

AI Voice Avatar as an ESP Speaking Partner: An Autoethnographic Exploration of Business English Vocabulary Acquisition

Nur Fatimah¹,

^{1,2}English Education Department, Postgraduate Program,
Universitas Muhammadiyah Gresik, Indonesia
fatimah000023@gmail.com

Dwita Laksmi Rachmawati²

²English Education Department, Postgraduate Program,
Universitas Muhammadiyah Gresik, Indonesia

Received: 21st June 2026

Accepted: 9th July 2026

Published: 30th Sept 2026

Abstract

This autoethnographic study investigates how an AI voice avatar functions as an English for Specific Purposes (ESP) speaking partner for Business English vocabulary acquisition. Specifically, this study used Sesame AI, a web-based AI voice-avatar application, as the speaking partner for simulated professional conversations. It aims to examine the researcher-participant's lived experience of using an AI voice avatar for professional speaking practice, to understand how AI-supported interaction facilitates contextual vocabulary learning, and to identify the business-related vocabulary that became more usable through repeated practice. The study employed a self-reflection design grounded in autoethnography. Data were generated across six speaking sessions over four weeks using a treatment log, vocabulary learning log, reflective journal, and critical incident reflection. The data were analyzed thematically by tracing patterns of vocabulary noticing, repetition, contextual use, confidence, autonomy, and technological constraints. The findings show that Sesame AI supported retention of Business English lexical chunks, encouraged autonomous rehearsal, and strengthened speaking self-efficacy by providing a low-pressure conversational environment. However, the interaction also revealed pragmatic limitations, including occasional generic responses, imperfect speech recognition, and limited emotional sensitivity. The study contributes a reflective learner-centered perspective to ESP and CALL by showing how web-based AI voice avatars may serve as supplementary speaking partners for self-directed professional language learning when supported by critical reflection and human pedagogical judgment.

Keywords: AI voice avatar; Sesame AI; Business English; English for Specific Purposes; vocabulary acquisition; autoethnography

Introduction

Business English vocabulary is a central resource for professional communication because workplace interaction requires more than general grammatical competence. Professionals are expected to explain positions, responsibilities, organizational services, customer profiles, schedules, and institutional values through precise lexical choices, appropriate collocations, and context-sensitive formulaic expressions. Within English for Specific Purposes (ESP), language instruction is therefore expected to be aligned with specific academic, professional, or occupational needs rather than treated as a decontextualized body of grammar and vocabulary (Dudley-Evans & St John, 1998; Hutchinson & Waters, 1987). In recent ESP work, AI-supported modules for discipline-specific learners also show that technology can support the design of specialized materials for fields such as accounting and business-related communication (Kurnia et al., 2025).

A persistent problem for adult EFL learners is the limited availability of suitable speaking partners. Classroom time may be restricted, peer interaction may not fully represent professional situations, and many learners hesitate to speak because they fear making errors in front of teachers, colleagues, or native speakers. This challenge is repeatedly noted in research on AI-assisted English learning, where conversational systems are valued because they provide individualized, repeated, and low-pressure practice opportunities (Crompton et al., 2024; Du & Daniel, 2024; Lai & Lee, 2024). In the Business English context, this issue is more specific: learners do not merely need someone to speak with; they need a partner who can sustain simulated workplace exchanges involving professional roles, organizational descriptions, service explanations, and customer-oriented discourse.

AI voice avatars provide a possible response to this need. Unlike text-only chatbots, AI voice avatars combine generative conversation, synthetic speech, and often visual presence, allowing learners to hear model language,



respond orally, and rehearse professional scenarios in a way that resembles spoken interaction. Recent studies and reviews show that generative AI can serve as a feedback provider, learning tutor, cognitive stimulator, interaction facilitator, and conversation partner, although its implementation also raises concerns about feedback quality, academic integrity, overreliance, and pedagogical limitations (Deng S Jamaludin, 2026; Lee, Choe, Zou, S Jeon, 2026). AI-enabled assistants may also provide adaptive learning pathways and instant support for learners in higher education (Sajja et al., 2024).

