ORIGINAL ARTICLE

INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE

ISSN: 2692-5206, Impact Factor: 12,23
American Academic publishers, volume 05, issue 11,2025

Journal: https://www.academicpublishers.org/journals/index.php/ijai

A COMPREHENSIVE TAXONOMY AND PEDAGOGICAL FRAMEWORK FOR AI-DRIVEN TOOLS IN EFL LISTENING AND SPEAKING INSTRUCTION: INTEGRATING MULTIMODAL COMPOSING AND ADAPTIVE LEARNING

Nosirboyeva Shakhzoda Bektosh qizi

SamSIFL

Abstract: This paper proposes a comprehensive framework integrating Artificial Intelligence (AI) into English as a Foreign Language (EFL) instruction, specifically targeting listening and speaking skills. We position AI as a "semiotic mediator" that bridges receptive listening and productive speaking through a "Listening-Composing Cycle." This model guides learners from deconstructing audio texts with AI analysis tools to creating their own spoken and multimodal outputs with AI assistants. A taxonomy classifies AI tools by their pedagogical role. Supported by theories like Social Semiotics and Flow Theory, the framework offers principles for developing multimodal communicative competence, emphasizing critical AI literacy. While empirical studies show AI reduces speaking anxiety and boosts confidence, challenges like technological limitations and ethical concerns require a balanced, human-centric approach, advocating for a "Critical-Digital Humanism" in language education.

В данной статье предлагается комплексная система интеграции искусственного интеллекта (ИИ) в обучение английскому как иностранному, с особым акцентом на навыки аудирования и говорения. Мы позиционируем ИИ как «семиотического посредника», который связывает рецептивное аудирование и продуктивную устную речь через «Цикл Аудирования-Создания». Эта модель направляет учащихся от деконструкции аудиотекстов с помощью инструментов анализа ИИ к созданию их собственных устных и многомодальных работ с помощью ИИ-ассистентов. Таксономия классифицирует инструменты ИИ по их педагогической роли. Подкрепленная такими теориями, как социальная семиотика и теория потока, система предлагает принципы для развития многомодальной коммуникативной компетенции, подчеркивая критическую грамотность в области ИИ. В то время как эмпирические исследования показывают, что ИИ снижает тревожность при говорении и повышает уверенность, такие проблемы, как технологические ограничения и этические вопросы, требуют сбалансированного, ориентированного на человека подхода, выступая за «Критически-цифровой гуманизм» в языковом образовании.

Ushbu maqola Sun'iy Intellekt (SI) ni chet tili sifatida ingliz tilini o'qitishga, ayniqsa tinglab tushunish va gapirish ko'nikmalariga qaratilgan holda integratsiyalash uchun keng qamrovli asosni taklif etadi. Biz SI ni "Tinglash-Yaratish Sikli" orqali qabul qiluvchi tinglash va mahsuldor nutqni bog'lovchi "semiotik vositachi" sifatida joylashtiramiz. Ushbu model o'quvchilarni SI tahlil vositalari yordamida audio matnlarni dekonstruksiya qilishdan tortib, SI yordamchilari yordamida o'zlarining og'zaki va ko'p modalli ishlarini yaratishgacha olib boradi. Taksonomiya SI vositalarini ularning pedagogik roli bo'yicha tasniflaydi. Ijtimoiy semiotika va Oqim nazariyasi kabi nazariyalar tomonidan qo'llab-quvvatlangan asos, ko'p modalli

ORIGINAL ARTICLE

INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE

ISSN: 2692-5206, Impact Factor: 12,23
American Academic publishers, volume 05, issue 11,2025

Journal: https://www.academicpublishers.org/journals/index.php/ijai

kommunikativ kompetensiyani rivojlantirish uchun tamoyillarni taklif etadi, SI bo'yicha tanqidiy savodxonlikni ta'kidlab. SI gapirish vahimasini kamaytiradi va ishonchni oshiradi, degan empirik tadqiqotlar mavjud bo'lsa-da, texnologik cheklovlar va axloqiy muammolar kabi qiyinchiliklar muvozanatli, inson markazidagi yondashuvni talab qiladi va til ta'limida "Tanqidiy-Raqamli Gumanizm" ni targ'ib qiladi.

