Proceeding of ICSPHS 2026 Symposium: Critical Perspectives on Global Education and Psychological Development
DOI: 10.54254/2753-7048/2026.HT29552

The Application Classification of Artificial Intelligence in Second Language Oral Learning and Its Impact on Oral Performance: A Systematic Review

Zhenyang Wang[1*], Zichen Zhou[2]

[1]*School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu, China*
[2]*School of Foreign Languages, Soochow University, Suzhou, China*
**Corresponding Author. Email: wwwangzhenyang@gmail.com*

Abstract. The swift development of artificial intelligence (AI) technology is giving rise to a profound shift in the modes of second language (L2) oral skills learning. The proposed research will contribute to the deliberate review of the current state of application of AI technology in L2 oral skills learning, its effectiveness on the learner's oral skills, and an insight into the current issues by employing a method of systematic literature review. Empirical research from the period of 2020 to 2025 is identified from ScienceDirect, Scopus, and ERIC following the PRISMA guidelines. The current literature states that there is a possibility of systematically dividing AI applications into four broad types of applications, including intelligent pronunciation feedback tools, conversational AI interfaces, automated speaking proficiency tools, and learning path recommendations. The effectiveness of AI technology is profound in improving L2 learners' fluency skills, accuracy of pronunciation, or both. However, the enabling effect of AI technology in improving HOLS skills of L2 learners is not significant. The application of AI technology provides opportunities for both personalised learning paths and immediate feedback. However, there still exist issues in application, including stability issues, a lack of emotional engagement, or the occurrence of bias. The proposed classification tool will offer theoretical foundations of research on this context in the forthcoming research. The upcoming research must focus on human-AI collaboration models or examine the longer-term role of this technology.

Keywords: second language oral learning, artificial intelligence in education, human-AI collaborative framework, oral performance

1. Introduction

Second language oral proficiency is core to cross-cultural communication. However, its acquisition has long been constrained by the limitations of traditional instruction. Traditional classrooms, restricted by time, teacher availability, and resources, often fail to provide sufficient personalized guidance and practice opportunities, easily leading to learner anxiety and the predicament of “dumb foreign language”.

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-7048/2026.HT29552

The developments in artificial intelligence (AI) technology, including automatic speech recognition, natural language processing, and large language models, offer revolutionary approaches in finding a solution to this problem. The use of AI technology in developing intelligent chatbots, virtual tutors, or pronunciation feedback tools provides a flexible, efficient, and stress-free environment for practising orals by enabling interactive dialogues, processing instantaneous and accurate feedback, as well as customised learning experiences. The technology has been effectively used in both offline learning settings, individual learning, and blended learning, helping learners accomplish their goals of enhanced fluency, pronunciation accuracy, or pragmatic skills.

However, there still exist some gaps in the current research. Firstly, most research on this issue pertains to short-term impacts of individual AI tools, which lack a generalised classification in addition to an overall assessment of the application of AI technology. Moreover, current research lacks consideration of differences in the application of AI technology in various learning settings. Moreover, there still lacks adequate discussion concerning its possible negative impacts, including its limitations (unstable feedback responses, lack of emotional engagement, biased algorithms).

Hence, this research will attempt to incorporate all the existing empirical data through a systematic review in order to fill the research gaps identified in the questions below. RQ1: What are the application domains of AI technology in second language (L2) speech learning? RQ2: What effect does AI technology have on L2 speech proficiency? RQ3: What are the key issues, as well as the strategies, involved in the application of AI technology? Through the answering of the aforementioned questions, this review seeks to provide definitive guidelines on the research.

2. Research methods

This study strictly followed the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines for systematic reviews to ensure transparency, rigor, and reproducibility.

2.1. Literature search strategy

The literature search encompassed three academic databases: ScienceDirect, Scopus, and ERIC. It covered publications from 2020 to August 2025. As Table 1 shows, keywords were combined and categorized into three groups, with Boolean operators “AND” and “OR” used to integrate these groups.

Table 1. Boolean search terms

Search section	Search terms
AI-technology related	“artificial intelligence” OR “AI” OR “chatbots” OR “speech recognition” OR “AI-based language learning apps” OR “virtual tutors”
L2-learning related	“second language” OR “L2” OR “foreign language”
Oral-skill related	“speaking” OR “oral proficiency” OR “fluency” OR “pronunciation” OR “pragmatic competence”

To supplement potentially omitted literature, the reference lists of key journals like Computer Assisted Language Learning, Pedagogy: Journal of English Language Teaching, Edelweiss Applied Science and Technology, Humanities and Social Sciences Communications were also manually searched.

