THE ROLE OF AI-POWERED FEEDBACK IN DEVELOPING EFL LEARNERS' SPEAKING ACCURACY

Sobirjonova Maqsuda Saydimqul qizi
Specialist at the Center for the Implementation of
Webster University Educational Programs in Tashkent

Abstract. Artificial intelligence (AI) has expanded the possibilities for providing immediate, individualized and repeatable feedback to learners of English as a foreign language (EFL). This paper examines the role of AI-powered feedback in developing speaking accuracy, understood as accurate and intelligible pronunciation, grammatical control, appropriate lexical selection and effective self-correction during oral production. The analysis synthesizes evidence from international research on automatic speech recognition (ASR), AI speech-evaluation systems, speech-enabled corrective feedback and generative AI. The literature indicates that AI can strengthen the feedback cycle by making errors more noticeable, shortening the time between performance and correction, increasing opportunities for deliberate repetition, and supporting learner autonomy. Meta-analytic evidence shows a moderate positive effect of ASR on L2 pronunciation, while recent studies report gains in speaking performance and favorable learner perceptions. At the same time, the pedagogical value of AI feedback depends on recognition accuracy, the quality and specificity of explanations, learner proficiency, task design and teacher mediation. A hybrid model in which AI handles high-frequency diagnostic feedback and teachers validate, prioritize and contextualize corrections is therefore proposed as the most educationally sound approach.

Keywords: artificial intelligence, EFL, speaking accuracy, corrective feedback, automatic speech recognition, pronunciation, self-repair, generative AI.

1. INTRODUCTION

Speaking is one of the most cognitively demanding components of foreign-language competence because learners must retrieve vocabulary, organize grammar, control pronunciation and prosody, and respond to an interlocutor under severe time constraints. Accuracy in this context does not mean the absence of every non-native feature; rather, it concerns the extent to which the learner's spoken forms are sufficiently precise, stable and intelligible to communicate the intended meaning. Persistent pronunciation errors, unstable grammatical forms and inappropriate lexical choices can reduce comprehensibility even when the learner is willing to communicate. Consequently, high-quality corrective feedback is a central condition for developing accurate oral production.

Traditional teacher feedback remains pedagogically important, but in large classes it is difficult to provide immediate, individualized correction for every learner after every speaking attempt. AI-powered systems address this limitation by processing spoken input through automatic speech recognition, acoustic analysis, natural-language processing and, increasingly, large language models. The resulting feedback can identify mispronounced words, transcription mismatches, grammatical problems, lexical choices, pauses or hesitations and may also provide model answers, explanations and opportunities for repeated practice. Earlier ASR-based computer-assisted language learning already demonstrated that formative feedback can improve articulation and help learners monitor progress (1). More recent AI systems extend this function by integrating diagnostic scoring, personalized prompts and conversational interaction (2,3).

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.

The relevance of these tools has increased because speaking practice requires frequency and repetition. A learner who receives only occasional teacher correction may notice an error but have too few immediate opportunities to test an alternative form. AI feedback can reduce this temporal gap: the learner speaks, receives a response within seconds, compares the output with a target form, modifies the utterance and tries again. This repeated micro-cycle is particularly compatible with theories of noticing, pushed output and self-regulated learning. However, technological immediacy does not guarantee pedagogical validity. Recognition systems can misinterpret accented speech, scoring algorithms may privilege easily measurable features, and generative systems may produce feedback that sounds convincing but is linguistically or contextually inappropriate. Therefore, the central issue is not whether AI should replace teachers, but how AI-generated feedback can be used to improve speaking accuracy without narrowing the complexity of oral communication.

The aim of this paper is to analyze the mechanisms through which AI-powered feedback contributes to EFL learners' speaking accuracy, summarize current international evidence, identify limitations, and propose a practical hybrid model for classroom and independent speaking practice. The discussion is based on eight international studies and reviews published between 2015 and 2026, including research on ASR, AI speech-evaluation programs, speech-enabled corrective feedback and generative AI interaction (1–8).

