

Arab World English Journal (AWEJ) Special Issue on CALL Vol. 17, No.12. July 2026
DOI: <https://dx.doi.org/10.24093/awej/call12.2>

Pp.23-39

Artificial Intelligence in English Language Teaching and Learning: A Scoping Review of Intelligent Computer-Assisted Language Learning (2015–2025)

Ese Emmanuel Uwosomah 

Department of Didactics of Language and Literature and Social Sciences,
Faculty of Education, Universitat Autònoma de Barcelona, Spain
Email: EseEmmanuel.Uwosomah@autonoma.cat

Received:02/27/2026

Accepted: 04/28/2026

Published:07/15/2026

Abstract

This scoping review primarily aims to synthesize empirical research on artificial intelligence in English language teaching and learning published from 2015 to 2025. The main question investigates how the integration, applications, and pedagogical roles of AI have evolved over the past decade. The significance of this study is that it uses Intelligent Computer-Assisted Language Learning as an interpretive lens to make sense of a rapidly shifting field, offering a framework to help educators navigate modern generative tools. Following Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews (PRISMA-ScR) guidance, 129 empirical studies were identified and analyzed using descriptive mapping and thematic analysis. The main findings indicate three overlapping evolutionary phases: an early system construction phase focused on tutoring, a mobile-and-voice phase emphasizing speech practice, and a generative phase dominated by large language models. The evidence base remains heavily concentrated in higher education, where AI frequently acts as a tutor, practice partner, or co-writer. For further use, this study recommends adopting teacher-mediated task designs, shifting assessments to focus on the learning process, and prioritizing longitudinal research in primary and under-resourced educational settings.

Keywords: Artificial intelligence, English language learning, English language teaching, Intelligent computer-assisted language learning, Scoping review

Cite as: Uwosomah, E. E. (2026). Artificial Intelligence in English Language Teaching and Learning: A Scoping Review of Intelligent Computer-Assisted Language Learning (2015–2025). *Arab World English Journal (AWEJ) Special Issue on CALL 17*. (12)23-39.
<https://dx.doi.org/10.24093/awej/call12.2>

Introduction

Artificial Intelligence (AI) is now a routine presence in many English language teaching and learning environments (Yassin et al., 2026). Across the last decade, AI research in language education has increasingly shifted toward empirical and implementation-focused questions, with growing attention to learner engagement, ethical use, and the need for teacher guidance and pedagogical regulation in real classroom settings (Li et al., 2025; Zainuddin et al., 2024). That said, the field has become difficult to read as a single story because studies draw on overlapping traditions, including Computer-Assisted Language Learning (CALL), Mobile-Assisted Language Learning (MALL), and broader artificial intelligence in language education, particularly in foreign language teaching and learning. Recent syntheses and field-level analyses note accelerated adoption of AI in English Language Teaching (ELT) and English Language Learning (ELL), especially following wider access to generative tools, while also highlighting persistent challenges related to academic integrity, unequal access, and teacher preparedness for responsible integration (Dwivedi et al., 2023; Kohnke et al., 2023; Daud et al., 2025).

These studies consistently show that the practical value of AI is linked to teacher decision-making about task design, constraints, and feedback norms, rather than tool capabilities alone. This perspective positions teacher governance and pedagogical orchestration as central conditions for meaningful learning, particularly in contexts where AI output must be verified, adapted, and integrated into learner voice and accountability (Memarian & Doleck, 2023; Daher, 2025). In this review, CALL is the umbrella term for technology-mediated language learning, MALL is a setting-specific subset organized around mobile or wearable devices, and AIED is the broader educational engineering tradition without a specific language or disciplinary focus. Within this landscape, Intelligent Computer-Assisted Language Learning (ICALL) is used as the conceptual anchor because it foregrounds how language-aware intelligence changes the pedagogical affordances of a system (Schulze & Heift, 2012; Ward, 2017; Elahi Shirvan & Taherian, 2025).

