

A Review of Research on Strategic Competence in AI-Mediated EFL Speaking Practice

Yuxian Zhang

Xi'an International Studies University, Xi'an, Shaanxi, 710128

Abstract: This review synthesizes empirical research (2021-2025) on AI-mediated EFL speaking practice through a language learning strategies lens. It examines how conversational AI—particularly voice chatbots, generative AI (GenAI), and automatic speech recognition (ASR)—reshapes learners' strategic behaviors and speaking outcomes. The synthesis identifies three research hotspots: task-based voice chatbots as conversation partners, GenAI's effects on willingness to communicate (WTC) and speaking anxiety, and ASR-based feedback for pronunciation development. A strategy-centered model is proposed, encompassing prompting, interaction management, repair, feedback uptake, and affect regulation. Pedagogical implications emphasize explicit strategy instruction for AI-mediated environments, assessment considerations, and teacher mediation. The review concludes with future research directions, including longitudinal designs, process data analysis, and comparative studies across interaction formats.

Keywords: L2 speaking; learning strategies; communication strategies; generative AI; chatbots

DOI:10.12417/3029-2344.25.12.027

1. Introduction

In recent years, English-speaking practice has evolved significantly, driven by advances in artificial intelligence (AI). Specifically, the integration of AI tools, such as conversational chatbots, generative AI, and automatic speech recognition (ASR), has introduced new ways for learners to practice and improve their speaking skills. This shift has expanded the range of speaking opportunities beyond the traditional classroom setting, providing learners with more frequent and less intimidating environments to engage in oral communication. AI-mediated speaking practice offers learners the opportunity to interact with AI systems that can simulate real conversations, offer instant feedback, and facilitate repeated practice. These features are particularly beneficial in English as a Foreign Language (EFL) context, where speaking opportunities are often limited, and anxiety can hinder learners' willingness to participate.

As AI-mediated speaking practice becomes more widespread, it is essential to explore the strategic behaviors that learners employ to maximize their speaking performance in these environments. The use of strategies such as planning, monitoring, self-repair, and reflection has long been a key factor in improving speaking skills. In AI-mediated contexts, learners not only use these strategies to manage their interactions with AI systems but also to overcome challenges related to feedback interpretation, interaction management, and affective factors such as anxiety and confidence.

Recent empirical studies have begun to examine how different forms of AI—chatbots, generative AI, and ASR-based systems—affect learners' speaking strategies and outcomes. These studies have explored how learners engage with AI systems, what strategies they employ to manage their speaking performance, and how these strategies influence their speaking ability. Additionally, AI systems provide new forms of feedback, which can play a critical role in learners' strategy use. While some studies have demonstrated the effectiveness of AI tools in enhancing learners' speaking skills, others have highlighted the complexities of feedback processing and the need for learners to develop specific strategies to make the most of these technologies.

This review aims to synthesize recent empirical research on AI-mediated speaking practice, focusing on the strategies that learners use to interact with AI systems and the outcomes of these strategies. By examining how different AI technologies influence learners' speaking strategies, this review provides insights into the potential of AI to support effective speaking practice and highlights the strategic competencies that learners need to develop to succeed in these new learning environments.

2. Conceptual frame

To systematically examine how learners engage with AI-mediated speaking environments, it is necessary to first clarify the key constructs that underpin strategic performance in oral communication. This conceptual framework brings together three interrelated perspectives: the nature of speaking performance and the role of strategic competence within it; the interactional and affective mechanisms through which strategies operate; and a strategy taxonomy adapted to the affordances and constraints of AI mediation. Together, these lenses provide a foundation for interpreting the empirical findings reviewed in subsequent sections.

2.1 Speaking performance and the role of strategic competence

Oral proficiency is commonly discussed in terms of complexity, accuracy, and fluency (CAF), reflecting trade-offs in attentional allocation during performance (Skehan, 1998). Learners cope with these constraints through strategies—both cognitive (e.g., reformulation, chunking) and interactional (e.g., clarification requests, confirmation checks). In classic models of communicative competence, strategic competence is the capacity to manage breakdowns and maintain communication under limitations. In AI-mediated speaking, strategic competence is not reduced; rather, it may be redistributed across new tasks, such as designing prompts, steering topic development, and evaluating AI feedback.

