org/10.1177/13621688251317786>
- [5] Klimova, B., & Chen, J. H. (2024). The impact of AI on enhancing students' intercultural communication competence at the university level: A review study. *Language Teaching Research Quarterly*, 43, 102–120. <https://doi.org/10.32038/ltrq.2024.43.06>
- [6] Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). New York: Academic Press.
- [7] Ngo, C.-L., Vu, H. & Duong, T. T. H. (2026). Translanguaging in generative AI prompting among Vietnamese EFL learners: An exploratory mixed-methods study. *The Language Learning Journal*. Advance online publication. <https://doi.org/10.1080/09571736.2025.2612180>
- [8] Nguyen, L. A. D., & Le, T. T. P. (2025). Exploring the effects of an AI chatbot on emotional engagement in English speaking lessons: Insights from Call Annie. *International Journal of AI in Language Education*, 2(2), 79–99. <https://doi.org/10.54855/ijaile.25225>
- [9] Sarwari, A. Q., Javed, M. N., Adnan, H. M., & Abdul Wahab, M. N. (2024). Assessment of the impacts of artificial intelligence (AI) on intercultural communication among postgraduate students in a multicultural university environment. *Scientific Reports*, 14, Article 13849. <https://doi.org/10.1038/s41598-024-63276-5>
- [10] Vo, T. K. A., & Nguyen, H. (2024). Generative artificial intelligence and ChatGPT in language learning: EFL students' perceptions of technology acceptance. *Journal of University Teaching and Learning Practice*, 21(6), 199–218. <https://doi.org/10.53761/fr1rkj58>
- [11] Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- [12] 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, Article 103533. <https://doi.org/10.1016/j.system.2024.103533>
- [13] Wang, C., Zou, B., Zhang, W., Du, Y., & Hu, W. (2026). Understanding EFL teachers' affective and cognitive responses to ChatGPT in higher education. *Humanities and Social Sciences Communications*, 13, Article 822. <https://doi.org/10.1057/s41599-026-07360-3>
- [14] Weng, Z., & Fu, Y. (2025). Generative AI in language education: Bridging divide and fostering inclusivity. *International Journal of Technology in Education*, 8(2), 395–420. <https://doi.org/10.46328/ijte.1056>

- [15] Wu, T.-T., Hapsari, I. P., & Huang, Y.-M. (2025). Effects of incorporating AI chatbots into think-pair-share activities on EFL speaking anxiety, language enjoyment, and speaking performance. *Computer Assisted Language Learning*, 1-39. <https://doi.org/10.1080/09588221.2025.2478271>
- [16] Xia, Y., Shin, S.-Y., & Kim, J.-C. (2024). Cross-Cultural Intelligent Language Learning System (CILS): Leveraging AI to facilitate language learning strategies in cross-cultural communication. *Applied Sciences*, 14(13), Article 5651. <https://doi.org/10.3390/app14135651>
- [17] Xueqing, X. & Li, R. (2024). Unraveling effects of AI chatbots on EFL learners' language skill development: A meta-analysis. *Asia-Pacific Education Researcher*. Advance online publication. <https://doi.org/10.1007/s40299-024-00853-2>
- [18] Yang, H., Kim, H., Lee, J. H., & Shin, D. (2022). Implementation of an AI chatbot as an English conversation partner in EFL speaking classes. *ReCALL*, 34(3), 327-343. <https://doi.org/10.1017/S0958344022000039>
- [19] 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, Article 100134. <https://doi.org/10.1016/j.caeai.2023.100134>
- [20] Zhai, C., Wibowo, S., & Li, L. D. (2024). Evaluating the AI dialogue system's intercultural, humorous, and empathetic dimensions in English language learning: A case study. *Computers and Education: Artificial Intelligence*, 7, Article 100262. <https://doi.org/10.1016/j.caeai.2024.100262>