This study focuses on several core components. The main aim of this review is to map and synthesize a decade of empirical evidence on AI in English Language Teaching and Learning (ELT/L) published from 2015 to 2025. The significance of the study lies in its response to the fragmentation of the field; by offering a functional role framework, it supports theory building and classroom design, while explaining how generative language models both extend and disrupt dialogue-oriented ICALL. The research objectives are to chart a decade of empirical evidence, use ICALL as an interpretive lens to connect diverse tool families to recurring instructional mechanisms, and establish a single empirical map of the field. Based on PRISMA-ScR and Population-Concept-Context (PCC) guidance (Tricco et al., 2018; Munn et al., 2018), the review addresses the following research questions:

- **RQ1.** How has the integration of AI in English language teaching and learning evolved from 2015 to 2025?
- **RQ2.** How has AI been applied in ELT/L in the empirical studies included in this review?
- **RQ3.** At which educational levels and in which locations is research on AI in ELT/L most prevalent?
- **RQ4.** Which AI tools and technologies are most frequently reported in this empirical literature?

Figure 1 below presents a conceptual map used throughout the manuscript. The map connects three layers: AI tool families, ICALL application types, and reported functional roles in

ELT/L. In the first layer, early work centers on tutoring systems and automated scoring. In the second layer, mobile, voice, and social robotics bring interaction and affect to the foreground. In the third layer, large language model tools add open-ended dialogue and generative co-authoring, which expands dialogue-based ICALL toward what can be described as large language model-mediated ICALL. These tools extend dialogue-based traditions, but they also introduce new instructional problems around authorship, assessment validity, and the explicit teaching of AI literacy.

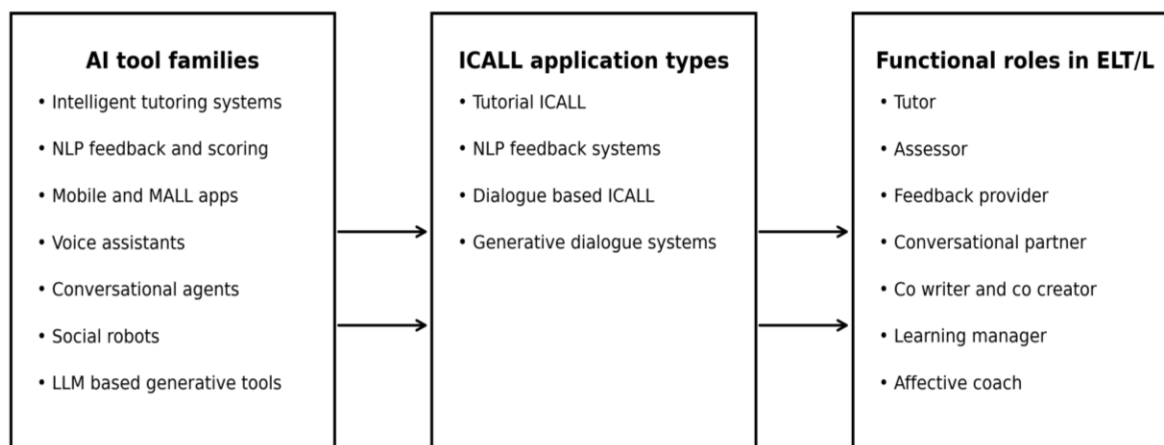


Figure 1. Conceptual map of AI tool families, ICALL categories, and functional roles in ELT/L. Source: Author's own conceptual synthesis based on ICALL literature and the reviewed studies.

Unlike prior syntheses, this review charts a full decade of empirical evidence while using ICALL as an interpretive lens, an approach that makes it possible to offer a functional role framework grounded in actual classroom mechanisms.

Literature Review

ELT and ELL are evolving rapidly as emerging educational technologies reshape classroom practice and learner experience, and AI now plays an increasingly important role in language education (Huang et al., 2023; Zhu & Wang, 2025). The expansion of open-access AI tools has accelerated teacher and learner experimentation, pushing the field to define how AI can add value rather than simply add novelty (Dwivedi et al., 2023; Kreijkes et al., 2026). At the same time, AI use is becoming common in academia, yet its integration in language teaching and learning is still relatively new and positioned as high potential but unevenly systematized (Sapuan et al., 2025). Post-pandemic pressure has further intensified AI adoption in teaching practice (Albedah, 2025).

