English Language & Literature International Conference
Vol. 8 No. 1
https://jurnal.unimus.ac.id/index.php/ELLIC/index
p-ISSN: 2579-7549
e-ISSN: 2579-7263

AI for All Voices: A Scoping Review on Inclusive Pronunciation and Speaking AI Tools

Ervinda Dwi Meidyana[1],
Fayza Achsina Salsabila[2],
Syahzanan Harris[3]
Universitas Airlangga
Surabaya, Indonesia
ervinda.dwi.meidyana-2024@fib.unair.ac.id

ABSTRACT

English as a Foreign Language (EFL) learners often face challenges in developing speaking skills, including limited practice opportunities, pronunciation difficulties, and speaking anxiety. As artificial intelligence (AI) tools become more integrated into language education, they offer new ways to support speaking and pronunciation practice. However, multilingual and non-native learners remain underrepresented in developing and designing many of these tools. This scoping review examines how AI-assisted pronunciation technologies address inclusivity in EFL contexts, more specifically for multilingual learners. A total of 15 peer-reviewed articles published between 2021 and 2025 were retrieved from Scopus, ScienceDirect, and Web of Science. The findings show that AI tools generally enhance pronunciation accuracy, fluency, and learner motivation while offering personalized and flexible learning environments. Several tools also support affective inclusivity by reducing anxiety during speaking tasks. Yet, inclusivity is often a by-product rather than an intentional design goal. Persistent challenges include accent bias and a reliance on native-speaker training data, which can disadvantage multilingual learners. This review underscores the need for more inclusive and culturally responsive AI systems and provides a foundation for future research in equitable EFL pronunciation support.

Keywords: Artificial Intelligence; EFL learners; multilingual learners; inclusive education; pronunciation training

p-ISSN: 2579-7549
e-ISSN: 2579-7263

INTRODUCTION

Artificial Intelligence (AI) is a machine's capability to perform cognitive tasks. AI tools are useful because they can help with reasoning, learning, and decision-making by collecting data from datasets and analyzing it in a split second (Feng et al., 2024; *Going Digital: Making the Transformation Work for Growth and Well-Being Going Digital*, 2017; Király, 2024). AI acquires its data from uploaded data sets in the system. The accessibility and quick response of AI significantly increase its role across many sectors, such as healthcare, finance, and even education. In education, AI technologies have been used a lot to help with learning processes as they enable automated assessment and support personalized instruction (Luckyardi et al., 2024). In language education, AI helps to assist both learners and teachers with its technologies, like speech recognition, chatbots, and intelligent tutoring systems (Du & Daniel, 2024; Király, 2024).

In language learning, using AI can be helpful and beneficial. It has been used a lot to help improve language skills such as writing, reading, speaking, and listening. Speaking is considered a difficult skill to master by second and foreign language learners. Not only does it need high language proficiency, but practicing speaking may also cause anxiety. One of the tools that can be used in speaking and pronunciation training is the Automated Speech Recognition (ASR). It works by recording learners' speech and providing targeted feedback. With this tool, learners can practice their speaking skills and receive immediate feedback on their pronunciation, fluency, intonation, and speech performance. It often uses acoustic feature extraction (e.g., Mel-frequency cepstral coefficients) and language modeling to analyze and transcribe spoken input (Do et al., 2024). More advanced systems include Goodness of Pronunciation (GOP) scoring and phoneme-level alignment, which are prominent in detecting mispronunciations and providing learners with targeted feedback. AI tools with automated speech recognition systems (ASR) can simulate real-time conversations and interactions, which are helpful for speaking practice. EFL learners with limited chances to interact with target language-speaking partners could also use AI tools.