2. MAIN PART

2.1. Speaking accuracy as a multidimensional construct

Speaking accuracy is multidimensional. Pronunciation accuracy includes segmental production (individual consonants and vowels), word stress, rhythm, connected speech and intonation. Grammatical accuracy involves the appropriate use of tense, agreement, word order, articles, prepositions and other morphosyntactic forms. Lexical accuracy concerns selecting words and collocations that express the intended meaning without inappropriate transfer from the first language. A fourth dimension is self-repair: the learner's ability to detect and correct an error during or immediately after speech. AI feedback can contribute to all four dimensions, but not equally. ASR systems are particularly strong at detecting segmental and word-level deviations, while generative AI can provide broader grammar and lexical explanations. Suprasegmental and discourse-level judgments remain more difficult and often require human interpretation (2,4).

The most useful AI feedback is therefore diagnostic rather than merely evaluative. A score such as "78/100" tells a learner that performance is imperfect but does not show what should change. Diagnostic feedback identifies the problematic form, explains why it is problematic, supplies a comprehensible target and encourages another attempt. This distinction is consistent with findings that explicit corrective feedback in ASR environments tends to be more effective than indirect transcription-only feedback (2).

Table 1. Main forms of AI-powered feedback for speaking accuracy

AI feedback mode	Primary target	Pedagogical contribution	Typical limitation
ASR transcription feedback	Pronunciation; word recognition	Makes mismatches visible; supports repeated pronunciation and self-repair	May confuse accented speech or noisy input

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.

AI feedback mode	Primary target	Pedagogical contribution	Typical limitation
Pronunciation scoring / acoustic feedback	Segments, stress, timing	Provides rapid, repeatable diagnosis and progress tracking	Scores may oversimplify intelligibility or prosody
NLP-based corrective feedback	Grammar and vocabulary	Highlights form errors and offers corrected versions	Accuracy depends on context and model design
Generative AI dialogue	Grammar, lexis, interaction	Creates adaptive speaking prompts, explanations and reformulations	Can hallucinate, over-correct or give inconsistent advice
Multimodal AI assessment	Integrated speaking performance	Combines audio, transcript, analytic criteria and feedback	Requires careful validation and human oversight

2.2. How AI feedback can improve speaking accuracy

The first mechanism is noticing. When the learner sees that a spoken word has been transcribed incorrectly or receives a targeted pronunciation flag, attention is directed to the gap between intended and perceived output. The second mechanism is comparison: the learner contrasts the original production with a model pronunciation, corrected sentence or AI-generated reformulation. Third, the learner performs self-repair by producing the revised form. Finally, immediate repetition helps stabilize the corrected pattern before attention shifts to another task. This sequence transforms feedback from information into action.

The speed of the cycle matters. In teacher-led lessons, feedback may be delayed because a teacher must distribute attention among many learners. AI can provide feedback after every attempt and thereby increase the density of practice. Wang and Young found that formative feedback in an ASR-based CALL system supported articulation, while summative feedback helped learners reflect on progress (1). The meta-analysis by Ngo, Chen and Lai synthesized 15 studies and 38 effect sizes and reported a medium overall effect of ASR on ESL/EFL pronunciation ($g = 0.69$); explicit corrective feedback and medium-to-long interventions were especially beneficial (2). This suggests that the value of AI is not simply automation but the possibility of sustained, explicit and repeatable corrective practice.

A third mechanism is learner autonomy. Because AI tools are available outside class, learners can practice privately and repeat difficult items without fear of taking classmates' time. Studies of AI speech-evaluation systems report that learners value convenience, access and the possibility of repeated practice (3,4). A fourth mechanism is affective support. Speech-enabled corrective feedback has been associated with improved speaking outcomes together with changes in confidence and speaking anxiety, indicating that an environment where errors can be rehearsed privately may encourage more experimentation (5). Nevertheless, autonomy should not become isolation; peer discussion and teacher interpretation remain important when feedback is ambiguous.

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.

