

*International Journal of Learning, Teaching and Educational Research*  
Vol. 25, No. 3, pp. 853-873 March 2026  
<https://doi.org/10.26803/ijlter.25.3.37>  
Received Nov 19, 2025; Revised Jan 18, 2026; Accepted Mar 19, 2026

## Artificial Intelligence in English-Speaking Education: A WoS-Based Bibliometric Analysis of Technology, Pedagogy, and Ethics (2021-2025)

**Bo Zhou** , **Lim Seong Pek\*** 

Faculty of Education and Liberal Arts  
INTI International University  
Nilai, Negeri Sembilan, Malaysia

**Nahdia Kabir** 

Faculty of Business and Communications  
American International University-  
Bangladesh Dhaka, Bangladesh

**Jiaying Yang** 

Library, South China Normal University,  
Guangzhou, Guangdong Province, China

**Mohamed Bouteraa** 

Faculty of Business Administration  
Qatar Foundation-American University of Qatar  
Doha, Qatar

**Xizi He** 

Fundamental Education Department, Hebei Academy of Fine Arts,  
Shijiazhuang, Hebei, 050700, China

**Abstract.** This study provides a bibliometric overview of artificial intelligence (AI)-assisted English-speaking education based on 248 WoS-indexed articles published between 2021 and 2025 (2,046 citations; H-index = 24). Using co-citation, co-occurrence, and temporal analyses, the study identifies major research patterns. The findings grounded and shift from an early emphasis on speech technologies to learner-centered, pedagogically grounded, and affect-sensitive approaches. The analysis identifies a critical gap, as ethical and governance concerns remain insufficiently addressed. Ethics-related terms do not appear among high-frequency or high-centrality keywords, despite their growing relevance

---

\*Corresponding author: Bo Zhou; [i24028791@student.newinti.edu.my](mailto:i24028791@student.newinti.edu.my)

in AI-enhanced learning. By synthesizing structural and temporal patterns, the study clarifies interactions among technological, pedagogical, and learner-related dimensions, while highlighting the under-theorization of ethics and governance. It also reveals a fragmented use of theoretical models. Frameworks such as the Technology Acceptance Model (TAM), Social Cognitive Theory (SCT), and Self-Determination Theory (SDT) are typically applied in isolation rather than being integratively linked to instructional design. Beyond descriptive mapping, this study advances an integrative perspective linking technology, pedagogy, and learner experience, with implications for teachers, researchers, and policymakers. The findings align with UNESCO's SDG 4, highlighting the need for inclusive and equitable AI-supported speaking education.

**Keywords:** large language models; co-citation analysis; keyword co-occurrence; TAM; EFL

## 1. Introduction

The rapid emergence of large language models such as ChatGPT has reshaped English-speaking education by transforming how speaking practice and feedback are delivered. Artificial intelligence-based (AI-based) tools now offer individualized practice, immediate feedback, and interactive learning environments, with reported benefits for learner engagement, speaking fluency, and pronunciation accuracy (Zou et al., 2025a). Since 2023, scholarly interest in AI-assisted English-speaking education has expanded rapidly, reflecting the growing pedagogical relevance of generative AI.

Despite this momentum, challenges remain in the effective integration of AI into speaking instruction. Limitations in semantic understanding, contextual adaptability, pragmatic accuracy, and instructional alignment may constrain learners' autonomous speaking development (Rudolph et al., 2023). More broadly, existing research has tended to prioritise technical performance, while pedagogical integration and ethical considerations have received less systematic attention (Dergaa et al., 2023; Zawacki-Richter et al., 2019). Recent bibliometric evidence suggests that technology-oriented keywords dominate the literature, whereas ethics- and pedagogy-related terms appear far less frequently.

Against this background, this study employs a bibliometric approach to map the research landscape of AI-assisted English-speaking education. Using co-citation, co-occurrence, and temporal analyses, it examines major research trends, influential publications, theoretical frameworks, and emerging topics across technological, pedagogical, and ethical dimensions. Bibliometric analysis is particularly suitable for this purpose, as it enables the systematic synthesis of large-scale and interdisciplinary literature.

The general aim of this study is to provide a systematic bibliometric overview of research on AI-assisted English-speaking education. While the study is primarily descriptive rather than theory-building, it adopts an integrative perspective to examine how technological, pedagogical, and ethical dimensions are represented

in existing research. The findings are intended to inform educators, researchers, and policymakers concerned with the responsible integration of AI in speaking instruction. The study addresses the following research questions:

**RQ1:** What are the major research trends, influential publications, and theoretical frameworks in AI-assisted English-speaking education?

**RQ2:** What emerging research themes and interdisciplinary connections across technological, pedagogical, and ethical dimensions are evident in AI-assisted English-speaking education?

**RQ3:** How is pedagogical integration of AI tools conceptualized and implemented in English-speaking education research?

Nevertheless, bibliometric findings should be interpreted with caution, as results may be influenced by database coverage, keyword selection, and the rapidly evolving nature of AI research. Moreover, adoption-oriented frameworks in prior research primarily emphasize user acceptance, offering limited insight into pedagogical integration and ethical governance in AI-supported speaking instruction. This study aligns with Sustainable Development Goal 4 by promoting inclusive and ethically responsible AI integration in education (UNESCO, 2021).

## 2. Literature Review

In recent years, scholarly interest in the integration of AI tools into English language education has increased, as reflected in bibliometric evidence on technology-enhanced learning and AI-supported instruction (Bo et al., 2025; Lim et al., 2025). Empirical studies indicate that AI-powered applications can enhance learners' speaking fluency, pronunciation accuracy, and engagement by providing automated feedback and interactive practice environments (Ma et al., 2025). Such tools are particularly valued in English as a Foreign Language (EFL) context for offering low-anxiety and immersive speaking practice (Zhou et al., 2025; Tai & Chen, 2024).

Despite these technological benefits, pedagogical challenges remain. AI-generated feedback often lacks sufficient semantic depth, pragmatic sensitivity, and contextual adaptability, which may lead learners to over-rely on technology and constrain the development of autonomous speaking abilities (Rudolph et al., 2023). Moreover, existing research tends to prioritise technical performance indicators, such as speech recognition accuracy and feedback responsiveness, while paying limited attention to instructional design, curriculum integration, and teacher mediation (Kasneci et al., 2023).

Ethical considerations have also become increasingly salient in AI-assisted speaking education. Issues related to data privacy, algorithmic bias, and unequal access to technology pose challenges to fairness and inclusivity in language learning contexts (Dergaa et al., 2023). However, these concerns are frequently addressed in a fragmented manner and are rarely integrated into pedagogical or theoretical frameworks. Commonly used adoption frameworks (e.g., TAM and UTAUT) have shaped prior work, yet recent methodological syntheses argue that such models require stronger pedagogical contextualization and ethical positioning in AI-mediated learning environments (Öztürk et al., 2024).

Recent studies therefore call for more integrative perspectives that connect technological affordances with pedagogical practices and ethical responsibility. Mohamed (2024) highlights the importance of aligning AI tools with learner autonomy and critical thinking, while Dwivedi et al. (2023) stress the role of institutional policy and contextual factors in responsible AI implementation. Nevertheless, a systematic synthesis of research trends, theoretical foundations, and thematic evolution in AI-assisted English-speaking education remains lacking, underscoring the need for a comprehensive bibliometric mapping aligned with the research objectives of the present study.