Despite their potential to improve pronunciation and speaking in the second language learning context, ASR tools are commonly developed using datasets composed primarily of native English speakers (Feng et al., 2024). For example, many ASRs often take their data from standard American or British varieties. As a result, it creates bias when recognizing speech from a particular age, gender, fluency, and accent (Feng et al., 2024). Language speakers come from various

p-ISSN: 2579-7549
e-ISSN: 2579-7263

backgrounds, and if some of their speech cannot be recorded and analyzed in the system, second language learners or people with disabilities would not receive adequate feedback and have limited sources of improvement (Feng et al., 2024). These issues are particularly pressing in the global use of English as a lingua franca, where learners across diverse contexts use English for international communication. When AI tools are not inclusive in design, they risk marginalizing learners who do not conform to standardized or native-speaking norms.

Language learners already face multiple challenges when developing speaking skills. These include limited exposure to spoken English outside the classroom for foreign language learners, pronunciation difficulties from L1-L2 phonological differences, and affective cognitive barriers such as speaking anxiety, self-consciousness, and fear of making mistakes or being judged (Almineeai et al., 2025; Ding & Zhu, 2025; Jing, 2024; Liu et al., 2025; Pituxcoosuvarn et al., 2024). These obstacles are even more visible for learners from marginalized or underrepresented linguistic backgrounds. When AI tools are not developed to recognize non-standard or accented speech, they may misinterpret correct utterances, give inaccurate feedback, or penalize learners based on pronunciation variance rather than intelligibility (Feng et al., 2024). It can discourage learners' confidence in speaking English and promote linguistic discrimination.

Many AI-based tools' technical and design requirements can limit accessibility. Tools that require high-speed internet, consistent power supply, or advanced devices may be out of reach for learners in low-resource environments (Mingyan et al., 2025). In addition, AI tools with complex design interfaces would not be accessible to all learners. Learners unfamiliar with these tools or who have a particular disability would need further assistance to use them. These design gaps highlight the need for context-aware, mobile-friendly, and low-bandwidth solutions that make AI tools accessible to many people.

Inclusivity in the context of AI-assisted language learning refers to whether or not the tool is able to accommodate and support a broad spectrum of learner identities. Learners come from varied linguistic backgrounds and have specific speech styles, proficiency levels, learning needs, and cognitive or affective profiles (Feng et al., 2024). Inclusive AI systems should be able to recognize diverse accents, provide adaptive and meaningful feedback, and personalize

p-ISSN: 2579-7549
e-ISSN: 2579-7263

learning based on each learner's profile. It also involves offering flexibility in how, when, and where learning occurs across platforms and devices, including mobile and offline use, to reduce participation barriers (Jing, 2024; Mingyan et al., 2025).

Affective inclusivity is equally important. Tools that offer private, low-pressure environments for speaking practice can help reduce speaking anxiety and boost learner confidence (Almineeai et al., 2025; Ding & Zhu, 2025; Pituxcoosuvarn et al., 2024). It is critical for those who may feel judged or anxious in traditional speaking contexts. Game-based environments, chatbots, and self-paced tasks have shown promise in creating emotionally safe learning spaces where learners can build speaking skills without fear of embarrassment (Ding & Zhu, 2025; Mingyan et al., 2025; Pituxcoosuvarn et al., 2024).

Despite these advantages, most current AI tools do not explicitly implement inclusivity as a design principle (Du & Daniel, 2024). Instead, inclusivity is often an incidental benefit, emerging from user experience rather than intentional development (Király, 2024).

The implementation of AI tools has its benefits and challenges. Its advancement has kept growing in recent years. This scoping review aims to map the research of AI tools to improve pronunciation and speaking to learners with diverse backgrounds. By synthesizing recent studies from 2021 to 2025, this review maps the types of AI tools and their implementations in speaking. It also identifies patterns, gaps, and areas for future development. The research question proposed in this study is:

"*To what extent do AI-assisted pronunciation and speaking tools support inclusive language education?*"

METHOD

This scoping review followed Arksey and O'Malley's (2005) scoping review framework. It outlines a flexible approach for synthesizing research across broad and complex topics. The goal of this review was to explore the extent to which AI-assisted pronunciation and speaking tools support inclusive language education. There were four key stages that guided the review process: (1) identifying the research question, (2) selecting relevant studies, (3) charting the data, and (4) summarizing and reporting the results.

p-ISSN: 2579-7549
e-ISSN: 2579-7263

1. *Scoping Review Research Question*

The research question proposed in this study served as the guide to choosing the flow of this scoping review. The research question guided the search strategy and the screening process of this scoping review. Broad conceptual definitions were used for core terms such as "inclusivity," "pronunciation," and "AI-assisted speaking" to allow for a wide-ranging and comprehensive examination of the literature.