However, much of the existing research on generative AI in language education focuses on general EFL classrooms, academic writing, broad learner perceptions, or quantitative outcomes. Less attention has been given to the inner learning experience of a single adult learner using an AI voice avatar to acquire Business English vocabulary. Existing work on GenAI and learner voice emphasizes the importance of agency, authenticity, and critical AI literacy (Chan S Hu, 2023; Kim et al., 2025; Smirnova, 2025; Sterling et al., 2025; Wang, 2025), but the affective and cognitive micro-processes of noticing, rehearsing, retaining, and using Business English vocabulary through voice-avatar interaction remain underexplored. This study addresses that gap through autoethnographic self-reflection.

The study is guided by three research questions: (1) How does the researcher-participant experience the use of an AI voice avatar as an ESP speaking partner for Business English practice? (2) How does the AI voice avatar support the noticing, practice, and contextual use of Business English vocabulary? (3) What Business English vocabulary items and expressions become more usable through AI voice-avatar speaking practice, and what challenges appear during the process?

English for Specific Purposes and Business English Vocabulary

ESP is a needs-based approach to language learning in which content, tasks, and communicative objectives are shaped by learners' target situations (Dudley-Evans S St John, 1998; Hutchinson S Waters, 1987). Business English is one branch of ESP because it prepares learners to communicate in professional and organizational settings. Its vocabulary includes not only technical terms but also semi-technical words, collocations, and lexical chunks such as target market, professional background, student advisory department, high-quality educational services, and oversee students' goals and achievements. These expressions carry professional meanings that are difficult to master through isolated memorization.

For ESP learners, vocabulary acquisition is strengthened when words are embedded in meaningful communicative tasks. A learner who repeatedly uses the phrase "I oversee students' academic goals and achievements" in simulated professional conversation is more likely to understand its functional and pragmatic value than a learner who only translates the word "oversee." This view is consistent with recent work on GenAI-assisted multimodal ESP pedagogy, which argues that ESP should move beyond passive reception of teacher-prepared materials and support learners as active designers of professional meaning (Christoforou, 2026).

Generative AI, Conversational AI, Personalization, and Learner Autonomy

Generative AI has expanded the possibilities of language learning by enabling contextually responsive interaction, instant feedback, and flexible task generation. Systematic reviews indicate that GenAI has been used in language education as a tutor, feedback provider, interaction facilitator, and conversation partner, with higher education and English learning as dominant contexts (Lee, Choe, Zou, S Jeon, 2026). In EFL instruction, GenAI has also been shown to play multiple roles across writing, speaking, grammar, reading, and assessment, although the literature remains uneven and teacher feedback is still preferred in many situations (Deng S Jamaludin, 2026).

Conversational AI research in ELT shows positive outcomes in affective and cognitive domains, especially in Asian EFL contexts where learners often have limited opportunities to use English outside the classroom (Lai S Lee, 2024). Reviews focusing specifically on AI chatbots for English speaking practice report benefits for speaking confidence, engagement, motivation, pronunciation, and anxiety reduction (Du S Daniel, 2024). Similarly, an AI-powered conversational system for college students learning English as a second language produced higher engagement, richer vocabulary use, and greater speaking test gains compared with a traditional listen-and-repeat interface (Dai S Wu, 2025).

AI-enabled personalized assistants are relevant to this study because they can generate adaptive learning pathways, provide instant responses, and support self-directed learning (Sajja et al., 2024). In ESP, AI-powered personalization can help learners focus on professional goals, select scenario-specific vocabulary, and rehearse at their own pace (Mansor, 2025). Such personalization is particularly valuable in Business English, where learners need repeated practice with situation-specific expressions rather than generic conversational fluency.

Synthetic Speech, Voice Assistants, and Affective Experience

The voice dimension of AI-mediated learning is especially important because spoken interaction involves timing, intonation, stress, pause, rhythm, and perceived social presence. Research on voice assistant conversation design demonstrates that linguistic features influence user preferences and the perceived naturalness of voice interaction (Lee, Sargsyan, Choi, S Kim, 2026). Studies on AI voice assistants also show that language style should fit the service context; concrete language may be more effective in utilitarian contexts, while abstract language may be preferred in hedonic contexts because the match influences processing fluency and continued use intentions (Lan et al., 2024).

Text-to-speech (TTS) tools also support pronunciation learning by allowing learners to listen, imitate, pause, and repeat target expressions. In pronunciation learning, TTS can provide models for segmental features, such as individual sounds, and suprasegmental features, such as stress and intonation (Amin, 2024). In the present study, the AI voice avatar's synthetic speech created opportunities for shadowing-like rehearsal: the learner could notice the rhythm of phrases such as high-quality educational services, middle-to-upper-income families, and operate Monday to Saturday, then reuse them orally.