### 3. Methodology

#### 3.1 Bibliometric Research Design

This study adopts a bibliometric research design to examine research on AI-enhanced English language acquisition. Bibliometric analysis is a quantitative, data-driven approach that maps scientific knowledge using citation- and keyword-based techniques. Citation- and keyword-based bibliometric techniques were employed to identify influential publications, core themes, collaboration patterns, and research trajectories (Zhou et al., 2026; Wider et al., 2024). Data were retrieved from the Web of Science (WoS) Core Collection due to its rigorous indexing standards and suitability for bibliometric analysis in education and AI.

#### 3.2 Bibliometric Data Collection

To enhance the accuracy and academic rigour of data retrieval, this study included peer-reviewed journal articles published in English and indexed in SSCI, SCI-Expanded, or ESCI. The inclusion of ESCI allowed high-quality emerging studies to supplement the more established outputs from SSCI and SCI-Expanded. The Web of Science Core Collection was searched using the Topic (TS) field with a Boolean strategy combining terms related to artificial intelligence (e.g., AI, chatbot, automatic speech recognition, NLP) and English-speaking skills (e.g., English speaking, oral English, spoken English, speaking fluency, pronunciation training). The time frame covered 2021–2025. Table 1 summarizes the inclusion criteria, including database coverage, document type, language, and indexing requirements.

**Table 1: Inclusion Criteria for Bibliometric Analysis**

WoS Database	ALL
Time period	2021 to 2025
Search field	TS
Search keywords	("artificial intelligence" OR AI OR chatbot* OR "automatic speech recognition" OR NLP) AND ("English speaking" OR "oral English" OR "spoken English" OR "speaking fluency" OR "pronunciation training")
Document type	Article
Language	English
WoS Index	SSCI or SCI-EXPANDED or ESCI

A systematic screening procedure was implemented to ensure the inclusion of high-quality, peer-reviewed studies relevant to the research focus. The preliminary dataset comprised 401 records obtained from the WoS Core

Collection. The screening process comprised four primary steps, as illustrated in Figure 1. To enhance screening reliability, two researchers independently reviewed the records at each screening stage. Any discrepancies were resolved through discussion until consensus was reached. Although the exclusion of non-English publications may introduce language bias, this decision was made to ensure consistency in data interpretation and analytical reliability across bibliometric indicators.

- Identification: A total of 401 records were identified through thematic searches.
- Screening: A total of 98 records that did not satisfy the search criteria were excluded, resulting in 303 articles for additional review.
- Eligibility: Thirty-six records not categorized as research articles were excluded, leaving 251 full-text articles evaluated for eligibility.
- Exclusion: Ultimately, three articles composed in non-English languages were omitted, yielding a total of 248 peer-reviewed journal articles for bibliometric analysis.

This rigorous filtering process ensured that all selected studies met the criteria of linguistic uniformity, quality, and relevance.

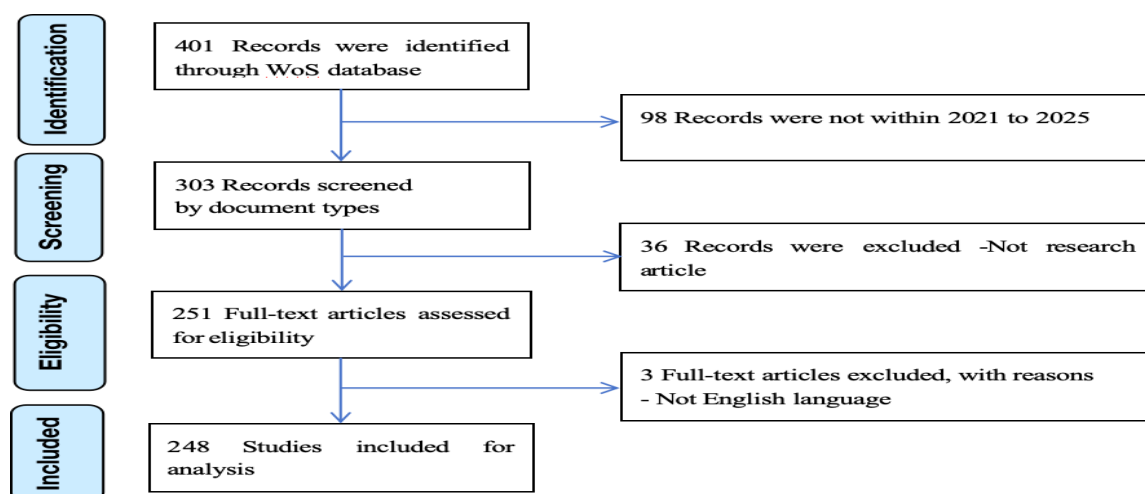


Figure 1: PRISMA Flowchart

### 3.3 Bibliometric Data Analysis

The dataset of 248 peer-reviewed articles was analyzed using VOSviewer and CiteSpace, two widely used bibliometric mapping tools for constructing and visualizing scientific knowledge networks. The study employed five complementary bibliometric techniques (Van Eck & Waltman, 2017), including performance analysis, co-citation analysis, keyword co-occurrence analysis, time evolution analysis, and burst detection analysis. For keyword co-occurrence, time evolution, and burst detection analyses conducted in CiteSpace, default parameter settings were adopted, allowing the software to automatically generate networks based on keyword frequency distributions and ensuring analytical objectivity and reproducibility. For co-citation analysis, VOSviewer was used to construct the co-citation network based on WoS bibliographic data. The Leiden algorithm was applied with a co-citation threshold of 47 to generate distinct

thematic clusters grounded in shared citation linkages and representative literature. Following the automated clustering generated by the software, cluster labels were manually reviewed and refined by examining representative keywords, highly cited documents, and article titles within each cluster to ensure conceptual coherence and interpretability.

## 4. Results

### 4.1 Bibliometric Performance and Annual Trends

The bibliometric analysis of AI-assisted English-speaking education provided useful metrics that highlight the influence and development of this field. A total of 248 yielded 2,046 citations, with 1,960 excluding self-citations. These articles were cited by 1,733 different documents (1,689 excluding self-citations), for an average of 8.25 citations per item. The H-index of 24 indicates that at least 24 of these articles have received 24 or more citations.

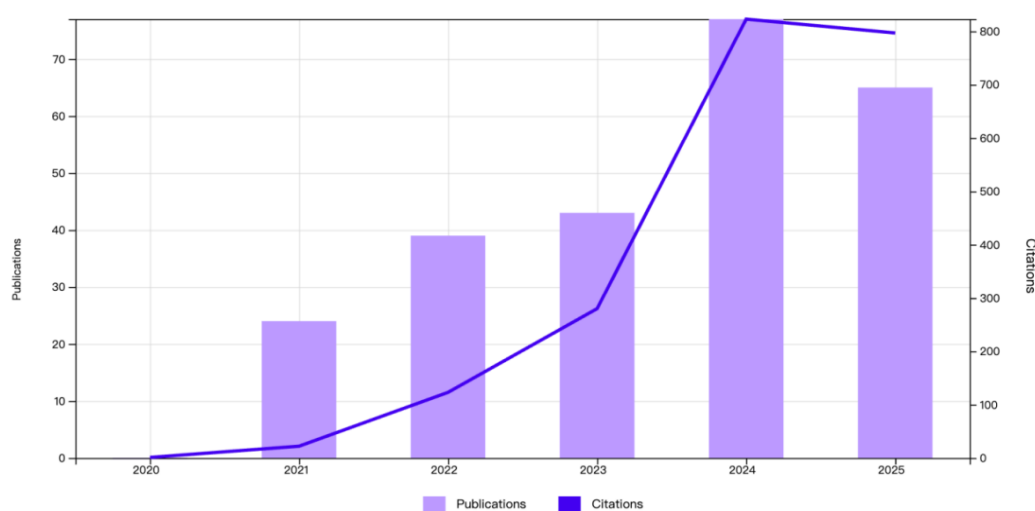


Figure 2: Publications and citations (2021-2025)