This inclusive approach enabled the review to capture studies involving various types of AI technologies (e.g., automated speech recognition, chatbots, mobile apps, LLMs) and diverse learner populations, especially multilingual, EFL, and non-native English speakers. As the review progressed, the inclusion criteria were refined based on the emerging patterns and content of the literature. It allowed for both flexibility and rigor in addressing the nuances of the field while maintaining consistency with the scoping review framework.

2. *Scoping Review Research Question*

Relevant keyword searches based on the conceptualized key terms were retrieved from Scopus, ScienceDirect, and Web of Science. These databases are chosen because they are the most widely used databases of reputable and peer-reviewed scientific publications. Initially, Eric was also a database used to search for related articles. However, only one result showed up, and it was a duplicate from Scopus. The articles chosen were from 2020-2025. Choosing this time range ensures that the AI tools used are the most recent and advanced. The search terms used for this review were as follows:

a. Scopus: (TITLE-ABS-KEY ("artificial intelligence" OR "AI" OR "machine learning" OR "speech recognition" OR chatbot OR "intelligent tutor")) AND (TITLE-ABS-KEY (pronunciation OR "speaking skills" OR "oral fluency" OR "accent training")) AND (TITLE-ABS-KEY (inclusive OR inclusivity OR "inclusive education" OR accessibility OR "universal design" OR "diverse learners" OR "language equity")) AND (PUBYEAR > 2020 AND PUBYEAR < 2025)

p-ISSN: 2579-7549
e-ISSN: 2579-7263

b. Web of Science: (TS = ("artificial intelligence" OR "AI" OR "machine learning" OR "speech recognition" OR chatbot OR "intelligent tutor")) AND (TS = (pronunciation OR "speaking skills" OR "oral fluency" OR "accent training")) AND (TS= (inclusive OR inclusivity OR "inclusive education" OR accessibility OR "universal design" OR "diverse learners" OR "language equity")) AND (PY= (2021-2024))
c. Science Direct: ("artificial intelligence" OR "AI" OR "speech recognition") AND (pronunciation OR "speaking skills" OR "accent training") AND (inclusive OR accessibility OR "diverse learners")

Table 1

Summary of the Search

Databases	Limiters	Search Results
Scopus	2020-2025	7
Web of Science	2020-2025	4
ScienceDirect	2020-2025	208

3. Study Selection: Inclusion and Exclusion Criteria

As shown in Table 1, a total of 219 references were initially retrieved through keyword-based searches across multiple academic databases, including Scopus, Web of Science, ERIC, and ScienceDirect. The search results were exported and organized using Mendeley reference management software, which was also used to identify and remove duplicate entries.

Following the removal of duplicates, the remaining articles underwent a pilot screening of titles and abstracts guided by the review's inclusion and exclusion criteria. The inclusion criteria were designed to ensure that selected studies:

- focused on the use of AI-assisted tools for pronunciation or speaking practice
- targeted language learners, particularly EFL, ESL, or multilingual speakers
- addressed or mentioned aspects of inclusivity, such as accessibility, learner diversity, or affective support (e.g., anxiety reduction)

p-ISSN: 2579-7549
e-ISSN: 2579-7263

- were published in peer-reviewed journals or conference proceedings
- were written in English and published between 2021 and 2024
- were open access

Studies were excluded if they:

- did not involve AI technologies or speaking/pronunciation tasks
- were duplicate records or inaccessible full texts