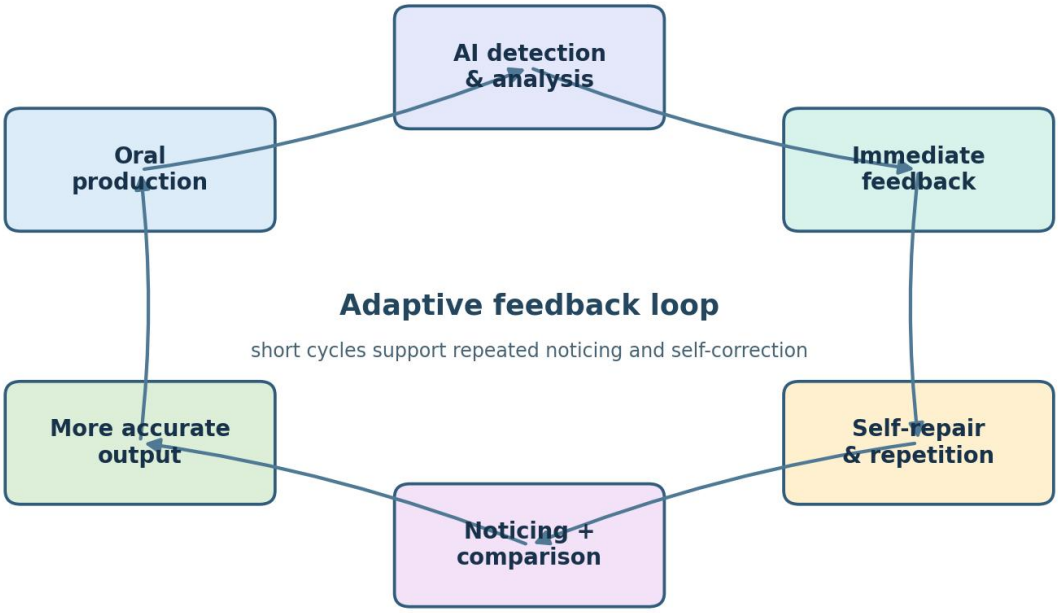

Figure 1. AI-powered corrective feedback cycle for speaking accuracy (conceptual model).

2.3. Evidence from international research

Recent international research provides converging, although not uniform, evidence that AI-assisted feedback can support oral development. Zou and colleagues investigated an AI speech-evaluation program with 40 EFL learners and reported significant pre–post improvement in speaking scores alongside generally positive perceptions of automated feedback (3). Sun compared an ASR-plus-peer-correction condition with conventional teacher-led instruction among 61 intermediate Chinese EFL learners; the study found benefits for pronunciation and broader speaking outcomes and highlighted the value of combining technology with collaborative correction (6). These studies indicate that AI feedback is more powerful when it is embedded in an instructional sequence rather than used as an isolated scoring device.

Large-sample perception data also help clarify what learners find useful. Zou et al. examined 366 learners across five universities using the EAP Talk speech-evaluation system. Participants valued the technology's convenience and its support for pronunciation, grammar-related accuracy, read-aloud and presentation practice, yet they also reported problems related to voice-recognition accuracy, limited explanations and the alignment of feedback with course needs (4). Such findings are important because inaccurate or opaque feedback can reduce trust. In other words, feedback must be both technically reliable and pedagogically interpretable.

Generative AI introduces a different form of support: interactive language use rather than only speech scoring. Sok and Shin found that EFL university students who completed oral tasks through ChatGPT-mediated interaction outperformed a control group on global speaking task performance, suggesting that conversational AI can expand opportunities for meaningful oral practice (7). At the assessment level, Li, Huang and Sheeran examined ChatGPT-4o as an AI peer assessor in EFL speaking classrooms and found potential for analytic scoring and actionable feedback, while also emphasizing the

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.

need to examine reliability across criteria and raters (8). Together, these studies show that generative AI may be useful not merely as a "corrector" but as an interlocutor, rehearsal partner and supplementary assessor.