At the same time, synthetic speech cannot be viewed as neutral. Studies of AI-generated voice and media show that human voices may still produce stronger cognitive effects than AI voices, suggesting that trust, fluency, and perceived credibility are shaped by voice quality and affective resonance (Gong, 2023). This is relevant to AI voice avatars in ESP: even when an avatar provides useful language input, its tone, prosody, and pragmatic sensitivity may influence whether learners experience the interaction as supportive, mechanical, or socially meaningful.

Learner Agency, Self-Efficacy, and Ethical Considerations

The learner's sense of agency is central to AI-assisted language learning. Studies of GenAI in higher education show that learners often perceive AI as a source of personalized support, brainstorming assistance, writing support, and research guidance, while also expressing concerns about accuracy, privacy, and ethical use (Chan S Hu, 2023). In EAP and writing contexts, structured AI use can enhance student agency and confidence, whereas unstructured AI use may increase anxiety or weaken learning outcomes (Smirnova, 2025). Research on GenAI-assisted writing similarly emphasizes the need to preserve authentic learner voice and avoid uncritical reliance on AI output (Kim et al., 2025; Sterling et al., 2025; Wang, 2025).

Self-efficacy theory is useful for interpreting speaking confidence in this study. Bandura (1997) argues that successful experiences can strengthen learners' beliefs in their ability to perform a task. For an adult EFL learner practicing Business English, repeated low-risk interaction with an AI avatar may function as a mastery experience because the learner can try, fail, reformulate, and repeat without immediate social judgment. However, ethical and pedagogical concerns remain. AI tools may generate generic responses, reproduce linguistic bias, standardize language norms, produce inaccurate feedback, or encourage overreliance (Choi, 2025; Crompton et al., 2024; Lee, Choe, Zou, S Jeon, 2026). Therefore, AI voice avatars should be framed as guided practice tools rather than replacements for teachers, peers, or professional human feedback.

Method Research

Design

This study employed a qualitative self-reflection design using autoethnographic inquiry. Autoethnography treats personal experience as a legitimate source of qualitative data while connecting that experience to broader educational, social, and technological issues (Ellis et al., 2011). The approach was appropriate because the purpose was not to measure the effectiveness of an AI tool experimentally, but to understand how the researcher-participant experienced vocabulary learning through repeated interaction with an AI voice avatar.

The analysis followed thematic analysis, which supports systematic identification of patterns across qualitative data (Braun S Clarke, 2006). The unit of analysis included not only the spoken task itself but also the meanings constructed through reflection after each session.

Participant, Context, AI Tool, and Duration

The participant was the researcher, an adult EFL learner who needed Business English for professional communication. The practice context focused on workplace-oriented speaking tasks, including professional self-introduction, describing an organization, explaining work responsibilities, identifying customers and target markets, describing service schedules, and responding to business-related questions.

The AI tool used in this study was Sesame AI (<https://app.sesame.com/>), a web-based AI voice-avatar application used as a supplementary speaking partner. In the treatment, Sesame AI functioned as the voice-avatar platform through which the researcher-participant conducted simulated ESP and Business English speaking practice. The platform provided voice-based interaction, synthetic speech input, and opportunities for oral

rehearsal of workplace-oriented vocabulary in a low-pressure environment. The treatment consisted of six speaking sessions over four weeks. Each session required the researcher-participant to speak, listen, respond, record vocabulary, and reflect on the learning experience. Table 1 presents the speaking treatment.

Table 1. Overview of the AI Voice-Avatar Speaking Treatment

Session	Business English Focus	Speaking Task	Expected Vocabulary Focus
1	Professional self-introduction	Introducing background, position, and department	position, professional background, lead teacher, student advisory
2	Describing an organization	Explaining the education-sector organization and its services	education sector, high-quality services, platform, academic goals
3	Customer and target market	Describing customers and target market characteristics	customers, elementary students, high school students, urban areas, middle-to-upper-income families
4	Responsibilities and achievements	Explaining main responsibilities and student outcomes	oversee, achievement, learning goals, student progress, support
5	Service operation and schedule	Explaining operational days and service hours	operate, Monday to Saturday, schedule, service hours, availability
6	Integrated role-play and reflection	Responding to a simulated professional conversation	reviewed vocabulary from Sessions 1-5; fluency and appropriateness