In CALL, many of these developments can be interpreted through ICALL, where AI, computational linguistics, and NLP enable language aware systems that provide adaptive support beyond conventional digital practice (Elahi Shirvan & Taherian, 2025). Classic ICALL work distinguishes tutorial systems and dialogue-based systems, with dialogue-oriented approaches aiming to support interaction for meaning-focused practice (Schulze & Heift, 2012; Ward, 2017; Bibauw et al., 2022). Recent attention to Generative AI and LLMs extends this trajectory by foregrounding personalization, responsive scaffolding, and engagement through conversational support and content generation (Shuliar et al., 2023). Empirical discussions commonly reference tools such as ChatGPT and intelligent tutoring systems as pathways to more interactive and

individualized learning experiences (Allehyani & Algamdi, 2023). Another prominent strand concerns AI-supported assessment and feedback, often framed as a way to scale formative guidance when embedded within sound pedagogy (Alshumaimeri & Alshememry, 2023).

However, the literature also highlights persistent gaps that justify synthesis. The primary research gap addressed in this study is the lack of a cohesive, empirical synthesis that connects disparate AI tool families to underlying instructional mechanisms across different educational contexts. Effective implementation often depends on new teacher competencies and professional development that support principled task design and classroom orchestration (Kundu & Bej, 2025). Another area that remains comparatively unexamined is ethical concerns, including privacy, transparency, fairness, and the risk that algorithmic bias may reproduce inequities in language learning environments (Baker & Hawn, 2022). For this reason, recent syntheses argue for mapping AI in ELT and ELL across contexts, tools, and educational levels while keeping interpretation grounded in empirical reporting (Daud et al., 2025; Van Horn, 2024). This mapping work also consolidates the pedagogical roles commonly attributed to AI, including tutor, co-designer, assessor, affective facilitator, and support for teacher professional learning (Balraj, 2025; Sharadgah & Sa'di, 2022; Tan et al., 2025). What emerges across these discussions is less a consensus on which tools work than a shared insistence that tools alone are insufficient. Responsible integration, the literature repeatedly argues, requires human oversight, teacher-mediated task design, and frameworks that develop AI literacy, conditions that remain unevenly realized across the contexts studied (Sharma & George, 2026).

Methodology

This scoping review follows PRISMA-ScR reporting principles and the PCC framework to structure the search, selection, and charting process (Tricco et al., 2018; Pollock et al., 2023; Munn et al., 2018). The purpose is to map the empirical terrain and synthesize patterns of use, not to compute pooled effect sizes. PRISMA 2020 and related reporting guidance (Page et al., 2021) also informed reporting clarity and transparency in the review.

Information sources and search strategy

Sources included multidisciplinary research databases, publisher platforms, and specialized repositories commonly used in CALL and applied linguistics. Searching combined keyword queries with backward citation searching of key papers. The core date range was from 2015 to 2025. The data sources include but are not limited to the Autonomous University of Barcelona virtual library, Google Scholar, ERIC, MDPI, Frontiers in Psychology, ScienceDirect, Wiley, ResearchGate, Semantic Scholar, Springer Open, Excellencia, Hindawi, Sciendo, Springer Link, Taylor and Francis Online, and Brain, Broad Research in Artificial Intelligence and Neuroscience. Additional relevant studies were identified through backward citation searches of key papers. To support reproducibility, Table 4 (Appendix C) summarizes the main sources and an example of the search logic.

Eligibility criteria

Studies were included when they were published between 2015 and 2025, were peer reviewed or editorially reviewed, reported empirical data, and had a clear focus on English language teaching or learning in EFL, ESL, EAL, or EMI contexts. Studies were excluded when they were reviews, conceptual papers without empirical data, or did not focus on English language

education. The review focuses on publications written in English. This restriction was applied for feasibility and because the review aims to represent the segment of the literature most accessible to the international CALL readership. This choice may exclude relevant evidence in other languages and is therefore treated as a source of language bias in the limitations section. Although scoping reviews often include grey literature, this review limited inclusion to empirical studies published in editorially reviewed conference proceedings, which was used as a credibility threshold to confirm that proceedings papers had undergone formal academic selection and were part of recognized conference venues. This criterion was not intended to equate editorial review with peer review, but to ensure that included studies met a minimum standard of scholarly oversight and were not drawn from unverifiable sources. Because the purpose of the review was to map the applications and roles of AI in ELT/ELL rather than to evaluate intervention effectiveness, inclusion decisions prioritized empirical relevance and traceability over the hierarchy of evidence. Publication type and study design were recorded during data charting to preserve interpretive transparency.