Figure 2 depicts the annual distribution of WoS-indexed publications and citations in AI-assisted English-speaking education from 2021 to 2025. The bars represent the number of peer-reviewed publications, while the line represents the corresponding citation counts. The field has steadily grown, with publications rising from 24 in 2021 to 77 in 2024, with a provisional total of 65 expected in 2025. Citation counts followed a similar pattern, beginning at a low level in 2021 and exceeding 800 in 2024, with only a minor decrease observed in 2025, according to data available as of July 20. This upward trend reflects growing scholarly interest and research productivity in AI-assisted English language learning. As a bibliometric study, the present analysis focuses on structural patterns, thematic prominence, and temporal dynamics rather than effect sizes in the experimental sense; practical significance is therefore interpreted through citation impact, keyword centrality, and thematic evolution.

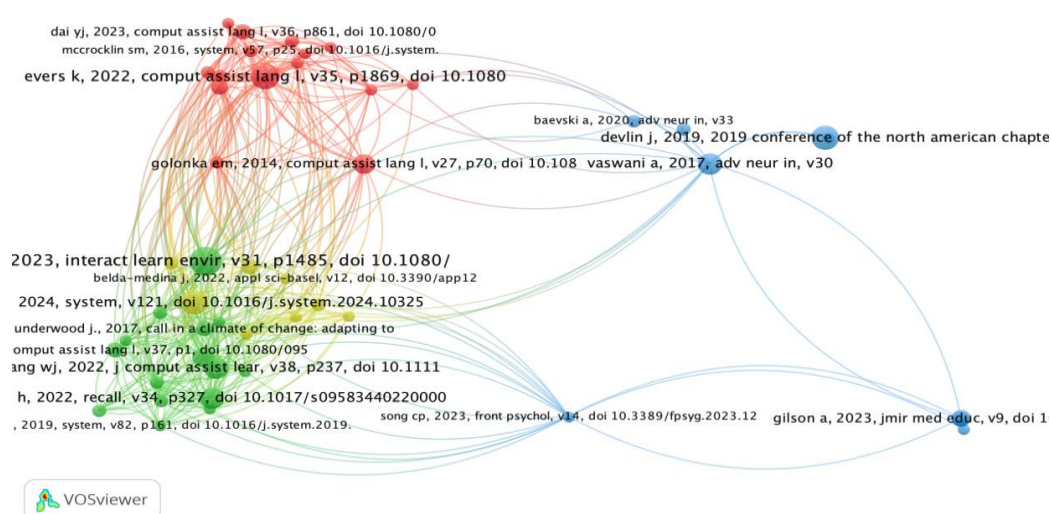
## 4.2 Most cited documents

**Table 2: Top 5 Most Cited Articles**

Rank	Authors	Title	Citations
1	Fathi et al. (2024)	Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions	124
2	El Shazly, R. (2021)	Effects of artificial intelligence on English speaking anxiety and speaking performance: A case study	86
3	Oruh et al. (2022)	Long short-term memory recurrent neural network for automatic speech recognition	81
4	Hsu et al. (2023)	Proposing a task-oriented chatbot system for EFL learners' speaking practice	63
5	Rogerson-Revell (2021)	Computer-assisted pronunciation training (CAPT): Current issues and future directions	54

Table 2 shows the top five most-cited publications on AI-assisted English language learning from 2021 to 2025. Fathi et al. (2024), with 124 citations, found that AI-mediated interactions improved EFL learners' speaking skills and willingness to communicate. Oruh et al. (2022), cited 81 times, developed an LSTM-based speech recognition model that serves as the foundation for adaptive AI speaking systems. El Shazly (2021), with 86 citations, investigated how AI-assisted environments helped reduce speaking anxiety. Hsu et al. (2023), cited 63 times, proposed a task-oriented chatbot to help EFL speakers practice. These representative studies include the pedagogical, technological, and affective dimensions of current AI-assisted English-speaking research.

## 4.3 Co-Citation Analysis by Clusters



**Figure 3: Co-citations Analysis (VOSviewer Visualisation)**

The co-citation analysis generated four distinct thematic clusters. Figure 3 depicts the resultant co-citation network, while Table 3 summarizes four primary themes: AI in pronunciation and speech recognition, AI in learner engagement and motivation, human-AI interaction for oral communication assistance, and Interdisciplinary AI in academic and medical English. These clusters collectively form the intellectual structure of AI-assisted English-speaking research, highlighting learner attitudes, technological affordances, and AI-supported speaking development.

Cluster 1 (Red) focuses on AI-augmented pronunciation and speech recognition, emphasizing the transformative influence of ASR technologies on oral language instruction. Empirical studies, such as those by Fouz-González (2020) and Dai and Wu (2023), indicate that ASR tools significantly improve pronunciation accuracy and fluency via real-time, individualized feedback. Bashori et al. (2024) demonstrate that such immediacy enhances learner awareness and self-regulation. ASR enhances language proficiency and learner independence when integrated with peer-assisted strategies. This trajectory is supported by theoretical principles from Social Cognitive Theory (SCT) and Self-Determination Theory (SDT), which emphasize the significance of perceived competence and self-efficacy in sustained engagement. This cluster perceives AI-driven speech tools as pedagogically revolutionary, integrating technological functionalities with learner psychology to facilitate more adaptive and engaging pronunciation instruction.

Cluster 2 (Green) examines the relationship between learners' acceptance of AI technologies and their motivation to engage in English-speaking activities. Grounded in adoption-oriented models (e.g., TAM), this cluster shows that perceived usefulness and ease of use are associated with learner engagement (Zou et al., 2025c). In particular, perceived usefulness positively predicts learners' motivation and speaking performance in AI-supported contexts (Zou et al., 2025b). Rahimi and Fathi (2024) report that authentic, low-stress AI-supported tasks (e.g., e-tandem learning) enhance motivation to communicate, while Huang et al. (2022) demonstrate that perceptions of chatbot responsiveness support sustained engagement. Overall, these findings indicate that motivation evolves through interaction with AI-mediated environments, highlighting the need for adaptive models beyond TAM that capture shifting learner attitudes and communicative confidence.

Cluster 3 (Blue) focuses on human-AI interactions in the development of oral communication skills in English language learning. Chatbots, virtual assistants, and social media-based platforms are examples of AI-powered tools that enable EFL learners to practice simulated conversational skills. Research by Fathi et al. (2024) and Zou et al. (2023a) demonstrates that these tools offer organized, low-stress speaking opportunities that alleviate anxiety and enhance communication willingness. Belda-Medina and Calvo-Ferrer (2022) emphasize the ability of chatbots to provide accessible, immediate, and contextually relevant responses. Hsu et al. (2023) developed a task-oriented chatbot system that allows for extended AI-mediated conversational practice, supporting more authentic and meaningful language use. This cluster is theoretically grounded in SCT, which

emphasizes learning via social interaction and feedback. In this study, AI is described as an interactive partner rather than a passive teacher, boosting learners' self-efficacy, autonomy, and communicative engagement.