DOI: 10.54254/2753-7048/2026.HT29552

2.2. Inclusion and exclusion criteria

Inclusion criteria encompass empirical studies, whether quantitative, qualitative, or mixed methods, published between 2020 and 2025, that explicitly investigate the direct impact of AI tools on L2 oral performance and provide clear methodological and results data.

Exclusion criteria consist of non-empirical studies such as reviews and theoretical articles, research not focused on L2 oral skills, studies without AI technology involvement, non-English publications, or those with an excessively small sample size.

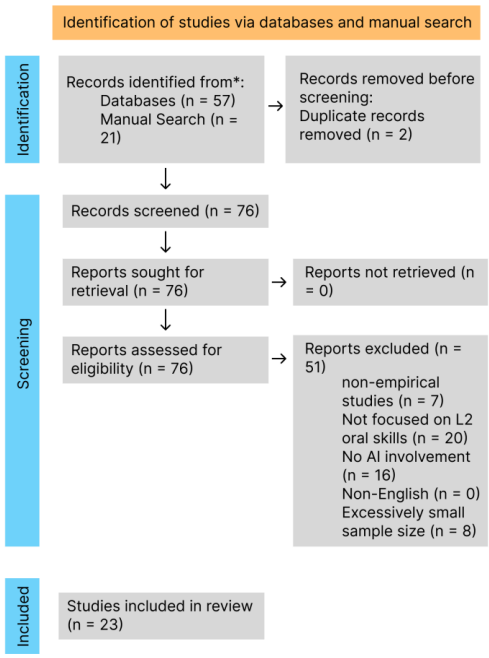

Figure 1. PRISMA flow diagram of the study selection process

Two independent reviewers assessed each record against the inclusion and exclusion criteria, achieving an initial interrater agreement of 95% via percentage agreement. Full consensus was reached after discussing discrepancies. In total, 51 records were excluded. The process resulted in 23 studies included in the review. Figure 1 shows the exclusions by criteria and the final number of studies.

2.3. Data extraction and analysis

The following information was extracted from the finally included literature to answer the research questions:

For RQ1: AI tool types, application scenarios, and technical functions.

For RQ2: Impact effects on oral fluency, pronunciation accuracy, and pragmatic competence, including measurement metrics and research design.

For RQ3: Challenges in implementation and corresponding coping strategies.

Data were extracted using standardized forms and analyzed using thematic analysis facilitated by NVivo software. Cross-validation among researchers was conducted to ensure reliability.

DOI: 10.54254/2753-7048/2026.HT29552

2.4. Quality assessment

The quality of each piece of research was assessed using a Critical Appraisal Skills Programme (CASP) checklist. The checklist helped consider how well the research was planned, how clearly the data were presented, and how reliable the findings were. Research with a high risk of bias (small population, unclear research methods, or potential conflicts of interest) was not considered. The use of this approach ensured that all quality evidence was considered.

3. Findings and discussion

For a thorough discussion, the authors chose and carefully analysed 23 research studies. Through this, the research was able to identify significant information, including types of AI tools, their effects on speech skills, among others. Based on the findings from this research, the authors respond to the three research questions posed.

3.1. RQ1: AI application classification framework

From the thematic analysis of the 23 studies encompassed in this literature, a classification framework of AI use in L2 Oral Learning was derived. The classification framework comprises four categories of AI tools that play specific roles.

The first category is known as intelligent pronunciation feedback systems. The system, which includes ELSA Speak or Google Read Along, relies on automatic speech recognition software which provides the learner with accurate feedback on pronunciation, enabling them to practice on their own [1]. Especially, this kind of system can be of immense help to beginners just learning pronunciation.

The second category is conversational AI agents. The use of applications such as ChatGPT or other chatbots can aid in the learning of languages by practising dialogues [2, 3]. The main benefit of this is that the dialogue can shift according to the levels of both individuals, making it simpler to increase fluency as well as speaking abilities [4, 5]. Practising all this can help the individual increase the confidence levels they need for carrying out a discussion.

Automated speech assessment tools, involving Liulishuo, examine key aspects of speech, including grammar, vocabulary, and rhythm [6, 7]. The detailed feedback provided by automated tools allows students to specifically address improvement opportunities identified by breaking down key issues into practice segments that can be followed across a protracted period of time.

To retain the interest of the learner, personalised learning tools like Duolingo or Babbel adapt according to an individual's level of performance as well as progress. What it seeks to do is ensure that the information provided is always up-to-date as well as of the right level of difficulty [8].

These four categories are summarised in Table 2 for clarity and quick reference.