Table 2. Selected international evidence on AI-assisted EFL speaking

Study	Participants / evidence	Technology	Key contribution	Implication for accuracy
Wang & Young (2015) (1)	18 adults + 16 school learners	ASR-based CALL	Formative + summative feedback supported articulation and reflection	Feedback form should match learner needs
Ngo et al. (2024) (2)	15 studies; 38 effect sizes	ASR meta-analysis	Overall g = 0.69; explicit feedback especially effective	Repeated explicit correction is preferable to transcription alone
Zou et al. (2023) (3)	40 EFL learners	AI speech evaluation	Significant pre–post speaking improvement; positive perceptions	Automated analytic feedback can support practice
Zou et al. (2024) (4)	366 learners; 5 universities	EAP Talk	Benefits plus concerns about recognition and feedback detail	Trust and explanation quality are essential
Shadiev et al. (2026) (5)	EFL learners in SECF intervention	Speech-to-text + automated corrective feedback	Targeted pronunciation, vocabulary and grammar correction	Integrated feedback can address multiple accuracy dimensions
Sun (2023) (6)	61 intermediate EFL learners	ASR + peer correction	Improved pronunciation and speaking outcomes	Technology is stronger when paired with peer interaction
Sok & Shin (2025) (7)	University EFL learners	ChatGPT oral interaction	AI groups improved global speaking task performance	GenAI can provide interactive rehearsal and reformulation

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.

Study	Participants / evidence	Technology	Key contribution	Implication for accuracy
Li et al. (2025) (8)	40 first-year English majors / peer-assessment design	ChatGPT-4o assessment	Examined scoring reliability and actionability of feedback	AI assessment should be calibrated with human judgment

2.4. From error detection to durable learning

For AI feedback to produce durable gains, learners need to do more than read or hear corrections. A productive sequence includes four actions: acknowledge the feedback, explain the error in simple terms, reformulate the utterance, and repeat the corrected form in a new context. This is particularly important for grammatical accuracy. If a learner says "She go to university every day" and receives "She goes…," repeating only the corrected sentence may produce short-term imitation. A stronger task asks the learner to generate two or three new third-person singular sentences and then use the target feature spontaneously in a brief response. Thus, AI feedback should be linked to retrieval and transfer.

Pronunciation requires a similar progression. Word-level correction may be appropriate for an unfamiliar consonant cluster or vowel contrast, but communicative accuracy also depends on stress, rhythm and comprehensibility in connected speech. The ASR meta-analysis found stronger effects for segmental pronunciation than for suprasegmental features (2). Teachers should therefore use AI for high-frequency micro-diagnosis while reserving classroom time for broader issues such as prominence, phrasing, pragmatic appropriacy and intelligibility across complete utterances.

Generative AI can additionally transform feedback into dialogic prompts. Instead of giving a single correction, it can ask the learner to explain what changed, contrast two forms, or repeat the message in a more precise way. This supports metalinguistic awareness. However, learner prompts should specify the feedback target. A broad command such as "Correct my English" can lead to excessive correction and interrupt fluency. More focused instructions—for example, "After I answer, identify only the two errors that most reduce grammatical accuracy and ask me to repeat the corrected sentences"—are more likely to preserve communicative flow.

2.5. Limitations and risks of AI-powered speaking feedback

Despite its advantages, AI feedback has important limitations. First, ASR accuracy is not identical to pronunciation accuracy. A correctly pronounced word can be mistranscribed because of background noise, microphone quality or accent variation, while an incorrectly pronounced form may still be recognized from context. Learners should therefore treat recognition mismatches as diagnostic signals rather than absolute judgments. Second, automated scores can create a false sense of precision. A numerical pronunciation or fluency score is useful for tracking change within the same system, but it should not be interpreted as a complete measure of communicative competence.

Third, generative AI may over-correct. Natural spoken language contains repetition, ellipsis, discourse markers and informal constructions that are acceptable in conversation but may be labeled as suboptimal by a model trained on edited text. Fourth, AI can provide inconsistent explanations or recommend a form that is grammatically possible but pragmatically unsuitable. These risks justify human validation, especially for advanced learners, high-stakes assessment and culturally sensitive communication. Fifth, data privacy matters because many systems process voice recordings in the cloud; institutions should clarify what is recorded, stored and reused.

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.