Cluster 4 (Yellow) investigates the application of AI in specialized English language acquisition, particularly in academic and medical contexts. Studies on ChatGPT's performance in the USMLE demonstrate its effectiveness in supporting terminology acquisition, content comprehension, and exam preparation (Gilson et al., 2023; Kung et al., 2023). Complementing this line of research, Sallam (2023) synthesizes evidence on ChatGPT in healthcare education, highlighting both its pedagogical potential and ethical implications.

Collectively, these studies indicate that large language models can enhance professional communication and engagement with discipline-specific content. Pedagogically, this cluster aligns with Task-Based Language Teaching (TBLT), as AI tools facilitate authentic, goal-oriented communicative tasks that strengthen learners' linguistic competence and vocational preparedness. Although ethical issues such as fairness and data security are acknowledged, they remain peripheral rather than constituting a coherent research focus within this cluster.

**Table 3: Co-Citation Cluster on AI-assisted English Speaking**

Cluster No. & Colour	Cluster Labels	No. of Articles	Representative Publications
Cluster 1 (Red)	AI-Enhanced Pronunciation and Speech Recognition in Language Learning	15	Evers & Chen (2022), Fouz-González (2020), Dai & Wu (2023), Bashori et al. (2024), Tejedor-García et al. (2020), El Shazly, R. (2021)
Cluster 2 (Green)	AI Acceptance and Learner Motivation in EFL Speaking	15	Rahimi & Fathi (2024), Jeon (2024), Huang et al. (2022), Tai & Chen (2023), Zou et al. (2025b), Zou et al. (2025c)
Cluster 3 (Blue)	Human-AI Interaction for Oral Communication Support	9	Fathi et al. (2024), Zou et al. (2023a), Belda-Medina & Calvo-Ferrer (2022), Hsu et al. (2023)
Cluster 4 (Yellow)	Interdisciplinary AI in Academic and Medical English	8	Gilson et al. (2023), Kung et al. (2023), Sallam (2023), Song & Song (2023)

Overall, the co-citation analysis indicates that research on AI-assisted English-speaking education is primarily structured around speech technologies, learner motivation, interactive AI systems, and domain-specific applications, while ethical considerations remain weakly integrated across clusters.

#### 4.4 Keyword Co-Occurrence and Cluster Analysis



Figure 4: Keyword Co-occurrence Network (Cite Space)

Figure 4 visualizes the conceptual structure of AI-assisted English-speaking research based on a keyword co-occurrence network created using CiteSpace (v6.3.R1). The network includes 181 nodes and 516 edges, indicating a moderately connected research landscape with multiple intersecting topics (Chen et al., 2020). High-frequency and high-centrality terms underline the dominant research foci. As shown in Table 4, “artificial intelligence” functions as the major hub (frequency = 73; centrality = 0.87), linking technological, pedagogical, and learner-oriented themes. Keywords such as “automatic speech recognition,” “English,” and “model” stress the field’s sustained dependence on core speech-processing technologies (Hwang et al., 2024).

In contrast, “English speaking” appears in peripheral positions, indicating that speaking outcomes are typically addressed indirectly via technology acceptance or system-focused constructs rather than as explicit analytical targets (Zou et al., 2025c). Ethical terms are likewise absent from the top 15 keywords, indicating that ethical and governance issues remain insufficiently theorized. The emergence of terms such as “Large Language Models” and “anxiety” reflects increasing interest in generative AI and affective variables, while the continued prominence of “speech recognition” highlights an ongoing reliance on core speech-processing technologies (Huang & Zou, 2024). This pattern likely reflects researchers’ tendency to operationalize speaking outcomes through more specific technical or pedagogical constructs—such as pronunciation accuracy, ASR-based feedback, or anxiety reduction—rather than the generic term “English speaking.”

Table 4: Top 15 Keywords by Frequency and Centrality, (2021-2025 Dataset)

Frequency	Centrality	Year	Keyword
73	0.87	2021	Artificial intelligence
20	0.09	2022	Natural language processing
15	0.24	2023	English
13	0.2	2021	Automatic speech recognition
12	0.14	2023	Language
12	0.09	2022	Model
11	0.09	2021	Speech Recognition
10	0.16	2021	Recognition
8	0.07	2024	Large Language Models
7	0.03	2021	Anxiety
7	0.03	2021	System
6	0.02	2022	Care
6	0.03	2022	2nd language
6	0.03	2023	Speech
5	0.01	2024	English speaking

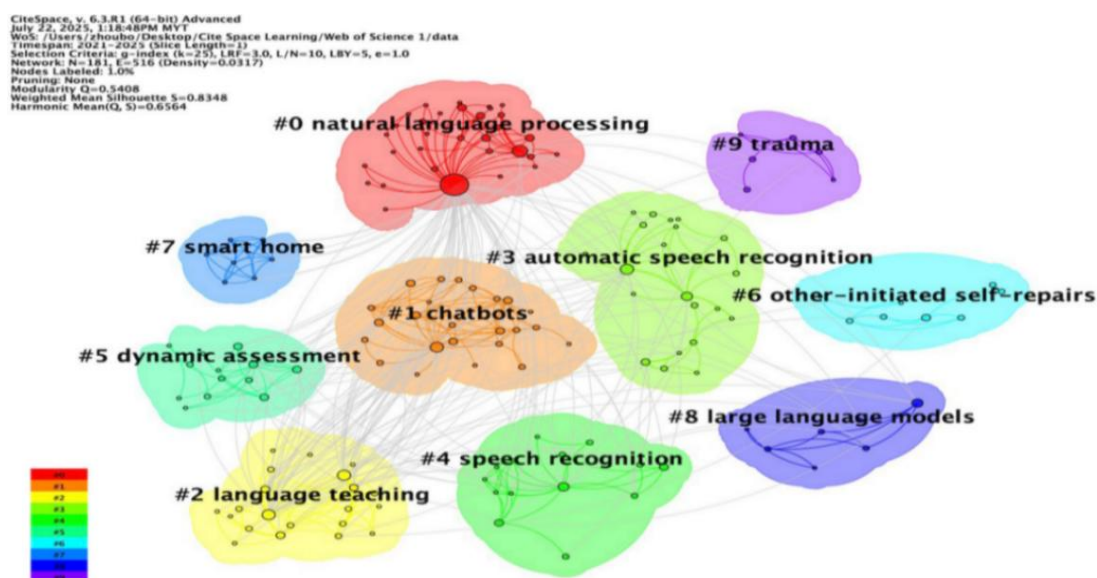


Figure 5: Keyword Cluster Map by LLR (Cite Space)

The LLR clustering (Figure 5; Table 5) offers a finer-grained thematic structure of the keyword network. Ten clusters were generated (modularity  $Q = 0.5408$ ; silhouette = 0.8348), indicating a well-structured and reliable clustering solution, with acceptable inter-cluster separation and strong intra-cluster consistency. To improve interpretability, cluster labels were reviewed manually. To enhance interpretability, cluster labels, node keywords, and cited abstracts were manually examined. Several clusters exhibited thematic overlap, requiring consolidation. Cluster #0 "Natural Language Processing" (32 nodes, silhouette = 0.752, 2023) and Cluster #1 "Chatbot-based" (24 nodes, silhouette = 0.749, 2023) represented AI-mediated oral communication systems, such as computer-assisted pronunciation

training (CAPT) and conversational agents, respectively. These were combined into the category of "AI-enabled speaking interfaces" (Huang et al., 2022; Yang et al., 2022). Cluster #3 "Automatic Speech Recognition" (22 nodes, silhouette = 0.802, 2022) and Cluster #4 "Speech Recognition" (14 nodes, silhouette = 0.814, 2022) were grouped as "ASR-based evaluation systems", reflecting their shared emphasis on speech-processing technologies used for pronunciation training and automated feedback (Rogerson-Revell, 2021).