Study Selection and Data Charting

Initial identification yielded 276 records broadly related to AI in education. Title and abstract screening retained 183 records that explicitly addressed AI in English language teaching or learning. Full text screening excluded 54 articles that did not meet the empirical focus, language focus, or accessibility requirements. The final corpus includes 129 empirical studies. To support interaction with full-text articles during screening and data charting, NotebookLM, an AI-powered research tool, was used as an auxiliary research assistant to query individual PDFs and identify key study attributes such as research design, participant characteristics, AI tool type, and reported outcomes. The tool facilitated targeted information retrieval from lengthy articles but did not replace manual reading. All eligibility decisions, inclusion judgments, and thematic coding were completed manually to preserve methodological rigor and minimize risks of automated misinterpretation.

Data Analysis and ICALL-informed Second-cycle Coding

The 129 studies were examined through descriptive mapping and thematic analysis (Braun & Clarke, 2006). Descriptive counts were produced for year, educational level, geographic location, target skill, and tool type. These counts establish the empirical landscape. Building on this map, thematic analysis identified recurring functional roles attributed to AI across studies. ICALL categories guided second-cycle coding in a targeted way. First cycle coding captured reported tool functions, pedagogical aims, interaction modes, and outcomes. In the second cycle coding, functional role labels were mapped to ICALL application types to clarify the instructional mechanisms implied by each role label. Four ICALL application types were used as organizing categories: tutorial ICALL, which emphasizes adaptive tutoring and structured practice; natural language processing feedback systems, which emphasize automated diagnosis and corrective feedback; dialogue-based ICALL, which emphasizes conversational interaction and meaning-focused practice; and generative dialogue systems, which emphasize open-ended dialogue and co-authoring with large language models. This mapping supports interpretive coherence when the same classroom role label appears across different tool families and time periods.

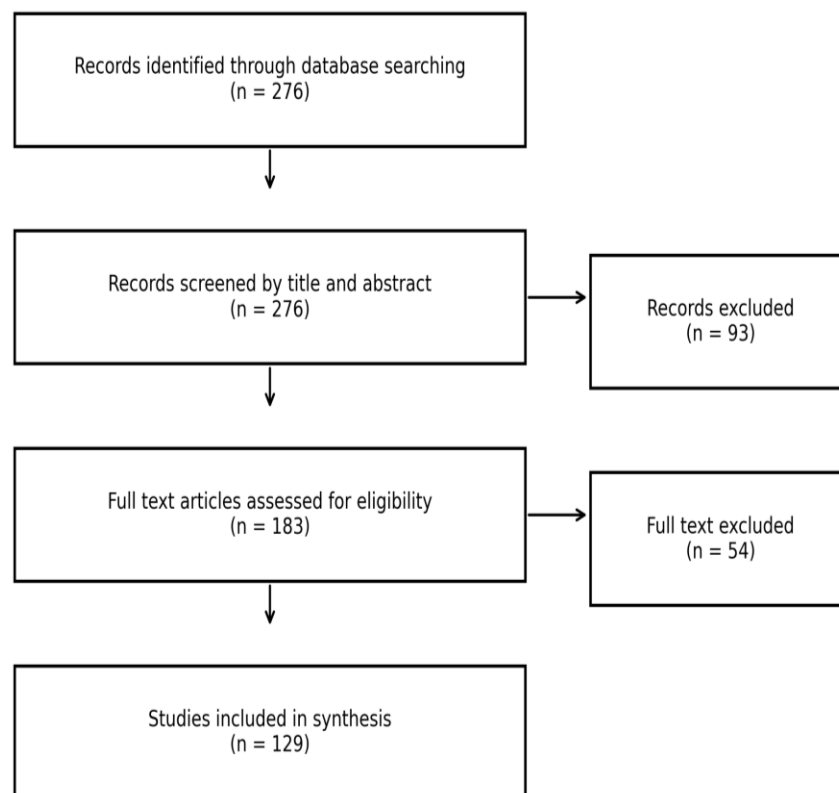


Figure 2. PRISMA-ScR flow diagram for study identification, screening, and inclusion. Source: Adapted from PRISMA-ScR guidelines (Tricco et al., 2018).