Keywords: Artificial Intelligence, EFL Listening, EFL Speaking, Digital Multimodal Composing, Pedagogical Framework, Semiotic Mediation, Multimodal Communicative Competence, AI-Assisted Language Learning, Speaking Anxiety, Learner Confidence, Learner Motivation, Adaptive Learning, Critical-Digital Humanism

Main Text

The pedagogical landscapes of teaching listening and speaking in English as a Foreign Language (EFL) have traditionally evolved along parallel but often separate tracks. Listening instruction has remained predominantly focused on decoding phonological and syntactic features, while speaking instruction frequently grapples with pronunciation accuracy, fluency, and overcoming anxiety. Simultaneously, the emerging field of Digital Multimodal Composing (DMC) emphasizes creative, productive language use through the orchestration of multiple semiotic modes such as text, image, and sound. This paper addresses the critical theoretical and practical gap between these domains by positioning Artificial Intelligence (AI) as the essential "semiotic mediator" that unifies receptive processing with productive design, particularly in the context of oral production.

Current AI applications in language learning often prioritize discrete skill development. In listening, tools like Automated Speech Recognition (ASR) systems focus narrowly on comprehension checking. In speaking, AI chatbots and pronunciation coaches provide practice but can underutilize the potential of critical listening as foundational input for multimodal creation and spontaneous speech. Our framework reconceptualizes this relationship, proposing that AI-driven tools can transform listening from passive reception into an active deconstruction of multimodal ensembles, which subsequently informs and enriches the construction of new spoken and multimodal texts. This is especially pertinent in contexts like the Defense Language Institute (DLI), where students, despite high proficiency in reading and listening, often lack the confidence and motivation to speak the target language, hindering their overall communicative competence and professional effectiveness. Research at DLI identified that linguistic, cognitive, psychological, and sociocultural factors contribute to this reluctance, with fear of making mistakes and criticism leading to a lack of confidence. The conventional methods in such contexts are often more teacher-centered than student-centered, failing to motivate students to convey their ideas or reflect the natural flow of speaking in everyday interactions.

This work synthesizes insights from a novel theoretical framework for listening and composing, empirical studies on AI's impact on speaking anxiety and motivation, and systematic reviews of AI integration in Asian tertiary EFL speaking instruction. It makes several significant contributions, namely a novel theoretical synthesis positioning AI as a semiotic mediator between listening and speaking-composing processes; an original multidimensional taxonomy classifying AI tools by their pedagogical roles in a listening-speaking continuum; and a

ISSN: 2692-5206, Impact Factor: 12,23
American Academic publishers, volume 05, issue 11,2025
Journal: https://www.academicpublishers.org/journals/index.php/ijai

comprehensive pedagogical framework for developing integrated multimodal communicative competence, supported by empirical evidence and addressing practical implementation challenges.

The proposed framework is underpinned by a confluence of robust theoretical perspectives that illuminate AI's multifaceted role as a mediator in language learning. Kress's social semiotic theory provides the foundational understanding that different modes such as aural, visual, and spatial offer specific affordances for meaning-making. The aural mode, central to listening, operates temporally and sequentially, while speaking and DMC involve orchestrating multiple modes. AI tools mediate this relationship by enabling "semiotic translation"—converting temporal speech into spatial representations through interactive transcripts, visual annotations, and discourse maps. This translation allows learners to analyze auditory information with the deliberation typically reserved for visual modes, creating crucial cognitive connections between reception, or listening, and production, or speaking and composing.

Building on Mayer's Cognitive Theory of Multimedia Learning, our framework recognizes AI's role in managing the complex cognitive demands of integrated listening-speaking tasks. AI tools function as cognitive scaffolds that strategically distribute processing across multiple channels. For instance, intelligent subtitling systems can reduce extraneous load during listening, while template-based composing platforms and AI conversation partners minimize cognitive overhead during production. This optimized load management enables deeper engagement with semantic content, rhetorical design, and fluency development. From a Vygotskyan perspective, AI tools serve as advanced psychological tools that mediate developing capabilities. In our framework, AI functions as a metacognitive scaffold that makes visible the often-implicit processes of both listening comprehension and multimodal design. When an AI system prompts a learner to identify rhetorical moves in a podcast or suggests alternative organizational structures for a digital story, it externalizes internal cognitive processes, facilitating their integration into the learner's developing skill set. This aligns with findings that AI can enhance self-efficacy and metacognition.

Furthermore, Flow Theory and Mindset Theory provide crucial psychological underpinnings for AI-enhanced environments. Flow theory emphasizes the state of optimal experience achieved when challenges balance skills. AI can personalize learning tasks to maintain this equilibrium, reducing the anxiety that disrupts flow and hinders speaking practice. Mindset Theory highlights how beliefs about ability impact learning. AI, by providing a safe, private space for practice and offering incremental, objective feedback, can help foster a growth mindset, convincing learners that their speaking skills can be improved through effort and strategy, thereby increasing motivation and willingness to communicate. This is particularly relevant for learners who hold a fixed mindset and believe their language skills cannot be substantially improved, making them more hesitant to speak due to fear of mistakes or judgment.