**Table 5: keyword Co-occurrence Cluster Summary (LLR, Cite Space)**

Label	No. of Nodes	Silhouette Score	Year
0	32	0.752	2023
1	24	0.749	2023
2	22	0.831	2023
3	22	0.802	2022
4	14	0.814	2022
5	13	0.861	2022
6	8	0.962	2025
7	7	0.989	2022
8	7	0.856	2024
9	6	0.933	2024

Unlike the merged clusters, several clusters were kept due to their thematic distinctiveness and internal cohesion. Cluster #2 "Language Teaching" (22 nodes, silhouette = 0.831, 2023) and Cluster #5 "Dynamic Assessment" (14 nodes, silhouette = 0.814, 2022) address fundamental pedagogical concerns such as task design and formative assessment. Smaller yet coherent clusters, such as #6 "Self-Repairs" (8 nodes, silhouette = 0.962, 2025) and #7 "Smart Home" (7 nodes, silhouette = 0.989, 2022), emphasize micro-interactional feedback and informal learning contexts (Jeon, 2024; Tai & Chen, 2023).

Cluster #9 "Trauma" (6 nodes, silhouette = 0.933, 2024) was classified as peripheral due to its low educational or technological relevance. Collectively, the refined clusters reveal three overarching strands in AI-assisted English-speaking research: AI-driven interaction design, pedagogical applications, and affective support mechanisms (Fathi et al., 2024). They also highlight developing interdisciplinary linkages between language education, psycholinguistics (self-repairs), mental health (trauma), and human-computer interaction (smart home), implying a broader integration with digital learning ecosystems.

## 4.5 Timeline and Burst Detection of Emerging Trends

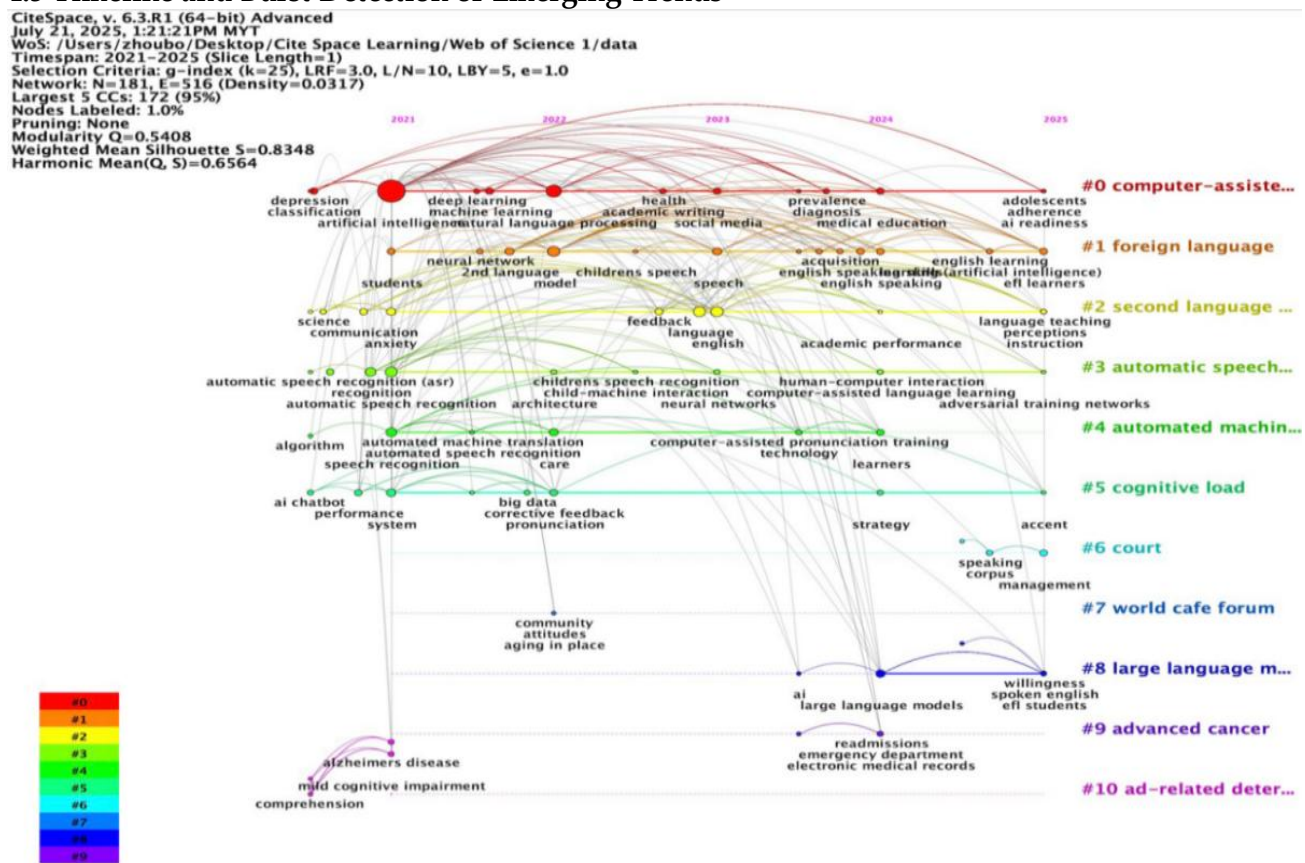


Figure 6: AI Speaking Keyword Evolution (2021– 2025, CiteSpace)

Figure 6 depicts the temporal evolution of keyword clusters in AI-assisted English-speaking research, from 2021 to 2025. Each horizontal line represents a thematic cluster, with labelled nodes indicating keywords that appeared or gained popularity during specific years. The timeline layout enables the identification of when clusters became active and how long they remained central to the scholarly conversation. Early-stage research between 2021 and 2022 focused heavily on Cluster #0 (computer-assisted pronunciation training) and Cluster #3 (automatic speech recognition), reflecting a strong emphasis on speech technologies such as ASR tools, CAPT systems, and pronunciation feedback (e.g., “automatic speech recognition,” “speech recognition,” “corrective feedback”) (Evers & Chen, 2022). These clusters form the technical backbone of early AI-driven speaking instruction, primarily emphasizing system performance and speech recognition accuracy.

From 2023 onwards, the field saw a thematic shift toward Cluster #2 (second language acquisition) and Cluster #1 (foreign language learning), with frequent keywords such as “language teaching,” “instruction,” and “academic performance.” This suggests growing attention to pedagogical frameworks and learner development in AI-speaking environments (Zou et al., 2025b). Additionally, Cluster #8 (large language models) and terms such as “AI” and “willingness” emerged in 2024, signifying a transition towards generative AI applications and learner affect. Minor yet more consequential clusters, such as

Cluster #5 (cognitive load) and Cluster #6 (court), indicate specialized research domains, including cognitive effort in AI verbal tasks and applications tailored to specific fields. Overall, the timeline shows a clear shift in research priorities: from early studies of speech technologies to broader concerns about pedagogy, learner engagement, and AI-human interaction. The thematic transitions seen between 2021 and 2025, reflect the field's adaptation to technological advancements and the emergence of large-scale, generative AI systems in language education.