Trustworthiness procedures

To enhance trustworthiness, the review maintained a transparent audit trail including a study charting sheet, coding memos, and iterative refinement notes. After initial coding, a stability check was conducted by recording a purposive subset of studies and refining code definitions when discrepancies emerged. This approach supports analytic consistency in single-reviewer thematic synthesis and aligns with guidance that scoping reviews should prioritize transparent charting and clear synthesis logic (Pollock et al., 2023). AI tools were used only to support document handling and attribute extraction, while eligibility decisions and final coding were completed manually.

Results

Corpus Overview and Temporal Distribution

Research activity increases after 2019 and rises sharply after 2023, coinciding with wide access to large language model tools. Across the decade, the center of gravity moves from system building and automation to interaction, learner experience, and then to generative copilot uses and AI literacy.

Table 1. *Yearly distribution of empirical studies on AI in ELT/L (n = 129)*

Year	Number of studies	Phase label
2015	2	System construction and automation
2016	4	System construction and automation
2017	6	System construction and automation
2018	3	System construction and automation
2019	14	Transition to mobile, voice, and affect
2020	12	Mobile, voice, and affect
2021	9	Mobile, voice, and affect
2022	10	Mobile, voice, and affect
2023	25	Generative AI and copilot
2024	24	Generative AI and copilot
2025	20	Generative AI and copilot

As shown in Table 1, the steady early growth reflects the high technical barriers of initial system construction. The sharp increase in studies from 2023 onward highlights a paradigm shift, driven by the widespread accessibility of large language model tools and classroom practitioners evaluating open-access models in real-time educational environments. This evolution marks a clear shift in phase labels from early closed-system automation to recent generative AI and co-pilot roles.

Where the evidence is located

Studies are unevenly distributed across regions and are dominated by Asian higher education contexts. Table 2 illustrates the most represented locations and typical foci.

Table 2. *Top locations of empirical studies in the review*

Country or region	Approximate number of studies	Common foci
China	36	Tutoring systems, automated testing, college English platforms, ChatGPT in EFL and EMI.
Indonesia	14	Mobile speaking and pronunciation tools, chatbots, and generative AI classroom studies.

Saudi Arabia	8	AI-assisted writing and assessment, student and teacher perceptions, acceptance studies.
India	6	AI-supported listening and speaking, and academic writing with generative tools.
South Korea	6	Tutoring systems and chatbots for conversation practice.
Taiwan	6	Speech recognition and machine translation, robots and voice assistants, AI writing evaluators.

The geographic picture that emerges from Table 2 answers RQ3 and is quite remarkable: over half the corpus originates from two countries, with China accounting for 36 studies and Indonesia for 14. This is not simply a distributional footnote. It means that current understandings of how AI functions in ELT/L are being built largely from East and Southeast Asian higher education contexts, raising serious questions about whether the patterns identified here would hold in under-resourced, multilingual, or non-university settings elsewhere.

Educational levels

Higher education is the primary site for AI research in the corpus, with less sustained evidence from primary, secondary, and adult community programs. Table 3 summarizes the educational levels.

Table 3. *Educational levels represented in the corpus*

Educational level	Approximate share	Typical foci
Higher education	68 percent	Writing support, feedback and assessment, speaking practice tools, EMI support, teacher education
Secondary education	9 percent	Chatbots for grammar and inquiry, mobile pronunciation, AI AI-supported speaking
Primary and early childhood	4 percent	Social robots as tutors, conversational systems for young learners
General, vocational, and teacher-focused contexts	19 percent	Teacher readiness, vocational English, platform adoption studies

The breakdown of educational levels in Table 3 demonstrates that AI interventions have been predominantly tested in higher education settings, which account for 68% of the corpus. The sparse coverage of secondary (9%) and primary education (4%) suggests that the field currently lacks sufficient evidence on how younger learners navigate AI tools, pointing to a vital area for future longitudinal research.

Three-phase evolution of AI in ELT/L

The evolution of AI integration (RQ1) is characterized by three distinct, overlapping phases that summarize shifts in dominant tools and pedagogical functions across 2015 to 2025. The phases are interpretive, and they capture changes in what AI is asked to do in classrooms and research designs, rather than claiming discrete technological eras.