Our taxonomy innovates by categorizing tools not in isolation, but according to their mediating functions within an integrated listening-speaking-composing process. The category of Analytical Deconstruction Tools includes AI-powered transcript generators, multimedia annotation platforms, and discourse analysis tools. These tools mediate the "freezing" of

ISSN: 2692-5206, Impact Factor: 12,23
American Academic publishers, volume 05, issue 11,2025
Journal: https://www.academicpublishers.org/journals/index.php/ijai

temporal speech into analyzable spatial representations. For example, students might use timestamped annotations to identify persuasive techniques in a TED Talk, creating a visual map of rhetorical strategies that informs their own speech composition and delivery choices. Adaptive Input Curators, such as YouTube's recommendation algorithm and curated podcast apps, mediate access to level-appropriate, interest-aligned multimodal texts. These systems provide the essential raw material for both listening practice and compositional inspiration, creating thematic connections between listening input and speaking output.

Performance Feedback Systems, including ASR-based pronunciation coaches, intonation analyzers, and fluency trackers, mediate the development of productive oral skills. These tools create direct bridges between responsive processing and productive performance by providing immediate, objective feedback on students' oral productions, which can be integrated into their speaking practice and multimodal compositions. Research with DLI students using the Talk Pal AI tool demonstrated its effectiveness in reducing communication apprehension, with observational data showing increased willingness to engage in spontaneous language production. Studies in Asian tertiary contexts have shown that AI voice chatbots were more effective than text chatbots or face-to-face practice in improving college learners' fluency and confidence. Research in Vietnam found that regular oral practice sessions using AI voice chatbots improved college students' grammar, vocabulary, pronunciation, and fluency. Studies in Kazakhstan demonstrated AI chatbots' positive influence on speakers' vocabulary, pronunciation, intonation, and grammar accuracy in higher education settings.

Generative Composition Assistants, like AI writing tools and image generators, mediate the creative process by providing starting points, overcoming blocks, and suggesting alternative approaches for scripting speeches or creating multimodal narratives. In our framework, these tools are positioned not as replacements for creativity but as collaborative partners that expand compositional possibilities. Research on AI-supported collaborative argumentation showed progress in Indonesian EFL learners' argumentative speaking skills when using platforms like ChatGPT. Finally, Immersive Context Simulators, such as VR platforms and augmented reality tools, mediate learning by situating both listening and speaking within authentic, contextualized frames. These tools enable "embodied practice," where students both comprehend and produce language within simulated real-world scenarios, which has been shown to decrease apprehension about speaking in public.

Our central innovation is the Listening-Composing Cycle, a recursive process framework that systematically connects reception and production, directly applicable to speaking skill development. The cycle begins with Critical Deconstruction, where students engage in deep analysis of mentor texts like podcasts and interviews using analytical deconstruction tools. This phase moves beyond basic comprehension to examine how multimodal meanings are constructed. For example, learners might use AI transcription to map the narrative structure and emotional cadence of a documentary narrator. The next phase, Strategic Planning, sees students building on their deconstruction analysis to plan their own spoken responses or multimodal compositions. Using generative assistants and adaptive input curators, they brainstorm ideas, outline arguments, and gather relevant media assets. This phase emphasizes the transfer of insights from analysis to design.

ORIGINAL ARTICLE

INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE

ISSN: 2692-5206, Impact Factor: 12,23

American Academic publishers, volume 05, issue 11,2025

Journal: https://www.academicpublishers.org/journals/index.php/ijai

The cycle then moves to Dialogic Composition and Speaking Practice, where students create their spoken or multimodal texts through an iterative process of drafting and revision. This phase is supported by performance feedback systems for pronunciation and fluency and generative assistants for content. It emphasizes the recursive nature of composition and speaking, where ideas and delivery evolve through continuous interaction with AI tools and peer feedback. AI chatbots provide a low-anxiety environment for practicing dialogs and presentations. Research shows that task-based learning is the most frequently integrated instructional method with AI tools in Asian tertiary education contexts. The flipped classroom model has also proven effective when combined with AI chatbots, allowing students to practice speaking skills outside class and arrive prepared for more interactive in-class activities. Other effective methods include peer learning facilitated by AI-based mobile apps, contextualized learning using tools like Youdao Dictionary and Baidu Map, and gamification through platforms like Duolingo to enhance engagement and motivation.