## Top 17 Keywords with the Strongest Citation Bursts



Figure 7: Keyword Burst Detection (2021–2025, CiteSpace)

Burst detection is an effective method for identifying emerging research frontiers, as it detects sudden surges in keyword usage over time (Öztürk et al., 2024). Figure 7 illustrates the 17 most prominent burst keywords in AI-assisted English-speaking research from 2021 to 2025. In this visualization, "Year" denotes the keyword's first appearance, "Strength" represents the magnitude of the surge, and the red bar depicts the period of increased citation growth.

The most significant increases occurred in the initial phase (2021-2022), with terms like "performance," "automatic speech recognition," and "computer-assisted pronunciation training" showing increased burst intensities. These represent early research priorities in speech technology, pronunciation feedback systems, and performance-based evaluation tools (Dai & Wu, 2023). More recent bursts, particularly "speaking skills" and "L2", began in 2024 and will continue into 2025, indicating a shift toward learner-centered outcomes and second language proficiency in AI-mediated environments.

These emerging trends reflect increased scholarly interest in how AI tools can improve speaking fluency, accuracy, and communicative competence (Fathi et al., 2024). This burst pattern reflects a larger trend: a shift from technology-centric systems to pedagogically driven, learner-centered applications of AI-assisted

English-speaking education. These findings are consistent with recent calls for integrated, personalized, and affect-sensitive AI interventions in language education.

## 5. Discussion

Based on the bibliometric patterns identified in the Results section, three integrative observations characterize the current state of AI-assisted English-speaking education. First, despite thematic diversity across speech technologies, learner motivation, interactive AI systems, and domain-specific applications, the field remains structurally fragmented. Earlier reviews likewise report isolated research strands, and the present analysis demonstrates this fragmentation at a structural level. For example, studies on automatic speech recognition prioritize system accuracy and feedback, whereas motivation-focused research emphasizes willingness to communicate with limited linkage to curriculum design or assessment. This separation restricts the translation of research into coherent classroom practice.

Second, the shift from technology-driven experimentation toward learner-centred and pedagogically informed inquiry mirrors trends in prior AI-in-education research but cannot be explained by technological maturation alone. The widespread availability of AI tools after 2023 reduced the novelty of system-level evaluations, while parallel developments in applied linguistics and CALL foregrounded affective and contextual dimensions. Third, the absence of ethics-related themes from high-frequency keywords suggests that ethical governance remains structurally marginalized. Early research was largely shaped by adoption-oriented models such as the Technology Acceptance Model, leaving issues of fairness, data privacy, accountability, and professional identity under-theorized. Overall, these findings highlight the need for an integrated analytical framework linking technology, pedagogy, learner psychology, and ethics.

### 5.1 Theoretical Implications

Building on the above discussion, this study offers several theoretical and practical implications for AI-assisted English-speaking education. At the theoretical level, the field exhibits a diverse yet fragmented landscape. Although existing studies address speech-recognition technology (Jiang et al., 2021), learner motivation, human-AI interaction, and domain-specific speaking skills, these strands largely develop in isolation. Frameworks such as TAM, SCT, SDT, and TBLT are typically applied independently. As a result, researchers face difficulties in explaining how technological, pedagogical, and affective processes interact within AI-mediated speaking environments.

From a research perspective, the observed temporal trends indicate a shift from technology-centred perspectives toward learner-centred and pedagogically driven approaches. Early work focused primarily on ASR accuracy and system performance, whereas recent studies emphasize interactional competence, L2 speaking development, and socio-emotional factors. This evolution suggests that static adoption-oriented accounts are insufficient for capturing the dynamic,

pedagogically situated nature of AI use, calling for more context-sensitive and integrative perspectives (Dwivedi et al., 2023; Öztürk et al., 2024).

For practitioners, these findings imply that effective AI integration should prioritize pedagogical alignment over technical sophistication alone. While AI-supported speaking tools can enhance engagement and feedback efficiency, their implementation must account for feasibility constraints such as institutional resources, teacher readiness, and curriculum compatibility. Resistance may arise when AI systems are perceived as increasing workload or threatening pedagogical autonomy. This highlights the need for scalable, low-threshold designs and teacher mediation (Chan & Hu, 2023).

Finally, the absence of ethics-related keywords in the co-citation and co-occurrence networks reveals a persistent ethical gap. Although issues such as data privacy, algorithmic bias, and fairness shape learner trust and engagement (Shin, 2020; Chiu, 2023), they remain weakly theorized. To address this gap, the integrated framework proposed in this study—linking learner engagement, technological efficacy, and pedagogical strategy (Figure 8)—explicitly positions ethical governance as a cross-cutting concern, thereby offering a more coherent basis for responsible and sustainable AI-assisted speaking instruction.

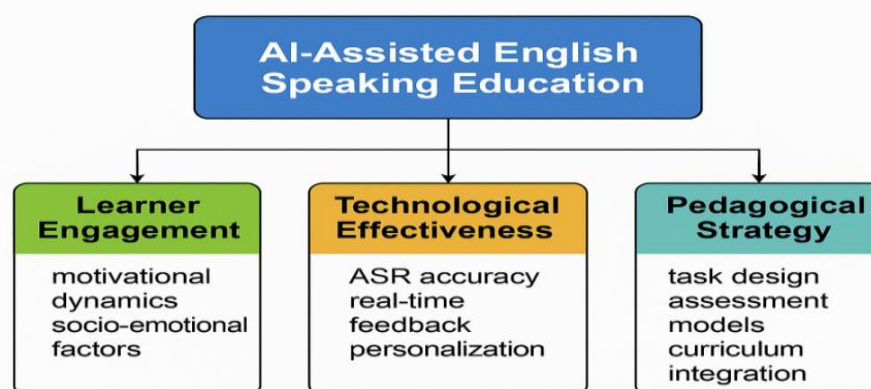


Figure 8: Core Dimensions of AI-Assisted English-speaking Education

## 5.2 Practical Implications

Based on the theoretical insights of this study, several practical strategies are proposed to support the responsible integration of AI technologies into English language instruction. These strategies address three interrelated dimensions—learner engagement, technological effectiveness, and pedagogical implementation—and are aligned with principles of sustainable development, with an emphasis on equity, ethical use, and scalability.

As a foundational step, educational institutions should implement tiered AI literacy programs for both teachers and students. Such programs should enhance technical competence while strengthening ethical awareness related to data privacy, algorithmic bias, and responsible AI use. At the basic level, learners are introduced to core concepts and ethical principles of AI-assisted English-speaking education. At the intermediate level, hands-on training with tools such as

ChatGPT and ELSA Speak can support pronunciation, fluency, and real-time feedback. At the advanced level, AI-integrated speaking tasks or collaborative peer-review projects can be used to evaluate the pedagogical value of AI in authentic communication contexts. Program effectiveness may be assessed through pre- and post-training measures, learner feedback, and usage analytics. Given the growing complexity of AI in education, universities should also promote interdisciplinary collaboration through AI-enhanced language learning laboratories and cross-disciplinary project teams. Such initiatives can align technological capabilities with instructional needs by integrating expertise from applied linguistics, educational technology, and computer science (Kasneci et al., 2023). Practical forms of collaboration include AI speaking laboratories for simulated practice and pronunciation feedback, interdisciplinary workshops on ASR tools and voice-based systems, and university–industry partnerships aimed at developing pedagogically aligned and ethically grounded AI speaking applications.