Finally, there is a pedagogical risk of accuracy becoming the only goal. Fluency, interactional competence, willingness to communicate and strategic competence are equally important. If learners stop after every minor error, speech may become hesitant and unnatural. Feedback should therefore be prioritized: errors that block meaning, recur frequently or represent current learning goals deserve immediate attention; minor slips can be delayed or ignored during fluency-focused tasks.

2.6. A hybrid AI–teacher model for EFL speaking practice

The reviewed evidence supports a hybrid model in which AI provides frequent, low-latency diagnostic feedback while teachers retain responsibility for pedagogical decisions. A practical lesson cycle begins with a communicative speaking task. The learner then completes one or two AI-assisted repetitions focused on a limited accuracy target such as final consonants, verb tense or article use. AI identifies patterns and proposes corrections; the learner self-repairs and repeats. The teacher subsequently checks a small sample of the feedback, resolves ambiguous cases and identifies which errors deserve class-level attention. Learners then transfer the corrected forms into a new peer or task-based activity. This approach avoids both extremes: total dependence on teacher correction and uncritical dependence on algorithms.

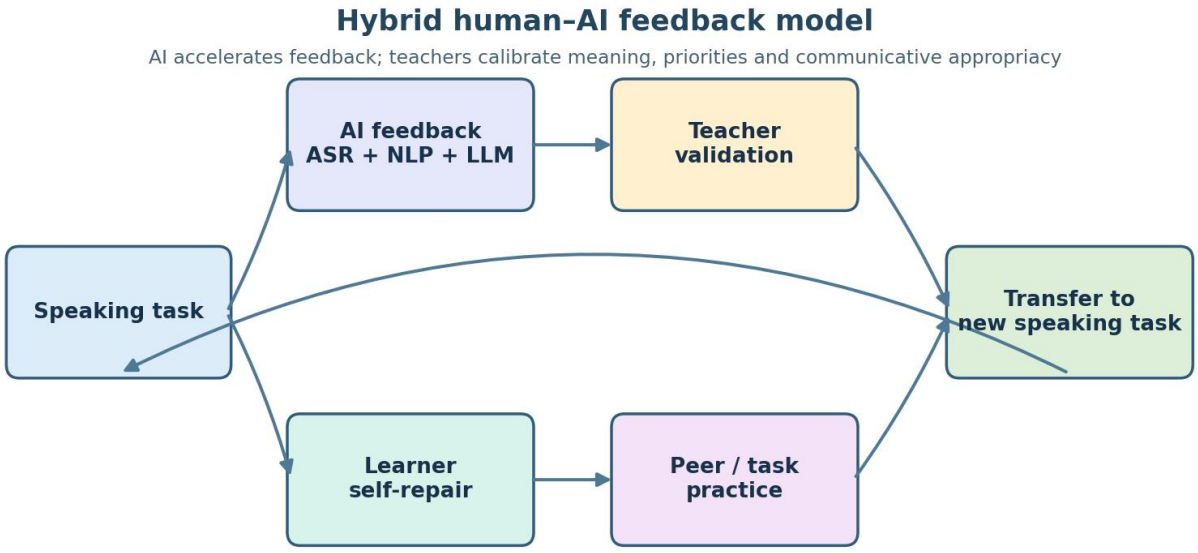

Figure 2. Proposed hybrid model integrating AI feedback, self-repair, teacher validation and transfer.

Table 3. Suggested indicators for monitoring speaking accuracy

Indicator	What it shows	How AI can support measurement
Pronunciation intelligibility	How easily target words/utterances are understood	ASR mismatch patterns + model comparison, verified by teacher listening

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.

Indicator	What it shows	How AI can support measurement
Grammatical error density	Frequency of recurrent morphosyntactic errors	Transcript-based detection of tense, agreement, article and word-order errors
Lexical appropriacy	Precision and suitability of word choice	Context-sensitive reformulations and collocation suggestions
Self-repair rate	Ability to notice and correct one's own output	Comparison of first attempt and revised attempt
Transfer accuracy	Use of corrected forms in new tasks	Delayed speaking prompts using the same target in a different context

A simple monitoring protocol can combine these indicators across time. The objective is not to maximize every score after each session, but to identify stable trends. For instance, a learner may show fewer grammatical errors in rehearsed answers but fail to transfer the target feature to spontaneous speech. In that case, the next instructional step should be contextual variation rather than additional explanation. Likewise, a learner whose ASR score improves while teacher-rated intelligibility remains unchanged may need work on rhythm or stress rather than individual sounds. Data are useful only when interpreted pedagogically.