Instruments and Data Analysis

Four instruments were used: a treatment log, a vocabulary learning log, a reflective journal, and a critical incident reflection. The treatment log documented the session focus, prompt, duration, topic, AI response pattern, and speaking task completion. The vocabulary learning log recorded target words and phrases, meanings, pronunciation notes, example use, and repeated use. The reflective journal captured feelings, confidence, anxiety, perceived improvement, and strategy use. The critical incident reflection recorded meaningful moments such as misrecognition, misunderstanding, successful reformulation, or increased confidence.

Data analysis was conducted in five stages. First, all logs and reflections were read repeatedly. Second, initial codes were generated, including vocabulary anxiety, lexical chunk noticing, repeated retrieval, pronunciation awareness, low-pressure speaking, generic AI response, and pragmatic limitation. Third, related codes were grouped into themes. Fourth, the themes were reviewed against the original data. Fifth, the themes were interpreted in relation to ESP vocabulary acquisition, CALL, and AI-assisted speaking literature.

Table 2. Data Sources and Analytical Purpose

Instrument	Main Data Recorded	Analytical Function
Treatment log	Session focus, prompt, duration, topic, AI response, speaking task completion	To reconstruct the learning process and compare sessions
Vocabulary learning log	Target words/phrases, meaning, pronunciation, example use, repeated use	To trace vocabulary noticing, uptake, and contextual use
Reflective journal	Feelings, confidence, anxiety, perceived improvement, strategy use	To identify affective and metacognitive patterns
Critical incident reflection	Meaningful learning moments, misunderstandings, technical issues, breakthrough experiences	To analyze benefits and constraints of AI-mediated practice

Trustworthiness and Ethics

Trustworthiness was strengthened through triangulation across four data sources, an audit trail of session logs and vocabulary records, and reflexive acknowledgment of the researcher-participant's subjectivity. Negative cases and limitations were included to avoid presenting AI as entirely beneficial.

Because the study used self-generated data, ethical risk was relatively low. Nevertheless, privacy and academic integrity were considered carefully. Sensitive personal information was excluded, and AI-generated responses were treated as learning materials rather than human research-participant data. The study also recognizes broader

concerns about data privacy, bias, feedback quality, and overreliance in AI-mediated language learning (Chan S Hu, 2023; Choi, 2025; Crompton et al., 2024).

Result and Discussion

Initial Business English Speaking Needs and Vocabulary Profile

The first theme shows that the researcher-participant initially experienced anxiety when trying to explain professional identity and responsibilities in Business English. The difficulty was not a lack of basic English, but the absence of precise lexical chunks suitable for professional self-representation. The participant could say “I manage students,” but this expression felt too general and less professional than “I oversee students’ academic goals and achievements.” This distinction reflects a key ESP issue: learners need vocabulary that is not merely correct but contextually appropriate for their target situation.

This finding is consistent with ESP theory, which emphasizes learners’ specific communicative needs (Dudley-Evans S St John, 1998; Hutchinson S Waters, 1987). It also aligns with recent AI-in-ELT reviews showing that AI can support speaking, pedagogy, and self-regulation, particularly when learners need practice beyond limited classroom opportunities (Crompton et al., 2024; Lai S Lee, 2024). In this study, the AI voice avatar became useful when prompts were explicitly connected to workplace tasks.

AI Voice Avatar Supported Vocabulary Noticing and Repeated Practice

The second theme concerns vocabulary noticing. Through repeated interaction, the participant became more aware of professional lexical chunks such as high-quality educational services, target market, student advisory department, and middle-to-upper-income families. The AI voice avatar supported this noticing process because the learner encountered vocabulary through spoken prompts, responses, and reformulations rather than through word lists only. This reflective finding is also aligns with Santoso et al. (2025), who reported that AI-assisted speaking practice enhanced pronunciation, fluency, vocabulary use, and communicative confidence through repeated interaction and immediate feedback.