To avoid overreliance on AI and preserve learners' critical speaking abilities, institutions should further establish AI-supported academic support systems that provide personalized and formative feedback (Tejedor-García et al., 2020). These may include automated feedback tools for pronunciation and fluency, AI-based conversational assistants for low-stress practice and learning analytics dashboards that enable instructors to monitor learner progress and adapt instructional strategies accordingly. Through the combined use of AI literacy training, interdisciplinary frameworks, and targeted feedback systems, AI-assisted English-speaking education can be implemented in ways that are both pedagogically effective and ethically sound (Hwang et al., 2024).

### **5.3 Limitations, Delimitations, and Future Directions**

This study has several limitations that also inform future research directions. First, reliance on the WoS Core Collection may have excluded relevant studies indexed in databases such as Scopus or Google Scholar, limiting dataset coverage. Second, the exclusive use of bibliometric methods restricts insight into contextual, pedagogical, and affective dimensions, which the results indicate remain underexplored. Third, the selected period (2021–2025) captures only short- to mid-term developments and does not reflect longer historical trends. In addition, differences in educational contexts and technological infrastructure may limit the generalizability of the findings.

In the short term (1–2 years), future studies should prioritize mixed-method and longitudinal designs to examine how AI-supported speaking tools influence learner engagement, affective factors, and classroom practices over time, thereby addressing the fragmentation identified in this study. In the longer term (beyond five years), cross-database bibliometric analyses and large-scale comparative research across regions are needed to capture structural evolution and ethical governance more comprehensively. Practice-oriented research should also examine how institutional constraints and teacher readiness affect sustainable AI integration, helping bridge bibliometric trends and pedagogically grounded evidence.

## 6. Conclusion

This study offers a bibliometric overview of research on AI-assisted English-speaking education based on co-citation, co-occurrence, and temporal analyses. The results indicate a clear development trajectory in the field. Early studies mainly focused on speech technologies, while more recent research has shifted toward learner-centred, pedagogically informed, and affect-sensitive approaches. At the same time, the analysis reveals a persistent gap. Ethical and governance issues remain weakly integrated, despite their growing importance in AI-enhanced learning contexts. By combining structural and temporal evidence, this study clarifies how technological affordances, pedagogical design, and learner-related factors interact in AI-mediated speaking education.

The findings also underline the role of teacher preparation and institutional support. These factors are essential for ensuring that AI tools are used in meaningful, equitable, and context-sensitive ways. Future research should extend this work through longitudinal and classroom-based studies. Such research can examine how AI-supported speaking practices develop over time and how ethical, pedagogical, and learner dimensions may be better aligned. This integrated perspective highlights the transformative potential of responsible AI adoption in English-speaking education and aligns with UNESCO's Sustainable Development Goal 4, which emphasizes inclusive, equitable, and sustainable educational innovation.

## Conflict of Interest

The authors declare no conflict of interest.

## 7. Acknowledgments

The authors acknowledge the use of Grammarly and QuillBot to support English language editing (grammar and clarity) during manuscript preparation. The manuscript remains an accurate representation of the authors' work and intellectual contributions.

## 8. References

- Bashori, M., van Hout, R., Strik, H., & Cucchiarini, C. (2024). 'Look, I can speak correctly': learning vocabulary and pronunciation through websites equipped with automatic speech recognition technology. *Computer Assisted Language Learning*, 37(5-6), 1335-1363. <https://doi.org/10.1080/09588221.2022.2080230>
- Belda-Medina, J., & Calvo-Ferrer, J. R. (2022). Using chatbots as AI conversational partners in language learning. *Applied Sciences*, 12(17), 8427. <https://doi.org/10.3390/app12178427>
- Bo, Z., Pek, L. S., Cong, W., Tiannan, L., Krishnasamy, H. N., Ne'matullah, K. F., & Arar, H. (2025). Transforming translation education: A bibliometric analysis of artificial intelligence's role in fostering sustainable development. *International Journal of Learning, Teaching and Educational Research*, 24(3), 166-190. <https://doi.org/10.26803/ijlter.24.3.9>
- Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20(1), 43. <https://doi.org/10.1186/s41239-023-00411-8>

- Chen, C., Song, M., & Heo, G. (2020). Revealing the changing knowledge structure of scientometrics: A bibliometric review from 1996 to 2018. *Journal of Informetrics*, 14(2), 101006. <https://doi.org/10.1016/j.joi.2020.101006>
- Chiu, T. K. (2023). The impact of generative AI (GenAI) on practices, policies and research direction in education: A case of ChatGPT and Midjourney. *Interactive Learning Environments*, 1–17. <https://doi.org/10.1080/10494820.2023.2253861>
- Dai, Y., & Wu, Z. (2023). Mobile-assisted pronunciation learning with feedback from peers and/or automatic speech recognition: A mixed-methods study. *Computer Assisted Language Learning*, 36(5-6), 861-884. <https://doi.org/10.1080/09588221.2021.1952272>
- Dergaa, I., Chamari, K., Zmijewski, P., & Saad, H. B. (2023). From human writing to artificial intelligence generated text: Examining the prospects and potential threats of ChatGPT in academic writing. *Biology of Sport*, 40(2), 615-622. <https://doi.org/10.5114/biolSport.2023.125623>
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., ... & Wright, R. (2023). Opinion paper: “So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>
- El Shazly, R. (2021). Effects of artificial intelligence on English speaking anxiety and speaking performance: A case study. *Expert Systems*, 38(3), e12667. <https://doi.org/10.1111/exsy.12667>
- Evers, K., & Chen, S. (2022). Effects of an automatic speech recognition system with peer feedback on pronunciation instruction for adults. *Computer Assisted Language Learning*, 35(8), 1869-1889. <https://doi.org/10.1080/09588221.2020.1839504>
- Fathi, J., Rahimi, M., & Derakhshan, A. (2024). Improving EFL learners’ speaking skills and willingness to communicate via artificial intelligence-mediated interactions. *System*, 121, Article 103254. <https://doi.org/10.1016/j.system.2024.103254>
- Fouz-González, J. (2020). Using apps for pronunciation training: An empirical evaluation of the English File Pronunciation app. *Language Learning & Technology*, 24(1), 62–85. <https://doi.org/10.10125/44709>
- Gilson, A., Safranek, C. W., Huang, T., Socrates, V., Chi, L., Taylor, R. A., & Chartash, D. (2023). How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. *JMIR Medical Education*, 9(1), e45312. <https://doi.org/10.2196/45312>
- Hsu, M. H., Chen, P. S., & Yu, C. S. (2023). Proposing a task-oriented chatbot system for EFL learners speaking practice. *Interactive Learning Environments*, 31(7), 4297-4308. <https://doi.org/10.1080/10494820.2021.1960864>
- Huang, F., & Zou, B. (2024). English speaking with artificial intelligence (AI): The roles of enjoyment, willingness to communicate with AI, and innovativeness. *Computers in Human Behavior*, 159, Article 108355. <https://doi.org/10.1016/j.chb.2024.108355>
- Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning – Are they really useful? A systematic review of chatbot-supported language learning. *Journal of computer assisted learning*, 38(1), 237-257. <https://doi.org/10.1111/jcal.12610>
- Hwang, W. Y., Guo, B. C., Hoang, A., Chang, C. C., & Wu, N. T. (2024). Facilitating authentic contextual EFL speaking and conversation with smart mechanisms and investigating its influence on learning achievements. *Computer Assisted Language Learning*, 37(7), 1632-1658. <https://doi.org/10.1080/09588221.2022.2095406>