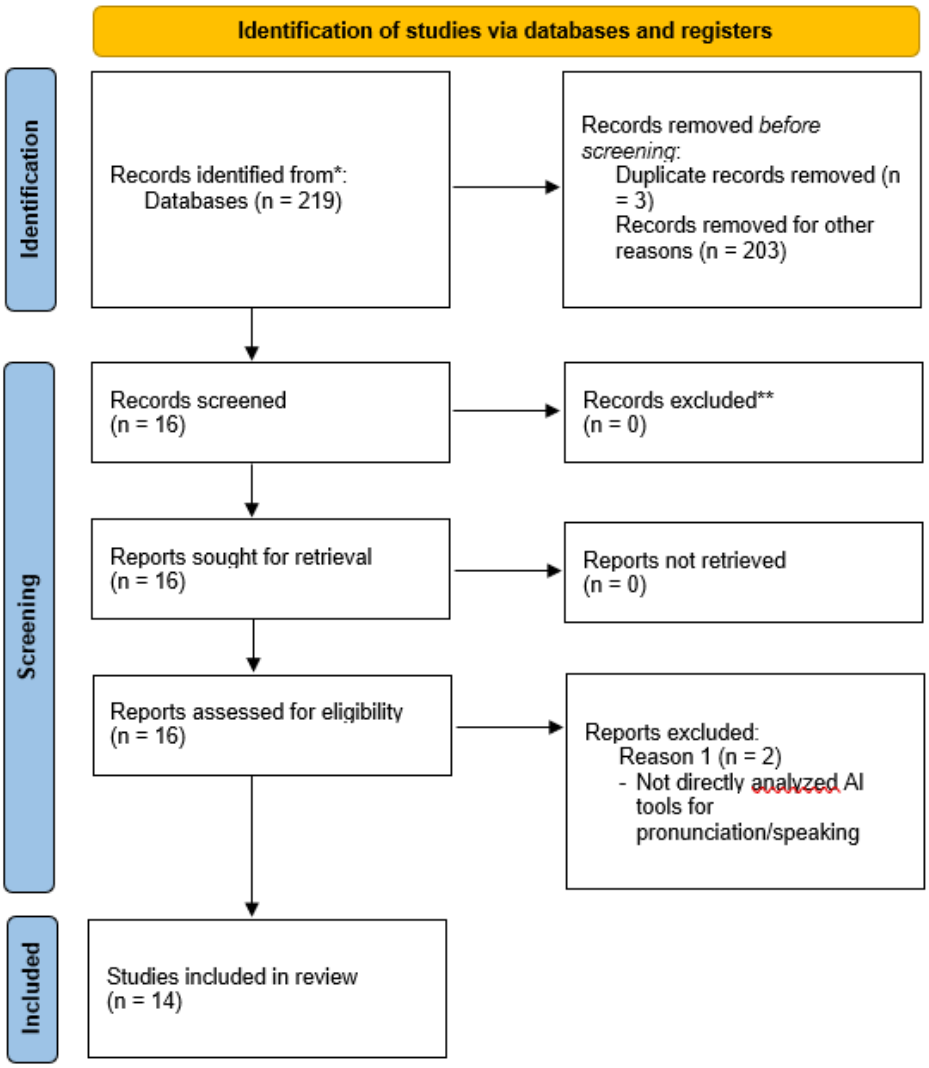
Full-text screening was then conducted on the articles that passed the initial screening to assess their eligibility based on the same criteria. Studies that aligned with the research question were included for final analysis.

After applying the criteria, 15 articles were selected to be included in the current review. The following flow chart was adapted from The PRISMA Group (PRISMA 2020 Flow Diagram — PRISMA Statement, n.d.).

p-ISSN: 2579-7549
e-ISSN: 2579-7263

Graphic 1:

PRISMA flow diagram of the literature searches and study selection process



FINDINGS AND DISCUSSION

A total of 15 studies were reviewed to investigate the extent to which AI-assisted tools support inclusivity in pronunciation and speaking instruction. These studies were systematically analyzed and summarized in Table 3, which includes each article's title, author(s), publication year, research aim and methodology, key findings, and relevance to the research question.