DOI: 10.54254/2753-7048/2026.HT29552

Table 2. AI tools classification framework

Category	Description	Examples
Intelligent pronunciation feedback systems	Provide real-time phoneme-level feedback on articulation and intonation.	ELSA Speak, Google Read Along
Conversational AI agents	Simulate dialogues to boost fluency and reduce anxiety.	ChatGPT, app chatbots
Automated speaking assessment systems	Evaluate oral production with diagnostic reports.	Liulishuo
Personalized learning path recommendations	Recommend tailored tasks based on learner data for engagement.	Duolingo, Babbel

This framework provides a good foundation upon which more research can be done. For teachers, this framework is beneficial in giving them an efficient means of selecting appropriate tools of AI in various educational settings [9]. For developers, it provides significant guidelines on the development of AI technology, focusing on particular skills such as pronunciation or dialogue practice.

3.2. RQ2: impact on oral performance

The studies included show that AI can contribute positively to improving L2 speaking abilities. There is evidence of its effectiveness in simple tasks, though it starts becoming less effective in more complex ones.

The significant benefit of AI is that it provides many opportunities to practice with rapid feedback. For instance, tools such as Alexa can effectively develop speaking skills by providing a calm environment where a student can practice [10]. Moreover, adopting AI on a continuous basis promotes an encouraging environment which pushes the student to practice regularly by helping them develop positive learning habits [11].

Regarding the pronunciation accuracy, the effectiveness of AI can be attributed to its micro-analytical abilities. Statistics from data of L2 students on ASR platforms, iSpraak, indicate that pinpointing specific pronunciation mistakes enhances L2 pronunciation remarkably [12]. Hence, it is not surprising that L2 students find AI pronunciation tools positive, accepting the feedback provided as useful and beneficial [13].

Aside from skills, another role of AI is in its affective aspects that play an important role regarding speaking fluency. One of the most significant findings from the studies is that interacting with AI-powered chatbots can help alleviate foreign language anxiety and increase the motivation of language learners effectively [4, 14, 15]. The increase in confidence gained from non-evaluative AI partners bolsters more attempts in speech, as seen from the effective use of AI-driven apps in after-class assignment completion [16, 17]. The Speeko application is another manifestation of the role of AI in transforming the self-perception of language learner proficiency [6].

However, the enabling role of AI reaches its limitations with regard to pragmatic competence, understanding of culture, or sophisticated strategies of communication. The effectiveness of AI in this regard is often dependent on the quality of the models it uses, which often lack proper awareness of society.

A structured abstract of this outcome, incorporating key findings on both the positive outcome measures as well as the intrinsic limitations, is presented in Table 3.

DOI: 10.54254/2753-7048/2026.HT29552

Table 3. Summary of AI's impact on L2 oral performance

Aspect	Positive effects	Limitations
Fluency & automaticity	Increased speech rate and reduced pauses through repeated and low-pressure practice [10].	Gains from scripted practice do not fully bridge the gap to unpredictable real-world communication [14].
Pronunciation accuracy	Significant error reduction in segmental and suprasegmental features via immediate ASR feedback [1, 12]. High perceived effectiveness among learners for pronunciation training [13].	Potential for algorithmic bias in evaluating non-standard accents, leading to unfair assessments [13].
Affective factors	Significant reduction in speaking anxiety through non-judgmental interaction [4]. Increased willingness to communicate and self-efficacy [6].	Superficial emotional support due to a lack of genuine empathy and limiting long-term engagement.
Pragmatic & interactive competence	(Not a primary strength of current AI)	Weak support for pragmatic rules, such as politeness and turn-taking [5]. Inability to deal with cultural differences and adapt to the changing environment [14].

3.3. RQ3: challenges and coping strategies

The review of 23 studies (as illustrated in Table 4) identifies significant concerns in the use of AI technology in L2 speaking practice. The concerns relate to emotional, technical, ethical, and learning matters.

The challenge here is the lack of emotional intelligence in AI. For instance, the use of AI tools like chatbots in this aspect is common in offering stress-free zones for practice. The challenge with this is that, since they do not understand emotional variations, students could become more stressed or less interested as time progresses, owing to tough learning [4]. The best approach in this field will thus be the application of both human teachers and AI. Here, AI can offer practice responses in repetitive activities, while human teachers can offer emotional intelligence [18].

The second issue is bias in algorithms. The problem is that most of the current AI models tend to perform poorly on limited or similar datasets concerning the assessment of various accents or dialects of speech. This issue can lead not only to ineffective feedback responses but also to demoralising the students. For example, ELSA Speak or iSpraak finds it difficult to judge the characteristics of speech of different learner populations [12, 13]. The problem can be resolved by training the AI models on more varied datasets. The bias of the algorithms can also be taken care of by checking it on a periodic basis. Many tools use bias detection algorithms that ensure all users receive fair feedback [19].