- Jeon, J. (2024). Exploring AI chatbot affordances in the EFL classroom: Young learners' experiences and perspectives. *Computer Assisted Language Learning*, 37(1-2), 1-26. <https://doi.org/10.1080/09588221.2021.2021241>
- Jiang, M. Y. C., Jong, M. S. Y., Lau, W. W. F., Chai, C. S., & Wu, N. (2021). Using automatic speech recognition technology to enhance EFL learners' oral language complexity in a flipped classroom. *Australasian Journal of Educational Technology*, 37(2), 110-131. <https://doi.org/10.14742/ajet.6798>
- Kasneci, E., Sessler, K., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Horvitz, E., ... & Schmidt, A. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, Article 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., ... & Tseng, V. (2023). Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLoS Digital Health*, 2(2), e0000198. <https://doi.org/10.1371/journal.pdig.0000198>
- Lim, S. P., Che Yob, F. S. S., Wong, R. M. M., & Camara, J. S. (2025). Mobile gaming in education: A bibliometric analysis of trends and performance. *International Journal of Evaluation and Research in Education*, 14(4), 2676. <https://doi.org/10.11591/ijere.v14i4.32991>
- Ma, 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*, 12, Article 163. <https://doi.org/10.1057/s41599-025-04688-0>
- Mohamed, A. M. (2024). Exploring the potential of an AI-based Chatbot (ChatGPT) in enhancing English as a Foreign Language (EFL) teaching: Perceptions of EFL faculty members. *Education and Information Technologies*, 29(3), 3195-3217. <https://doi.org/10.1007/s10639-023-11917-z>
- Oruh, J., Viriri, S., & Adegun, A. (2022). Long short-term memory recurrent neural network for automatic speech recognition. *IEEE Access*, 10, 30069-30079. <https://doi.org/10.1109/ACCESS.2022.3159339>
- Öztürk, O., Kocaman, R., & Kanbach, D. K. (2024). How to design bibliometric research: An overview and a framework proposal. *Review of Managerial Science*, 18(11), 3333-3361. <https://doi.org/10.1007/s11846-024-00738-0>
- Rahimi, M., & Fathi, J. (2024). Employing e-tandem language learning method to enhance speaking skills and willingness to communicate: The case of EFL learners. *Computer Assisted Language Learning*, 37(4), 924-960. <https://doi.org/10.1080/09588221.2022.2064512>
- Rogerson-Revell, P. M. (2021). Computer-assisted pronunciation training (CAPT): Current issues and future directions. *RELC Journal*, 52(1), 189-205. <https://doi.org/10.1177/0033688220977406>
- Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching*, 6(1), 342-363. <https://doi.org/10.37074/jalt.2023.6.1.9>
- Sallam, M. (2023). ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns. In *Healthcare*, 11(6), Article 887. <https://doi.org/10.3390/healthcare11060887>
- Shin, D. (2020). User perceptions of algorithmic decisions in the personalized AI system: Perceptual evaluation of fairness, accountability, transparency, and explainability. *Journal of Broadcasting & Electronic Media*, 64(4), 541-565. <https://doi.org/10.1080/08838151.2020.1843357>
- Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students.

- Frontiers in Psychology*, 14, Article 1260843.  
<https://doi.org/10.3389/fpsyg.2023.1260843>
- Tai, T. Y., & Chen, H. H. J. (2024). Improving elementary EFL speaking skills with generative AI chatbots: Exploring individual and paired interactions. *Computers & Education*, 220, 105112. c
- Tai, T. Y., & Chen, H. H. J. (2023). The impact of Google Assistant on adolescent EFL learners' willingness to communicate. *Interactive Learning Environments*, 31(3), 1485-1502. <https://doi.org/10.1080/10494820.2020.1841801>
- Tejedor-García, C., Escudero-Mancebo, D., Cámara-Arenas, E., González-Ferreras, C., & Cardeñoso-Payo, V. (2020). Assessing pronunciation improvement in students of English using a controlled computer-assisted pronunciation tool. *IEEE Transactions on Learning Technologies*, 13(2), 269-282. <https://doi.org/10.1109/TLT.2020.2980261>
- UNESCO. (2021). *AI and education: Guidance for policymakers*. United Nations Educational, Scientific and Cultural Organization. <https://doi.org/10.54675/PCSP7350>
- Van Eck, N. J., & Waltman, L. (2017). Citation-based clustering of publications using CitNetExplorer and VOSviewer. *Scientometrics*, 111(2), 1053-1070. <https://doi.org/10.1007/s11192-017-2300-7>
- Wider, W., Jiang, L., Lin, J., Fauzi, M. A., Li, J., & Chan, C. K. (2024). Metaverse chronicles: A bibliometric analysis of its evolving landscape. *International Journal of Human-Computer Interaction*, 40(17), 4873-4886. <https://doi.org/10.1080/10447318.2023.2227825>
- Yang, H., Kim, H., Lee, J. H., & Shin, D. (2022). Implementation of an AI chatbot as an English conversation partner in EFL speaking classes. *ReCALL*, 34(3), 327-343. <https://doi.org/10.1017/s0958344022000039>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – Where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1), 1-27. <https://doi.org/10.1186/s41239-019-0171-0>
- Zhou, B., Lim, S. P., Kabir, N., Asna, A., & Lu, T. N. (2026). From automation to socio-affective learning: A bibliometric analysis of artificial intelligence in English as a foreign language education. *Problems of Education in the 21st Century*, 84(1), 229-247. <https://doi.org/10.33225/pec/26.84.229>
- Zhou, Q., Hashim, H., & 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*, 30(1), 1-22. <https://doi.org/10.1007/s10639-025-13401-2>
- Zou, B., Guan, X., Shao, Y., & Chen, P. (2023a). Supporting speaking practice by social network-based interaction in artificial intelligence (AI)-assisted language learning. *Sustainability*, 15(4), 2872. <https://doi.org/10.3390/su15042872>
- Zou, B., Xie, S., & Wang, C. (2025a). Students' willingness to communicate (WTC) in using artificial intelligence (AI) technology in English-speaking practice. *International Journal of Information and Communication Technology Education*, 21(2), 1-14. <https://doi.org/10.4018/IJICTE.375387>
- Zou, B., Wang, C., Yan, Y., Du, X., & Ji, Y. (2025b). Exploring English as a foreign language learners' adoption and utilisation of ChatGPT for Speaking practice through an extended technology acceptance model. *International Journal of Applied Linguistics*, 35(2), 689-704. <https://doi.org/10.1111/ijal.12658>
- Zou, B., Lyu, Q., Han, Y., Li, Z., & Zhang, W. (2025c). Exploring students' acceptance of an artificial intelligence speech evaluation program for EFL speaking practice: an application of the Integrated Model of Technology Acceptance. *Computer Assisted Language Learning*, 38(5-6), 1366-1391. <https://doi.org/10.1080/09588221.2023.2278608>