p-ISSN: 2579-7549
e-ISSN: 2579-7263

1. *AI Tools' Impact on Pronunciation and Speaking Performance*

The majority of the studies reported positive impacts of AI tools on learners' speaking fluency, pronunciation accuracy, motivation, and self-confidence. AI technologies such as EAP Talk, AI-powered mobile applications, and interactive chatbots significantly enhanced learner engagement and speaking outcomes (Du & Daniel, 2024; Liu et al., 2025; Mingyan et al., 2025). AI tools accommodate controlled speaking tasks, such as reading aloud, and non-controlled speaking tasks, such as presentations. After recording the speech, they provide immediate and targeted feedback that the learners can use to improve their speaking performance. These tools are particularly beneficial for learners in EFL contexts who may have limited access to native speakers or face-to-face interaction.

Several studies emphasized the practical benefits of AI-assisted tools in speaking. With the correct implementation, these tools allow learners to practice speaking independently without the pressure of direct human interaction. The finding of the study implies that AI tools reduce speaking anxiety experienced by learners and boost their confidence to speak (Almineeai et al., 2025; Pituxcoosuvarn et al., 2024). It is a critical aspect of cognitive or affective inclusivity. It helps learners who experience embarrassment, social anxiety, or self-consciousness when speaking.

2. *The Role of Automated Speech Recognition (ASR)*

The role of Automated speech recognition (ASR) was prominent for pronunciation or speaking training. It was used to evaluate pronunciation and speaking performance (Jing, 2024) and provide supportive and immediate feedback for learners through its mechanism (Király, 2024; Qassrawi et al., 2024). However, there are some limitations of ASR. The effectiveness of ASR tools relies on the datasets of the program (Feng et al., 2024). These datasets often come from native speakers or standardized English, such as American and British English. With a limited dataset, AI tools may not be able to recognize and process speech from non-native speakers or people with regional accents. The limitation may cause bias and force second and foreign language learners to achieve native-like fluency. However, this limitation can be addressed by improving the system by making some adjustments in the dataset, such as adding more regional accents as comparison material.

3. *Personalization, Accessibility, and Affective Support*

Several studies explored the personalization and accessibility aspects offered by AI tools in the context of language learning. The

p-ISSN: 2579-7549
e-ISSN: 2579-7263

personalization of AI tools takes into account whether they can provide specific feedback, be accessed at their own pace, and be accessible regardless of time and place. ELSA Speak and game-based systems such as Taboo Talks are some of the studied AI tools that promote accessibility and personalization (Luckyardi et al., 2024; Pituxcoosuvarn et al., 2024). Outside of classroom activity, students can use these tools with self-study or as an after-class practice tool. Teachers could also use these AI tools in classrooms to improve language learning.

Many AI tools incorporated game-based designs and chatbot interactions to promote learner engagement and reduce the stress commonly associated with oral language production. Speaking anxiety is a common problem experienced by many language learners. These approaches fostered a more inclusive affective environment where learners felt more comfortable experimenting with language (Almineeai et al., 2025).

4. Limitations in Inclusivity and Ethical Concerns

From the previous studies, the findings showed inclusivity as an accidental gain. It was not the actual goal of the study, but the tools are beneficial in promoting inclusivity. Some studies raised concerns about systemic bias within AI tools. Feng et al. (2024) studied the bias in ASR. It revealed that ASR systems demonstrated significant bias against nativeness, accent, gender, and age groups. This bias results in unequal learning outcomes. Some groups' speech is easier to recognize than other groups. Similarly, Du & Daniel (2024) pointed out that current chatbot systems often lack personalized pronunciation feedback, making them less effective for learners with non-standard speech patterns.

Another limitation is the continued focus on native-like fluency as the standard for success in many AI tools. Native-like fluency could marginalize learners while learners have diverse linguistic identities (Luckyardi et al., 2024). Language learners have their own language quirks and uniqueness. They should not remove this identity just to fit the standardized native-like speaking.