The final phase, Metacognitive Curation, involves students compiling their work in digital portfolios accompanied by reflective commentaries that articulate their design choices, learning processes, and growth in speaking confidence. AI tools facilitate this reflection by providing analytics on pattern use, progression, and strategy development. Studies utilizing AI learning platforms like EAP Talk and Liulishuo have demonstrated their effectiveness in tracking student progress and providing data for self-assessment.

The pedagogical framework is guided by several core principles. Semiotic Awareness Design means instruction should explicitly highlight the affordances of different modes and their combinations. AI tools enable comparative analysis exercises where students experiment with conveying the same message through different modal combinations, such as a spoken explanation versus a narrated video, developing sensitivity to semiotic resource selection. The principle of Scaffolded Autonomy Development involves a progressive shift of agency from tool to learner through carefully sequenced scaffolding. Initial cycles might provide extensive AI support, such as full transcripts and detailed pronunciation feedback, with subsequent iterations gradually reducing automated assistance as students internalize processes and strategies, fostering independent speaking ability.

Critical AI Literacy Integration is essential, ensuring students develop a critical awareness of AI capabilities and limitations through guided evaluation of AI-generated content and reflection on human-AI collaboration dynamics. This includes analyzing biases in AI responses, understanding the ethical dimensions of data privacy, and recognizing when AI feedback may be inaccurate, thus preventing overreliance. Research shows that overreliance on AI is a significant challenge, as it can weaken learners' critical thinking and independent learning abilities, potentially hindering deeper cognitive engagement needed for mastering complex language skills. Finally, the principle of Trans-languaging and Multimodal Resource Integration actively encourages drawing on learners' full linguistic and semiotic repertoires. AI translation tools, multilingual transcription, and cross-cultural media resources facilitate this inclusive approach to meaning-making, validating learners' identities and supporting comprehension and production.

ORIGINAL ARTICLE

INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE

ISSN: 2692-5206, Impact Factor: 12,23

American Academic publishers, volume 05, issue 11,2025

Journal: https://www.academicpublishers.org/journals/index.php/ijai

Empirical research strongly supports this framework. Qualitative case studies, such as those involving DLI students using tools like Talk Pal AI, reveal that AI-assisted practice can reduce communication apprehension and boost confidence, especially among anxious learners. Systematic reviews of AI in Asian tertiary EFL contexts further validate the use of AI chatbots, apps, and platforms for improving speaking fluency, accuracy, and willingness to communicate. The research landscape shows dramatic growth, with publications increasing significantly since 2022, indicating rapidly growing scholarly interest and validation. Geographical analysis reveals concentrated research efforts in East Asia, particularly in Mainland China, Taiwan, and Korea. Methodologically, quantitative approaches utilizing pre-test and post-test designs to measure speaking improvement dominate the field, complemented by mixed-methods approaches that combine quantitative measures with qualitative insights from interviews and surveys.

However, successful implementation requires addressing several considerations. A key challenge is the misalignment between AI feedback and specific curricular goals or assessment criteria. AI systems must be integrated in a way that their feedback, particularly on speaking, is relevant to academic and real-world tasks. Research shows that learners often feel dissatisfied when AI feedback lacks relevance to their academic tasks or exams, and AI systems frequently provide superficial correction of pronunciation and grammar while lacking depth in textual structure and argumentation insights. Furthermore, educators need support in developing dual expertise in listening-speaking pedagogy and multimodal composition, along with technical proficiency in integrating AI tools meaningfully. Professional development is crucial to help teachers transition to facilitators in an AI-mediated classroom, with research indicating that many teachers lack adequate training opportunities to develop their abilities in using AI tools effectively.

Ethical and equity-focused implementation is paramount. Institutions must address data privacy concerns, algorithm bias awareness-particularly the documented problems with ASR accuracy when processing non-native accents-and equitable access to technology. Technical deficiencies present significant barriers, with studies highlighting problems with ASR systems accurately recognizing non-native accents, leading to confused feedback that can weaken student motivation. The high cost of some AI tools can exacerbate educational inequalities, particularly affecting students from low-income families and institutions with limited budgets. Other documented challenges include learner variability in cognitive abilities, cultural backgrounds, and learning preferences; emotional demotivation due to AI's lack of human empathy and social presence; and lack of AI literacy among both students and teachers, which can lead to blind reliance on AI's advice without critical evaluation.