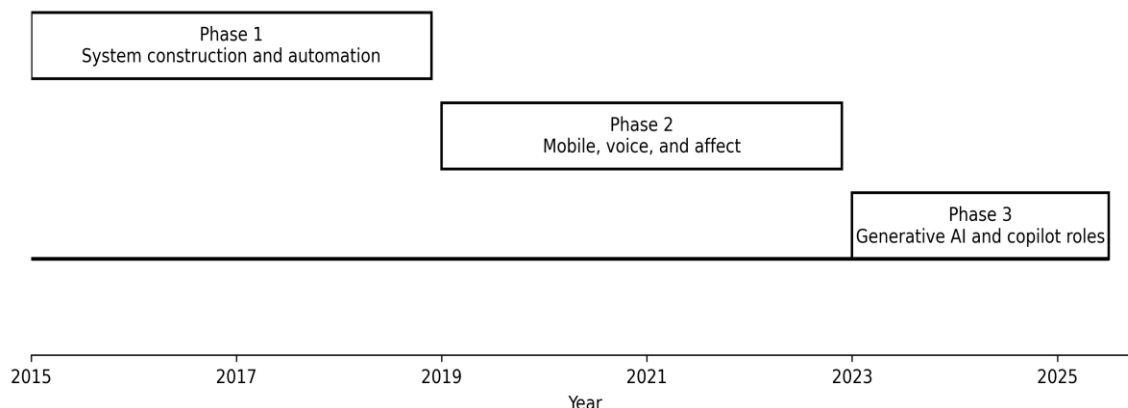


Figure 3. Timeline of three overlapping phases in empirical AI for ELT/L, 2015 to 2025.

Source: Created by the author using temporal patterns observed in the corpus.

Phase 1, system construction and automation, 2015 to 2018. Research aligns with tutorial ICALL and automated evaluation, with an emphasis on intelligent tutoring and rule-based feedback for grammar, vocabulary, and testing, plus early automated assessment systems (Sun, 2018). Reported outcomes focus on accuracy and efficiency, while classroom trials are typically short and small scale.

Phase 2, mobile, voice, and affect, 2019 to 2022. The corpus shifts toward interaction and learner experience through mobile applications, voice assistants, conversational agents, and social robots. Reported outcomes commonly include increased practice time, reduced anxiety, and higher motivation (Obari et al., 2020). This phase aligns strongly with dialogue-based ICALL because interactional practice and affective support become central.

Phase 3, generative AI and copilot roles, 2023 to 2025. Studies increasingly examine large language model tools, especially for writing support, revision, feedback, and AI literacy training. A consistent finding is that learning value depends on task constraints, transparency, and coaching on verification and voice (Duong & Chen, 2025; Zhang et al., 2025). This phase extends dialogue-based ICALL but adds a generative layer that reshapes authorship, assessment, and classroom accountability.

Functional Roles of AI and ICALL Alignment

In response to RQ2, the review brings together the recurring functional roles attributed to AI in ELT/L and presents them in Table 5 in Appendix D. The table defines each role and links it to a corresponding ICALL application type. Across the corpus, the roles identified were tutor, assessor, practice partner, affective support, co-writer, and teacher aid. These categories were used during second cycle coding to connect reported tool functions with underlying ICALL mechanisms. This made it possible to show how different technologies could serve similar instructional purposes across studies. The framework also has practical value, as it can

be used as a design checklist to help researchers and teachers specify the instructional mechanism involved, the mode of interaction, and the kind of evidence needed for each role.

Tools and Technologies most Frequently Reported

To respond to RQ4, the AI tools reported across the corpus were organized into broad categories and are presented in Table 6 in Appendix E. Grouping the tools in this way helps reduce repetition across individual studies while making longer term patterns in the literature easier to see. It also provides a clearer view of how the technological focus of the field has shifted over time. Throughout the decade, early studies were dominated by intelligent tutoring and automated feedback systems. In the middle period, mobile and voice-based tools became more visible. After 2023, generative large language model systems emerged as the most prominent category. This categorization, therefore, not only summarizes the tools discussed in the corpus but also helps show the broader movement in AI-supported ELT/L across the period under review.

Discussion

The findings of this review show a clear progression in the integration of AI into English language teaching over the past decade. The evidence moves from early tutoring and automated scoring systems to mobile and voice-based tools that support interaction and affective engagement, then to generative AI for drafting, revising, and AI literacy development. Across these stages, AI is repeatedly positioned as a tutor, assessor, practice partner, affective support, co-writer, and teacher aid, with these roles aligning closely with established ICALL mechanisms. Most studies are concentrated in higher education and in a limited number of regions, especially China and Indonesia, while primary and secondary contexts remain underrepresented. The tools discussed follow the same path, moving from intelligent tutoring and automated feedback to mobile and conversational agents, then to large language model tools.