More advancements and improvisation in the system can help with these limitations. For example, Sahed et al. (2025) introduced a lip-reading dataset for Bengali pronunciation mapping, while Shi et al. (2024) investigated phoneme recognition through direct articulatory observation. These tools represent a significant step forward in inclusive design, as they offer alternatives for learners with hearing or

p-ISSN: 2579-7549
e-ISSN: 2579-7263

speech impairments and could pave the way for more equitable pronunciation training technologies.

5. Observed Trends and Patterns

There are some notable patterns and trends across the reviewed studies. AI tools prioritized accessibility and learner autonomy in their application. Many AI tools are designed for use outside the classroom and on mobile devices with flexible schedules. They support learners from geographically or economically diverse backgrounds so they can use AI tools equally (Jing, 2024; Mingyan et al., 2025). Aside from that, affective inclusivity is also a trend observed in the studies. Affective inclusivity, particularly the reduction of speaking anxiety, was a recurring theme. Studies emphasized the importance of creating emotionally safe learning spaces through private, self-paced, AI-supported practice (Ding & Zhu, 2025; Pituxcoosuvarn et al., 2024).

Bias and underrepresentation in training datasets continue to be significant challenges. AI tools risk reinforcing inequities without intentionally including non-native or regionally varied speech in model development (Feng et al., 2024). A promising shift toward multimodality was observed, with tools exploring visual and articulatory input to enhance feedback and accessibility, especially for learners with disabilities (Sahed et al., 2025; Shi et al., 2024).

The findings suggest that AI-assisted speaking and pronunciation tools partially support inclusive language education. Many tools promote flexibility, personalization, and affective engagement, which are critical for inclusive learning. However, inclusivity is often a secondary outcome rather than a guiding design principle. Only a few studies explicitly addressed inclusive practice, such as accommodating learners with speech impairment (Sahed et al., 2025). Moreover, challenges such as accent bias, limited feedback for non-controlled speech, and the dominance of native-speaker norms still limit th e full implementation of inclusive AI-assisted language learning.

CONCLUSION

This scoping review examined the extent of AI-assisted pronunciation and speaking tools to support inclusive language education. It draws from fifteen recent studies published between 2021 and 2025. The review identified both AI's promising potential

p-ISSN: 2579-7549
e-ISSN: 2579-7263

and current limitations in addressing the diverse needs of language learners, particularly in EFL and multilingual contexts.

The findings reveal that AI technologies, such as automated speech recognition (ASR), chatbots, mobile applications, and large language models (LLMs), contribute positively to learners' speaking fluency, pronunciation accuracy, motivation, and engagement. Many tools offer real-time feedback, flexible access, and personalized learning paths, which can support learners who face limited opportunities for speaking practice or who experience speaking anxiety. These features reflect progress toward inclusivity, especially in affective and motivational dimensions.

From the findings of the studies, not all AI tools are inclusive, some still use dataset from native speakers. This could lead to bias against non-native speakers. The limited dataset of certain regional accent could also limit the inclusivity of these tools. Without the proper adjustments, AI tools would not be able to recognize regional accents and provide feedback to language learners. Moreover, while a few innovative studies explored multimodal approaches such as lip-reading or articulatory analysis to support learners with special needs, these developments remain underrepresented in mainstream tools.

Teachers should take into account the accessibility and responsiveness of AI tools to accommodate diverse learners' profiles. They could integrate AI as a complementary tool that enhance learners' speaking experience without making them feel anxious. This practice could help boost students' confidence and improve their overall speaking performance. For researchers, future studies should further explore how inclusive principles, such as Universal Design for Learning (UDL), can be applied to AI in language learning.

In conclusion, AI-assisted speaking and pronunciation tools greatly promise to enhance language learning in inclusive ways. However, their full potential will only emerge when inclusivity moves from an afterthought to a guiding value in how these technologies are conceived, built, and implemented.