The functionality of the AI tools can be hampered by limitations in technology. The unreliable speech recognition abilities of the tools and the resulting lag can contribute enormously to stoppages in learning. For instance, Chen stated that there was a considerable pause in the responses from the system, disrupting the flow of conversations [20]. Moreover, feedback error occurrence is frequent, identified by Kholis, and it was found by Ngo that background noise can erode the effectiveness of the tool [1, 21]. Enhancements in the reliability of fundamental algorithms, or carrying out offline preparations or testing on realistic scenarios, can help make the tools reliable.

DOI: 10.54254/2753-7048/2026.HT29552

The use of AI in education requires the processing of a substantial amount of data. However, an issue of great significance concerning data safety could result in a loss of trust among students [22]. For instance, there are voice-recording chattingbots that lack adequate safety, while there has been an issue with how data is collected by Duolingo [20, 23]. Trust can be preserved by following specific guidelines. The guidelines entail making data anonymous, clear usage of data, and specific ethics.

Teaching methodology still remains an issue. The training of teachers could be inadequate, and teachers' intentions in class could not align with the use of AI tools. This becomes an issue in blended classes [18]. The authors Polakova and Klimova referred to issues in the first testing classes [2]. The authors Mingyan, Noordin, and Razali showed that most of the students did not use the AI application in doing their assignment since they did not understand how [17].

Table 4. Challenges and strategies

Challenge	Description	Coping strategies
Lack of emotional interaction	AI lacks empathy, reducing motivation, especially for beginners.	Use emotion detection and human-AI hybrid systems.
Algorithmic bias	Biased datasets cause unfair accent or cultural evaluations.	Use diverse datasets and conduct bias audits.
Technical instability	Unreliable recognition or latency lowers trust.	Enhance ASR/NLP with updates and offline options.
Privacy concerns	Data collection risks breaches and ethical issues.	Anonymize data and enforce ethical guidelines.
Pedagogical integration	Teacher training gaps hinder AI alignment with goals.	Provide training and customizable AI features.

4. Conclusion

The current review relies on a set of 23 studies concerning the role of AI in L2 speaking skills. Based on classification (answer to RQ1), AI tools can be categorised into intelligent pronunciation feedback systems, conversational AI agents, automated speaking assessment tools, or learning route recommendations. Each category of tools focuses on supporting a different aspect of L2 speaking skills. It can also be seen that there are important benefits of AI in improving the fundamental level of pronunciation accuracy and fluency, as well as confidence-building in non-anxiety conditions (answer to RQ2), though there seem to be limitations of AI (answer to RQ2). The use of AI does not provide a complete solution. There appear to be various challenges of AI from the perspectives of technical issues, lack of emotional engagement, or ethics (answer to RQ3).

The implication of this finding is threefold. Turning to teachers first, this implies that AI must be incorporated in a complementary fashion in a blended learning approach, where it can be aptly used in repetitive drilling. Second, it implies that AI must be developed from an inclusive design perspective, emphasising the use of datasets from various sources, including affective computing. Third, it implies that this proposed framework of classification can effectively be used in comparative research on the application of AI among various demographic groups of students.

Although there are benefits, it is worth noting that this review has its own limitations. As it only considered English publications, there could be issues of publication bias, not including findings from non-English settings. Moreover, there could be an element of subjectivity with respect to the methodology of thematic analysis.

DOI: 10.54254/2753-7048/2026.HT29552

Future research avenues arising from this literature review are extending research into multilingual use of AI in L2 spoken learning, focusing on longitudinal research designs, exploring research approaches on ethics of data privacy, bias elimination, or both, or identifying human-AI collaboration approaches, perhaps involving co-teaching or virtual reality. Research efforts in this direction will be pivotal in unlocking the full potential of AI in the field of language learning.

Authors contributions

The authors equally contributed, and the names are listed in alphabetical order.