The TTS quality of the avatar supported pronunciation awareness. The learner could listen to the rhythm, stress, and intonation of target expressions and then repeat them. This resembles shadowing-based practice because the learner used the model voice as a basis for oral rehearsal. Studies on AI text-to-speech suggest that TTS tools can support independent pronunciation practice by helping learners attend to both segmental and suprasegmental features (Amin, 2024). The finding also connects with research on conversational AI and speaking practice, which reports that AI tools can improve confidence, motivation, and oral practice opportunities (Du S Daniel, 2024; Dai S Wu, 2025).

Table 3. Vocabulary Development Display Based on the Learning Log

Vocabulary Item / Phrase	Type	Context of Use	Learning Indication
lead teacher	Professional role phrase	Introducing current position	Reinforced because it was repeatedly used in self-introduction
student advisory department	Workplace noun phrase	Explaining department or division	Clarified the professional context of work
oversee students’ goals and achievements	Verb phrase / responsibility expression	Describing responsibilities	Improved accuracy compared with simpler expressions such as “manage students”
high-quality educational services	Business/service phrase	Describing organizational value	Added formality and professionalism to service description
target market	Business term	Describing customers	Supported clearer explanation of customer segmentation
middle- to- upper- income families	Market-segmentation phrase	Describing customer profile	Improved precision in discussing market characteristics
operate Monday to Saturday	Operational schedule expression	Explaining service hours	Improved fluency in describing institutional operation

Vocabulary Became Stronger Through Contextual Speaking Tasks

The third theme shows that vocabulary became stronger when used in contextual speaking tasks. In the final integrated role-play, the participant reused vocabulary from previous sessions to answer follow-up questions

about role, department, organizational services, customer profile, responsibilities, and schedule. Vocabulary learning therefore moved from recognition to contextual production. The AI avatar's value was not simply that it "gave vocabulary," but that it provided a conversational environment in which the learner had to retrieve and adapt vocabulary in response to prompts.

This finding supports the view that GenAI can function as an interaction facilitator and conversation partner (Lee, Choe, Zou, S Jeon, 2026). It also resonates with studies of learner agency in AI-mediated learning: structured AI use can encourage students to become more active, confident, and purposeful in completing language tasks (Smirnova, 2025). In addition, research on AI-assisted writing shows that learners often position AI as a virtual tutor or digital peer, but they still need to maintain their own voice and judgment (Kim et al., 2025; Sterling et al., 2025; Wang, 2025). In the present speaking context, the avatar similarly functioned as a rehearsal partner rather than an authoritative instructor.

Confidence Improved, but Pragmatic and Affective Challenges Remained

The fourth theme relates to affective experience. The participant reported greater willingness to speak after repeated practice because the avatar created a low-pressure environment. The absence of direct human judgment allowed the participant to experiment with vocabulary, pause, reformulate, and try again. This reflects Bandura's (1997) self-efficacy theory: successful repeated experiences can strengthen learners' belief in their ability to perform a communicative task.

Nevertheless, the AI voice avatar also produced challenges. Some responses were generic and did not always push the participant to elaborate. At times, the avatar's tone felt insufficiently empathic, and speech recognition did not fully capture pronunciation. This finding is important because voice-based interaction depends not only on content accuracy but also on social and affective perception. Research on voice assistants indicates that linguistic features influence user preferences and perceived naturalness (Lee, Sargsyan, Choi, S Kim, 2026), while language style should match the communicative service context to support processing fluency and continued use (Lan et al., 2024). Furthermore, synthetic voices may differ from human voices in cognitive and affective impact (Gong, 2023).

The findings also highlight the need for responsible integration. Reviews of GenAI in language education emphasize concerns about content quality, querying problems, feedback limitations, academic integrity, and overreliance (Deng S Jamaludin, 2026; Lee, Choe, Zou, S Jeon, 2026). Students themselves often recognize both the benefits and risks of GenAI, including personalization, efficiency, privacy concerns, and ethical uncertainty (Chan S Hu, 2023). Thus, while the AI voice avatar supported autonomy and confidence, human pedagogical judgment remains necessary for evaluating pragmatic appropriateness, professional tone, and ethical use.

Table 4. Thematic Summary of Findings

Theme	Data Source	Interpretation
Professional vocabulary needs	Treatment log; reflective journal	Speaking tasks revealed the need for more precise Business English expressions.
Vocabulary noticing and retrieval	Vocabulary learning log; treatment log	The AI avatar provided exposure and repeated opportunities to reuse professional vocabulary.
Contextual vocabulary use	Role-play session; reflective journal	Vocabulary became more meaningful when used in simulated professional conversation.
Confidence and constraints	Reflective journal; critical incident reflection	The AI environment reduced pressure but also produced generic responses and occasional misrecognition.