REFERENCES

Almineeai, A. M., Alrashed, M. A., Hezam, T. A., Alharthi, A. F., & Alharthi, S. F. (2025). The impact of ChatGPT on EFL medical track students' affective ilter. *Journal of Language Teaching and*

p-ISSN: 2579-7549
e-ISSN: 2579-7263

Research, *16*(1), 158–167. https://doi.org/10.17507/JLTR.1601.17

Arksey, H., & O'Malley, L. (2005). Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology*, *8*(1), 19–32. https://doi.org/10.1080/1364557032000119616

Ding, Y., & Zhu, J. (2025). Mobile based peer feedback in EFL speaking: learners' motivation, behavioral engagement in feedback provision, and achievement. *Asian-Pacific Journal of Second and Foreign Language Education*, *10*(1), 1–16. https://doi.org/10.1186/S40862-024-00314-9/FIGURES/3

Do, H., Lee, W., & Lee, G. G. (2024). *Acoustic Feature Mixup for Balanced Multi-aspect Pronunciation Assessment*. 312–316. https://doi.org/10.21437/interspeech.2024-2498

Du, J., & 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*, *6*, 100230. https://doi.org/10.1016/J.CAEAI.2024.100230

Feng, S., Halpern, B. M., Kudina, O., & Scharenborg, O. (2024). Towards inclusive automatic speech recognition. *Computer Speech and Language*, *84*. https://doi.org/10.1016/J.CSL.2023.101567

Going Digital: Making the Transformation Work for Growth and Well-being Going digital. (2017).

Jing, W. (2024). Speech recognition sensors and artificial intelligence automatic evaluation application in English oral correction system. *Measurement: Sensors*, *32*, 101070. https://doi.org/10.1016/J.MEASEN.2024.101070

Király, L. (2024). Artificial Intelligence as a virtual language teacher. *Magyar Nyelvor*, *148*(5), 649–665. https://doi.org/10.38143/Nyr.2024.5.649

Liu, X. J., Wang, J., & Zou, B. (2025). Evaluating an AI speaking assessment tool: Score accuracy, perceived validity, and oral peer feedback as feedback enhancement. *Journal of English for Academic Purposes*, *75*, 101505. https://doi.org/10.1016/J.JEAP.2025.101505

p-ISSN: 2579-7549
e-ISSN: 2579-7263

Luckyardi, S., Karin, J., Rosmaladewi, R., Hufad, A., & Haristiani, N. (2024). Chatbots as digital language tutors: Revolutionizing education through AI. *Indonesian Journal of Science and Technology*, *9*(3), 885–908. https://doi.org/10.17509/ijost.v9i3.79514

Mingyan, M., Noordin, N., & Razali, A. B. (2025). Improving EFL speaking performance among undergraduate students with an AI-powered mobile app in after-class assignments: an empirical investigation. *Humanities and Social Sciences Communications 2025 12:1*, *12*(1), 1–15. https://doi.org/10.1057/s41599-025-04688-0

Pituxcoosuvarn, M., Radhapuram, S. C. T., & Murakami, Y. (2024). Taboo talks: Enhancing ESL speaking skills through language model integration in interactive games. *Procedia Computer Science*, *246*(C), 3674–3683. https://doi.org/10.1016/J.PROCS.2024.09.189

Qassrawi, R. M., ElMashharawi, A., Itmeizeh, M., & Tamimi, M. H. M. (2024). AI-powered applications for improving EFL students' speaking proficiency in higher education. *Forum for Linguistic Studies*, *6*(5), 535–549. https://doi.org/10.30564/fls.v6i5.6966

Sahed, M. T. R., Aronno, M. T. I., Nyeem, H., Wahed, M. A., Ahsan, T., Islam, R. R., Ovi, T. B., Kundu, M. K., & Sadeef, J. A. (2025). LipBengal: Pioneering Bengali lip-reading dataset for pronunciation mapping through lip gestures. *Data in Brief*, *58*. https://doi.org/10.1016/J.DIB.2024.111254

Shi, X., Feng, T., Huang, K., Kadiri, S. R., Lee, J., Lu, Y., Zhang, Y., Goldstein, L., & Narayanan, S. (2024). Direct articulatory observation reveals phoneme recognition performance characteristics of a self-supervised speech model. *JASA Express Letters*, *4*(11), 114801. https://doi.org/10.1121/10.0034430