References

[1] Kholis, A. (2021) Elsa Speak App: Automatic Speech Recognition (ASR) for Supplementing English Pronunciation Skills. Pedagogy: Journal of English Language Teaching, 9, 01-14.
[2] Polakova, P. and Klimova, B. (2024) Implementation of AI-Driven Technology into Education—A Pilot Study on the Use of Chatbots in Foreign Language Learning. Cogent Education, 11, 2355385.
[3] Taeza, J. (2025) The Role of AI-Powered Chatbots in Enhancing Second Language Acquisition: An Empirical Investigation of Conversational AI Assistants. Edelweiss Applied Science and Technology, 9, 2616-2629.
[4] Yang, H., Kim, H., Lee, J.H. and Shin, D. (2022) Implementation of an AI Chatbot as an English Conversation Partner in EFL Speaking Classes. ReCALL, 34, 327-343.
[5] Tai, T.Y. and Chen, H.H.J. (2024) Improving Elementary EFL Speaking Skills with Generative AI Chatbots: Exploring Individual and Paired Interactions. Computers & Education, 220, 105112.
[6] Shafiee Rad, H. (2024) Revolutionizing L2 Speaking Proficiency, Willingness to Communicate, and Perceptions through Artificial Intelligence: A Case of Speeko Application. Innovation in Language Learning and Teaching, 18, 364-379.
[7] Takatsu, H., Suzuki, S., Eguchi, M., Matsuura, R., Saeki, M. and Matsuyama, Y. (2025) Gnowsis: Multimodal Multitask Learning for Oral Proficiency Assessments. Computer Speech & Language, 95, 101860.
[8] Dandu, G., Charyulu, G.M. and Ksnvl, L. (2024) AI-Driven Language Learning: The Impact of Rosetta Stone on ESL Students' Speaking Proficiency and Self-Control. Rupkatha Journal on Interdisciplinary Studies in Humanities, 16, 2.
[9] Wang, C. and Zou, B. (2025) D-ID Studio: Empowering Language Teaching with AI Avatars. TESOL Journal, 16, e70034. https: //doi.org/10.1002/tesj.70034
[10] Elmaadaway, M.A.N., El-Naggar, M.E. and Abouhashesh, M.R.I. (2025) Improving Primary School Students' Oral Reading Fluency through Voice Chatbot-Based AI. Journal of Computer Assisted Learning, 41, e70019.
[11] Ericsson, E. and Johansson, S. (2023) English Speaking Practice with Conversational AI: Lower Secondary Students' Educational Experiences over Time. Computers and Education: Artificial Intelligence, 5, 100164.
[12] Nickolai, D. (2024) Quantifying the Impact of ASR-Based Instruction: What Does the "iSpraak" Platform Learner Data Show? The EuroCALL Review, 31, 16-23.
[13] Abdel-Reheem Amin, E. (2024) EFL Students' Perception of Using AI Text-to-Speech Apps in Learning Pronunciation. SSRN.
[14] Zhang, D., Wu, J.G. and 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.
[15] Zou, B., Xie, S. and Wang, C. (2025) Students' Willingness to Communicate (WTC) in Using Artificial Intelligence (AI) Technology in English-Speaking Practice. International Journal of Information and Communication Technology Education, 21, 1-18.
[16] Choi, W.Y. (2025) The Role of Changes in Pronunciation Ability and Anxiety through Gen-AI on EFL Learners' Self-Directed Speaking Motivation and Social Interaction Confidence—A CHAT Perspective. System, 128, 103788.
[17] Mingyan, M., Noordin, N. and Razali, A.B. (2025) Improving EFL Speaking Performance among Undergraduate Students with an AI-Powered Mobile App in After-Class Assignments: An Empirical Investigation. Humanities and Social Sciences Communications, 12, 1-15.
[18] Zhou, Q., Hashim, H. and Sulaiman, N.A. (2025) Supporting English Speaking Practice in Higher Education: The Impact of AI Chatbot-Integrated Mobile-Assisted Blended Learning Framework. Education and Information Technologies.

DOI: 10.54254/2753-7048/2026.HT29552

[19] Abdelhalim, S.M. and Alsehibany, R.A. (2025) Integrating AI-Powered Tools in EFL Pronunciation Instruction: Effects on Accuracy and L2 Motivation. Computer Assisted Language Learning.
[20] Chen, K.T.C. (2022) Speech-to-Text Recognition in University English as a Foreign Language Learning. Education and Information Technologies, 27, 9857-9875.
[21] Ngo, T.T.N. and Chen, H.H.J. (2025) The Impact of ASR-Based Reading Tutor on Young EFL Learners' Oral Skills: A Case Study of Google Read Along. Computer Assisted Language Learning.
[22] Chen, M.R.A. (2024) The AI Chatbot Interaction for Semantic Learning: A Collaborative Note-Taking Approach with EFL Students. Language Learning & Technology, 28, 1-25.
[23] Kessler, M., Loewen, S. and Gönülal, T. (2025) Mobile-Assisted Language Learning with Babbel and Duolingo: Comparing L2 Learning Gains and User Experience. Computer Assisted Language Learning, 38, 690-714.