Synthesis of the Findings

Taken together, the findings suggest that AI voice-avatar interaction can support Business English vocabulary acquisition through three connected mechanisms: repeated exposure, contextual retrieval, and affective safety. Repetition helped the learner remember lexical chunks; contextual role-play helped transform vocabulary into usable professional expressions; and the low-pressure environment supported self-efficacy. The avatar therefore acted as a form of interactive scaffolding: it did not replace the teacher, but it created a space in which the learner could rehearse, notice, and revise language before using it in human interaction.

This study also adds nuance to CALL and ESP scholarship. In CALL, the finding supports the growing view that conversational AI can provide individualized, interactive, and immersive practice, but it also shows that learners require critical awareness of AI limitations (Crompton et al., 2024; Lai S Lee, 2024). In ESP, the study demonstrates that AI-supported practice is most valuable when prompts are tightly connected to target professional situations. A generic avatar conversation may produce generic language; a well-designed Business English scenario can generate more useful vocabulary and reflection.

Conclusion

This autoethnographic study examined the use of an AI voice avatar as an ESP speaking partner for Business English vocabulary acquisition. The findings show that the avatar supported vocabulary development by combining repetition, contextual speaking practice, and affective presence. The participant became more aware of professional expressions, reused lexical chunks across sessions, and reported increased confidence in speaking. Vocabulary items such as lead teacher, student advisory department, high-quality educational services, target market, middleto-upper-income families, and operate Monday to Saturday became more usable because they were practiced in meaningful professional scenarios.

The study contributes to ESP and CALL by providing a first-person account of the psychological and pedagogical experience of learning professional vocabulary through AI-mediated voice interaction. It shows that AI voice avatars can function as supplementary speaking partners, especially for adult EFL learners who lack access to human interlocutors for Business English practice. However, the study also confirms that AI should not be treated as an autonomous replacement for human teaching. Pragmatic nuance, professional appropriateness, ethical awareness, and emotional sensitivity still require human guidance.

The main limitation of this study is its single-participant autoethnographic design. The findings are not statistically generalizable and are shaped by the researcher-participant's reflective interpretation. Future research should involve larger groups of ESP learners, longer treatment periods, recorded speaking data, independent assessment of vocabulary gains, and comparisons between AI voice-avatar practice and human peer practice. Further studies may also examine how avatar voice quality, feedback style, emotional expression, and visual embodiment influence learner engagement, vocabulary retention, pronunciation development, and self-efficacy.

References

- Amin, E. A.-R. (2024). EFL students' perception of using AI text-to-speech apps in learning pronunciation. *Migration Letters*, 21(3), 887–903.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.
- Braun, V., S Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Chan, C. K. Y., S Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20, Article 43. <https://doi.org/10.1186/s41239-023-00411-8>
- Choi, S. (2025). Ethical issues of using AI in language education: A critical perspective. *English Teaching*, 80(5), 229–248. <https://doi.org/10.15858/engtea.80.5.202512.229>
- Christoforou, M. (2026). Gen AI-assisted multimodal meaning design: Exercising a pedagogic metalanguage of transposition. *Pedagogies: An International Journal*, 21(2), 276–300. <https://doi.org/10.1080/1554480X.2025.2522884>
- Crompton, H., Edmett, A., Ichaporia, N., S Burke, D. (2024). AI and English language teaching: Affordances and challenges. *British Journal of Educational Technology*, 55 (6), 2503–2529. <https://doi.org/10.1111/bjet.13460>
- Dai, L., S Wu, F. (2025). An AI-powered conversational system for college students learning English as a second language. *Education and Information Technologies*, 30, 23393–23417. <https://doi.org/10.1007/s10639-025-13640-3>
- Deng, L., S Jamaludin, K. A. (2026). Roles of generative artificial intelligence (GenAI) in English as a foreign language (EFL) instruction: A systematic literature review. *SAGE Open*, 16(1), 1–17. <https://doi.org/10.1177/21582440261418315>
- Ding, D., S Muhyiddin B Yusof, A. (2025). Investigating the role of AI-powered conversation bots in enhancing L2 speaking skills and reducing speaking anxiety: A mixed methods study. *Humanities and Social Sciences Communications*, 12, Article 1223. <https://doi.org/10.1057/s41599-025-05550-z>
- Du, J., S 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, Article 100230. <https://doi.org/10.1016/j.caeai.2024.100230>
- Dudley-Evans, T., S St John, M. J. (1998). *Developments in English for specific purposes: A multi-disciplinary approach*. Cambridge University Press.