2.2 Interaction, feedback, and affect as strategy-relevant mechanisms

Interactionist perspectives emphasize that learning is supported when learners receive comprehensible input, produce output, and engage in negotiation of meaning and form (Long, 1996). In speaking tasks, feedback—whether from teachers, peers, or technology—creates opportunities for noticing gaps and adjusting production. Affect matters because anxiety can reduce risk-taking and willingness to speak, while perceived competence and supportive interaction can increase WTC (MacIntyre et al., 1998). These affective variables are themselves targets of self-regulatory strategies (e.g., reappraisal, planning, rehearsal), and recent AI studies increasingly treat emotions and WTC as measurable outcomes in their own right (Wang et al., 2024; Zhang et al., 2024).

2.3 Strategy perspectives suited to AI-mediated speaking

A strategy lens highlights observable and reportable actions learners take to improve performance. Oxford's (1990) classification (e.g., cognitive, metacognitive, affective, social strategies) remains useful, but AI-mediated speaking introduces hybrid strategic behaviors: learners must interact with a system that can respond instantly yet may produce errors, hallucinations, or pragmatically odd turns. Thus, strategy use includes (a) interaction management (turn-taking, follow-up questions), (b) repair and clarification (detecting and correcting misunderstandings), (c) feedback uptake (using ASR/AI feedback to revise), and (d) prompting and task framing (specifying roles, constraints, and evaluation criteria for the AI).

3. Research hotspots

Building on the conceptual framework outlined above, this section turns to recent empirical research that illuminates how AI-mediated environments shape learners' strategic engagement in speaking. The review identifies four interconnected hotspots where empirical work has concentrated over the past five years. Each hotspot reflects not only a particular type of AI tool—voice chatbots, generative AI conversational agents, automatic speech recognition systems—but also a distinct set of strategic demands and learner responses. Together, these areas represent the current landscape of inquiry into how technology-mediated conditions interact with learners' strategic behaviors, affective experiences, and speaking outcomes. The following sub-synthesize key findings from each hotspot, with particular attention to the strategic implications that emerge from the designs and results of these studies.

3.1 Voice chatbots as task-based conversation partners

One influential line of empirical work investigates whether voice chatbots can function as feasible interlocutors

for classroom speaking tasks. A large-scale study with Korean EFL learners implemented a task-based voice chatbot (“Ellie”) and evaluated task success, interaction quantity, and learner perceptions. Learners completed three speaking tasks in class, produced substantial conversational turns, and achieved high task success rates, indicating that well-designed chatbot tasks can elicit sustained oral interaction that may be difficult to obtain in typical EFL classrooms (Yang et al., 2022). From a strategy perspective, this environment encourages rehearsal, strategic planning, and interactional maintenance strategies because learners must keep the conversation moving without relying on human empathy cues.

Another research has emphasized that such “human–machine symbiosis” requires explicit pedagogical design: teachers must decide how responsibilities are divided between human instruction and AI interaction, and how risks (e.g., overreliance, misinformation) are managed. Although this research is primarily methodological and conceptual, it frames a practical agenda for strategy-oriented pedagogy: learners need guidance not only on “speaking more” but on speaking strategically in human–AI collaboration.

3.2 GenAI chatbots, WTC, and speaking anxiety

A second hotspot focuses on whether GenAI can change learners’ psychological readiness to speak. Wang et al. (2024) compared a control condition (teacher/classmate interaction) with two chatbot conditions: one text+voice GenAI chatbot, and another that combined text+voice interaction with a human-like avatar. Using multilevel modeling, the study found that the avatar-supported condition significantly improved WTC and self-perceived communicative competence and reduced speaking anxiety, though speaking test performance did not significantly differ across groups. The qualitative findings suggested that immersion and perceived emotional support contributed to improved affective experiences (Wang et al., 2024). It worth noticing that strategies are not only linguistic but also affect-regulatory. Learners may deploy coping strategies differently when the interlocutor is an AI system rather than a human classroom audience.

Similarly, Zhang et al. (2024) compared two AI tools—an AI chatbot and an intelligent tutoring system (ITS)—and found divergent WTC trajectories: learners in the chatbot group increased WTC while the ITS group’s WTC declined. The authors argued that the conversational nature of the chatbot better supported communicative engagement than the tutoring logic of the ITS (Zhang et al., 2024). This finding suggests that interactional affordances drive strategy deployment: a conversational agent invites turn-management, topic development, and negotiation strategies, whereas an ITS may trigger more accuracy-focused, test-like behaviors that do not necessarily build willingness to speak.