Finally, this integrated framework presents more than a new method for teaching listening and speaking; it proposes a fundamental reorientation of EFL pedagogy for the digital age. By theorizing AI as a semiotic mediator and operationalizing it through the Listening-Composing Cycle, we provide a concrete pathway to transform learners from passive consumers of language into critical, agentive designers of multimodal meaning. The empirical evidence from diverse contexts—from defense language institutes to Asian universities—demonstrates consistent patterns of AI tools reducing speaking anxiety, boosting confidence, and improving specific speaking skills when properly integrated. Ultimately, this work advocates for an

ISSN: 2692-5206, Impact Factor: 12,23
American Academic publishers, volume 05, issue 11,2025
Journal: https://www.academicpublishers.org/journals/index.php/ijai

approach grounded in Critical-Digital Humanism. This philosophy resists the technocratic impulse to use AI merely for efficiency and skill-drill, reclaiming it instead for a humanistic pedagogy that prioritizes critical thinking, creative expression, and identity formation. The true objective shifts from mere communicative competence to rhetorical sovereignty—the learner's empowered right and ability to wield all available semiotic resources, including AI, to assert their voice, challenge dominant narratives, and shape their own representational destiny. The success of this framework will be measured not only in improved test scores but in the development of individuals who can harness the fusion of language, technology, and human creativity to critically engage with the world and articulate their place within it.

References:

OEH Elmahdi, AAS Balla, AH Abdelrady, E Osman, AOA Ahmed., 2025 " AI-Driven VocabularyAcquisition in EFLHigher Education: Interdisciplinary Insights into Technological Innovation, Ethical Challenges, and Equitable Access" Ethical Challenges, and Equitable Access,

Bjork, R.A., Bjork, E.L., 2019. Forgetting as the friend of learning: Implications for teaching and self- regu lated learning. Advances in Physiology Education. 43, 164–167.

Abulibdeh, A., Zaidan, E., & Abulibdeh, R. (2024). Navigating the confluence of artificial intelligence and education for sustainable development in the era of industry 4.0: Challenges, opportunities, and ethical dimensions. Journal of Cleaner Production, 140527. https://doi.org/10.1016/j.jclepro.2023.140527

Albreiki, B., Zaki, N., & Alashwal, H. (2021). A systematic literature review of student'performance prediction using machine learning techniques. Education Sciences, 11(9), 552. https://doi.org/10.3390/educsci11090552

Yu, M. Optimizing EFL vocabulary acquisition: a randomized controlled mixed-methods investigation of artificial intelligence-driven incidental, contextual, and multimodal strategies. Educ Inf Technol (2025). https://doi.org/10.1007/s10639-025-13803-2

Abdul Razak, R., Abdul Hamid, S., Ahmad, R., Samingan, A. K., & Mohd Nasir, N. A. (2021). Kurikulum Merdeka Belajar (KMB): Fostering Autonomous Learning Among Students in Higher Education Institutions. International Journal of Academic Research in Progressive Education and Development, 10(4), 256–270.
An, Y., & Mindrila, D. (2020). Strategies and tools used for learner-centered instruction. International Journal of Technology in Education and Science, 4(2), 133-143.
Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. Available at SSRN 4337484.
Bazeley, P. (2013). Qualitative data analysis: Practical strategies. Sage Publications.
Benson, P., & Voller, P. (2014). Autonomy and independence in language learning. Routledge.
Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative research in psychology, 3(2), 77–101.
Braun, V., & Clarke, V. (2020). Does one size fit all? What counts as quality practice in (reflexive) thematic analysis? Qualitative Research in Psychology, 18(3), 328–352.

ISSN: 2692-5206, Impact Factor: 12,23
American Academic publishers, volume 05, issue 11,2025
Journal: https://www.academicpublishers.org/journals/index.php/ijai

Bryman, A. (2012). Social research methods. Oxford University Press.
Budiarti, N. M., & Lusiana, I. G. A. S. (2019). Learner autonomy in speaking skill development: Indonesian EFL learners' perspective. Journal of Applied Linguistics and Language Research, 6(5), 180-194.
Cekaite, A., & Schubert, C. (2019). Developing learner autonomy in English language education: A case study of a Swedish high school. Studies in Second Language Learning and Teaching, 9(1), 33-56.
Chatib, M., Haryanto, E., Lailiyah, S., & Isnaini, H. (2021). Using Chatbot to Promote Students' Autonomy in English Learning during the Covid-19 Pandemic. International Journal of Emerging Technologies in Learning, 16(12), 107–116. https://doi.org/10.3991/ijet.v16i12.14849
Chen, C., & Wang, H. (2019). The effects of a mobile chatbot on student learning and engagement. Interactive Learning Environments, 27(8), 1036-1049.
Chik, A. (2010). Autonomy and CALL: A model of institutional implementation. In M. Thomas, H. Reinders, & B. Warschauer (Eds.), Contemporary computer-assisted language learning (pp. 98–118). Continuum.

CAFE E. F. L. Integrating AI into Listening Comprehension: Dynamic Approaches for EFL/ESL Learners.