- Ellis, C., Adams, T. E., S Bochner, A. P. (2011). Autoethnography: An overview. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 12(1), Article 10.
- Gong, C. (2023). AI voices reduce cognitive activity? A psychophysiological study of the media effect of AI and human newscasts in Chinese journalism. *Frontiers in Psychology*, 14, <https://doi.org/10.3389/fpsyg.2023.1243078>
- Hutchinson, T., S Waters, A. (1987). *English for specific purposes: A learning-centred approach*. Cambridge University Press.
- Kim, J., Yu, S., Detrick, R., S Li, N. (2025). Exploring students' perspectives on generative AI-assisted academic writing. *Education and Information Technologies*, 30, 1265–1300. <https://doi.org/10.1007/s10639-024-12878-7>
- Kurnia, L. I., Setyawan, B., Khoirunisa, R., Mustofa, A., Dja'far, V. H., S Purnomo, S. H. (2025). AI-based English module for accounting students. *Transformational Language, Literature, and Technology Overview in Learning*, 5(1).
- Lai, W. Y. W., S Lee, J. S. (2024). A systematic review of conversational AI tools in ELT: Publication trends, tools, research methods, learning outcomes, and antecedents. *Computers and Education: Artificial Intelligence*, 7, Article 100291. <https://doi.org/10.1016/j.caeai.2024.100291>
- Lan, H., Tang, X., Ye, Y., S Zhang, H. (2024). Abstract or concrete? The effects of language style and service context on continuous usage intention for AI voice assistants. *Humanities and Social Sciences Communications*, 11, Article 99. <https://doi.org/10.1057/s41599-024-02600-w>
- Lee, S., Choe, H., Zou, D., S Jeon, J. (2026). Generative AI (GenAI) in the language classroom: A systematic review. *Interactive Learning Environments*, 34(1), 335–359. <https://doi.org/10.1080/10494820.2025.2498537>
- Lee, Y., Sargsyan, L., Choi, S., S Kim, S.-H. (2026). Exploring user perceptions and preferences in voice assistant conversation design: The role of linguistic features. *International Journal of Human-Computer Interaction*, 42(4), 2524–2541. <https://doi.org/10.1080/10447318.2025.2530058>
- Mansor, N. S. (2025). AI-powered personalization in ESP: Enhancing learner autonomy and engagement in English for professional contexts. *International Journal of Research and Innovation in Social Science*, 9(10), 2239–2254. <https://doi.org/10.47772/IJRISS.2025.91100178>
- Sajja, R., Sermet, Y., Cikmaz, M., Cwiertny, D., S Demir, I. (2024). Artificial intelligence-enabled intelligent assistant for personalized and adaptive learning in higher education. *Information*, 15(10), Article 596. <https://doi.org/10.3390/info15100596>
- Santoso, M., Anwar, K., S Maruf, N. (2025). Artificial intelligence on EFL in vocational high school: The impact of Talkpal.AI on speaking skill. *Indonesian EFL Journal*, 11(3), 625–638. <https://doi.org/10.25134/991nty70>
- Sesame AI. (n.d.). *Sesame AI* [Web application]. <https://app.sesame.com/>
- Smirnova, L. (2025). Developing students' agency and voice by using generative AI in an online EAP module. *Innovation in Language Learning and Teaching*. <https://doi.org/10.1080/17501229.2025.2538781>
- Sterling, L., Ye, C., Ying, H., S Chen, Z. (2025). Finding your voice: Using generative AI to help international students improve their writing. *Information*, 16(4), Article 289. <https://doi.org/10.3390/info16040289>
- Wang, C. (2025). Exploring students' generative AI-assisted writing processes: Perceptions and experiences from native and nonnative English speakers. *Technology, Knowledge and Learning*, 30, 1825–1846. <https://doi.org/10.1007/s10758-024-09744-3>