An experimental study on GenAI-empowered oral teaching reflects a move from broad “impact talk” to measurable outcomes. Although methodological details and outcome measures vary across studies, this line of inquiry signals an emerging domestic research frontier: linking GenAI-supported oral instruction to observable changes in learners’ participation, confidence, and communicative behavior.

3.3 ASR and automated feedback for pronunciation and speaking development

ASR is not new, but recent research consolidates its effectiveness and clarifies boundary conditions. A meta-analysis by Yang, H. et al. (2022) synthesized 15 studies (38 effect sizes) on ASR and L2 pronunciation, showing a medium overall effect and stronger outcomes when feedback is explicit. Effects were larger for segmental pronunciation than suprasegmentals, and peer-supported practice in ASR conditions produced larger benefits than practicing alone (Ngo et al., 2023). This points to a strategic implication: feedback does not automatically improve speaking; learners need strategies to interpret, prioritize, and apply feedback (e.g., selecting targets, practicing minimal pairs, monitoring transfer to spontaneous speech).

Furthermore, an experiment was conducted focusing on improving pitch prominence ability to address constraints in oral English phonological development. Although not necessarily AI-based, such studies inform what ASR feedback should emphasize and where learners may need strategy instruction (e.g., attention to prosodic

salience rather than only segmental accuracy). In addition, the discussions of AI's role in foreign language education have noted that technological feedback must be integrated with curricular goals and teacher mediation, rather than treated as a standalone “replacement” for pedagogy.

3.4 Strategy-relevant affective variables: anxiety, WTC, and self-efficacy

A final hotspot concerns the psychological architecture underlying speaking participation and persistence. Empirical studies have increasingly operationalized constructs such as oral self-efficacy and WTC. A scale was developed and validated for Chinese university students' English oral self-efficacy, offering measurement infrastructure for strategy research: if strategies are taught, changes should be traceable not only in performance but also in self-beliefs about speaking capability. Relatedly, L2 grit and WTC was investigated through chain mediation involving communicative confidence and autonomous motivation, suggesting that willingness to speak is sustained by motivational self-regulation processes that can be shaped through strategic interventions.

Although these studies are not always situated in AI contexts, they are directly relevant because AI-mediated speaking practice plausibly changes the conditions under which self-efficacy and WTC develop: frequency of successful interaction, perceived safety, and opportunities for autonomy are likely to influence learners' self-beliefs and hence their strategic engagement.

4.A strategy-centered synthesis: what “speaking strategies” look like with AI

Based on the reviewed hotspots, AI-mediated speaking invites a re-specification of strategy categories. Classic communication strategies still apply (Dörnyei & Scott, 1997), but they interact with new “AI-facing” behaviors:

Task framing and goal-setting strategies: learners specify roles (“act as an IELTS examiner”), constraints (time, vocabulary), and evaluation criteria, effectively shaping the communicative task before speaking begins. This resembles metacognitive planning (Oxford, 1990) but is realized through prompt/task design.

Interaction management strategies: follow-up questioning, topic extension, turn-keeping, and coherence building become central because chatbots can sustain dialogue indefinitely; learners must avoid shallow, short turns.

Repair and verification strategies: learners must detect AI misunderstandings or questionable responses and practice clarification or correction. This is an interactional competence outcome aligned with negotiation of meaning (Long, 1996).

Feedback uptake strategies: ASR and GenAI provide feedback at varying levels (phonological, lexical, pragmatic). Learners need strategies to select priorities and build deliberate practice cycles (Ngo et al., 2023).

Affect regulation strategies: reductions in speaking anxiety and increases in WTC observed in GenAI conditions (Wang et al., 2024) suggest that learners may strategically engage more when perceived threat is lower. However, strategy instruction is still needed to prevent avoidance patterns (e.g., overreliance on AI-generated language or minimal-effort responses).

This synthesis aligns with “role transformation” and ethical/competency requirements in the ChatGPT era. Teachers are not merely “users” of tools; they are designers of strategic learning environments, responsible for making strategy use visible, teachable, and assessable.

5. Pedagogical implications for an L2 learning strategy course

The preceding analysis underscores that AI-mediated speaking environments demand new forms of strategic competence. These demands have direct implications for L2 strategy instruction, which must evolve to address the unique affordances and challenges of human–AI interaction. This section outlines three pedagogical priorities: explicit strategy instruction tailored to AI contexts, reconceptualized assessment practices, and the mediating role of teachers in human–AI collaboration.