3. CONCLUSION

AI-powered feedback can make an important contribution to EFL speaking accuracy by increasing the frequency, immediacy and individualization of corrective practice. Across ASR, AI speech-evaluation systems, speech-enabled corrective feedback and generative AI, the strongest pedagogical pathway is a repeated cycle of oral production, error detection, explicit feedback, comparison, self-repair and transfer. Meta-analytic and empirical studies indicate positive effects on pronunciation and broader speaking performance, while learners commonly value the convenience and autonomy of AI-supported practice (2–7).

Nevertheless, AI feedback should not be treated as an infallible substitute for expert judgment. Recognition errors, accent bias, limited suprasegmental analysis, over-correction and inconsistent generative explanations can reduce validity. The most effective implementation is therefore hybrid: AI manages high-frequency diagnostic feedback and rehearsal, learners actively self-repair, and teachers validate difficult cases, prioritize errors and preserve the communicative purpose of speaking. In this model, technological efficiency and pedagogical expertise reinforce each other rather than compete.

For EFL programs, the practical implication is clear: speaking technologies should be selected not only for attractive interfaces or automated scores, but for the quality of their feedback loop. Systems are most valuable when they help learners understand what is inaccurate, why it matters, how to improve it, and whether the corrected form can be used successfully in a new communicative context. Future research should continue to compare AI and human feedback across proficiency levels, accents and task types, with special attention to long-term transfer, fairness, privacy and the reliability of multimodal generative AI assessment.

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.

REFERENCES
1. Wang, Y.-H., & Young, S. S.-C. (2015). Effectiveness of feedback for enhancing English pronunciation in an ASR-based CALL system. Journal of Computer Assisted Learning, 31(6), 493–504. https://doi.org/10.1111/jcal.12079
2. Ngo, T. T.-N., Chen, H. H.-J., & Lai, K. K.-W. (2024). The effectiveness of automatic speech recognition in ESL/EFL pronunciation: A meta-analysis. ReCALL, 36(1), 4–21. https://doi.org/10.1017/S0958344023000113
3. Zou, B., Du, Y., Wang, Z., Chen, J., & Zhang, W. (2023). An investigation into artificial intelligence speech evaluation programs with automatic feedback for developing EFL learners' speaking skills. SAGE Open, 13(3), 1–9. https://doi.org/10.1177/21582440231193818
4. Zou, B., Liviero, S., Ma, Q., Zhang, W., Du, Y., & Xing, P. (2024). Exploring EFL learners' perceived promise and limitations of using an artificial intelligence speech evaluation system for speaking practice. System, 126, 103497. https://doi.org/10.1016/j.system.2024.103497
5. Shadiev, R., Feng, Y., Zhussupova, R., & Altinay, F. (2026). Effects of speech-enabled corrective feedback technology on EFL speaking skills, anxiety and confidence. Computer Assisted Language Learning, 39(4), 878–914. https://doi.org/10.1080/09588221.2024.2430761
6. Sun, W. (2023). The impact of automatic speech recognition technology on second language pronunciation and speaking skills of EFL learners: A mixed methods investigation. Frontiers in Psychology, 14, 1210187. https://doi.org/10.3389/fpsyg.2023.1210187
7. Sok, S., & Shin, H. W. (2025). Do interactions with ChatGPT influence L2 learners' oral speaking ability, summarization ability, and perceptions of generative AI tasks? TESOL Quarterly, 59(S1), S19–S51. https://doi.org/10.1002/tesq.70001
8. Li, J., Huang, J., & Sheeran, T. (2025). ChatGPT4o as an AI peer assessor in EFL speaking classrooms: Examining scoring reliability and feedback effectiveness. SAGE Open, 15(3). https://doi.org/10.1177/21582440251369938

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.