5.1 Strategy instruction for AI-mediated speaking

A strategy-oriented pedagogy should make learners' AI-mediated speaking cycles explicit: plan → interact →

notice → revise → reflect. In practical terms, this can be implemented through structured reflection logs and checklists aligned with Oxford's (1990) metacognitive and affective strategies, but adapted to AI contexts. For example, learners can be trained to (a) set two pronunciation targets per week using ASR data (Ngo et al., 2023), (b) rehearse interactional moves (clarification requests, reformulations), and (c) evaluate whether a GenAI dialogue genuinely increased communicative effort or simply produced comfortable but shallow exchanges.

5.2 Assessment and academic integrity

AI-mediated speaking raises assessment questions: if learners practice with GenAI, how should instructors interpret gains, and how can oral assessments maintain authenticity? It is emphasized that language education must anticipate ethical risks and design governance mechanisms, rather than adopting tools uncritically. For speaking, this implies that summative assessment should prioritize spontaneous interaction, strategic adaptability, and justification of choices (e.g., explaining how feedback was used), rather than only rehearsed scripts.

5.3 Teacher mediation in human–AI symbiosis

Human–AI symbiosis frameworks argue that AI's value emerges when teachers orchestrate roles, sequences, and safeguards. For strategy learning, teachers should model strategic prompting, demonstrate repair moves when the AI response is inappropriate, and teach learners to challenge outputs rather than accept them. This is consistent with the broader idea that effective technology integration is fundamentally pedagogical rather than purely technical.

6.Directions for future research

Recent empirical studies have established that AI-mediated speaking can meaningfully affect WTC and speaking anxiety (Wang et al., 2024; Zhang et al., 2024), and that ASR supports pronunciation learning under certain conditions (Ngo et al., 2023). The next stage for strategy research should emphasize: (a) process measures (dialogue logs, turn-level analyses, self-repair frequency), (b) longitudinal designs that trace strategy development over months rather than weeks, (c) comparative conditions (human–human vs human–AI vs hybrid), and (d) validated strategy instruments adapted to AI contexts, leveraging domestic measurement work such as oral self-efficacy scales. In short, the field needs to move from “Does AI help speaking?” toward “Which strategies mediate AI's effects, for whom, and under what designs?”

7.Conclusion

Empirical EFL speaking research shows a clear consolidation of AI-mediated speaking as a mainstream topic, with GenAI chatbots and ASR systems reshaping practice opportunities, emotional experiences, and interactional patterns. The strongest contribution of recent work is not simply that learners can “speak more,” but that they can engage in strategically regulated speaking under conditions that reduce anxiety and increase willingness to communicate (Wang et al., 2024; Zhang et al., 2024). For a learning-strategy perspective, this implies that AI-mediated speaking should be treated as a strategic ecology: learners plan, prompt, manage interaction, repair breakdowns, uptake feedback, and regulate affect. Effective pedagogy will therefore depend on explicit strategy instruction and careful task design, ensuring that AI becomes a catalyst for deeper communicative engagement rather than a shortcut around learning.

References:

- [1] Dörnyei,Z.,&Scott,M.L.(1997).Communication strategies in a second language:Definitions and taxonomies. *Language learning*,47(1),173-210.
- [2] 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).Academic Press.
- [3] MacIntyre,P.D.,Clément,R.,Dörnyei,Z.,&Noels,K.A.(1998).Conceptualizing willingness to communicate in a

- L2: A situational model of L2 confidence and affiliation. *The Modern Language Journal*, 82(4), 545–562.
- [4] Thi-Nhu Ngo, T., Hao-Jan Chen, H., & Kuo-Wei Lai, K. (2024). The effectiveness of automatic speech recognition in ESL/EFL pronunciation: A meta-analysis. *ReCALL*, 36(1), 4–21.
- [5] Oxford, R. L. (1990). *Language learning strategies: What every teacher should know*. Newbury House.
- [6] Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.
- [7] 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.
- [8] 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.
- [9] Zhang, D., Wu, J. G., & Fu, Z. (2024). From shy to fly: Facilitating EFL learners' willingness to communicate with an AI chatbot and an intelligent tutoring system. *System*, 127, 103501.
- [10] Du, Y., & Daniel, B. K. (2024). Transforming language education: A systematic review of AI-powered chatbots for English as a foreign language speaking practice. *Computers and Education: Artificial Intelligence*.
- [11] Klímová, B., & Ibna Seraj, P. M. (2023). The use of chatbots in university EFL settings: Research trends and pedagogical implications. *Frontiers in Psychology*.