The review also reflects the broader development outlined in earlier scholarship. Earlier CALL and ICALL work focused on rule-based tutoring and NLP-driven feedback (Schulze & Heift, 2012; Ward, 2017), which matches the first phase of the corpus. The rise of mobile, voice, and conversational tools reflects the growing attention given to interaction, motivation, and affect in the late 2010s (Obari et al., 2020; Bibauw et al., 2022). Recent attention to generative AI matches analyses showing how LLMs are reshaping writing support, feedback, and ethical debate in ELT/L (Dwivedi et al., 2023; Kohnke et al., 2023; Li et al., 2025). The geographic concentration found here echoes concerns in recent reviews about the limited evidence from under-resourced and non-university settings (Daud et al., 2025; Sapuan et al., 2025). The decade-long map developed in this review confirms broader trends and points to the empirical gaps that remain.

Interpreting change through ICALL

When ICALL is applied as an interpretive lens rather than a classification scheme, a more nuanced picture emerges. Many tools that appear new are, instructionally speaking, familiar: they deliver tutorial sequences and corrective feedback through updated interfaces while leaving the underlying pedagogical logic largely intact. What genuinely changes the picture is the arrival of large language models, which do not simply automate existing mechanisms but introduce a generative capability that redefines what interaction means in a language learning context. The key classroom issue becomes how learners verify, adapt, and take ownership of AI output, not only

whether the system can detect and correct errors. The functional role framework helps prevent category confusion. Tool families describe what technology is. On the other hand, ICALL application types describe the instructional mechanism, and functional roles describe what the tool is asked to do in a learning design. Keeping these layers distinct makes it easier to compare studies, explain mixed findings, and specify what kind of evidence is needed for each claimed benefit.

Research Implications

Pedagogical Implications

The findings indicate that the pedagogical value of AI depends heavily on teacher decisions regarding task constraints and feedback norms, rather than just the capabilities of the tools themselves. Generative AI shifts the teacher's role from a mere tool adopter to a "learning architect," requiring them to model verification practices, define acceptable use, and help learners maintain their voice and accountability. This means that assessment designs must pivot from product-focused evaluations to process-oriented ones, emphasizing reasoning through drafts with commentary, oral defenses, and reflective logs. What is also notable is that pedagogical frameworks must explicitly integrate AI literacy and ethical governance, addressing risks like hallucinated content, algorithmic bias, and privacy, so that learners are equipped to critically evaluate output quality and make responsible decisions (Baker & Hawn, 2022; Dakakni & Safa, 2023; Zhang et al., 2025). This ethical layer reinforces the need for explicit governance rather than uncritical adoption.

Contextual Implications

The review reveals that the current evidence base is highly concentrated in higher education and a small set of geographic regions, which significantly limits the generalizability of the findings. There is a pressing need to expand AI integration and research into underrepresented contexts, specifically primary and secondary education, as well as adult community programs. Crucially, when implementing AI in under-resourced contexts, educators and researchers must pay careful attention to local infrastructure constraints and the resulting impacts on teacher workload.

Methodological Implications

Future research needs to move beyond short-term interventions and small-scale pilot studies by evaluating AI-supported designs longitudinally over longer periods and across entire courses. Methodologically, this requires replication studies, delayed post-testing, and making the complete instructional sequence explicitly visible rather than treating AI use as a single, isolated treatment. Furthermore, study designs must be highly transparent; researchers should explicitly report the specific functional role being tested, the underlying ICALL mechanisms, and the human oversight procedures utilized for data integrity and pedagogy. For studies involving large language models, transparent reporting of prompts, task constraints, and validation steps is critical, as minor design alterations can significantly affect both learner behavior and outcomes.

This review has limitations. The search covers studies from 2015 to 2025, but data collection concluded during 2025 and therefore does not capture all publications from that year. The English-language restriction may exclude relevant evidence published in other languages and amplify geographic bias in the available corpus. Although multiple databases and platforms were searched, the review does not claim complete coverage of local or regional journals. Finally, AI

tools supported some logistical aspects of document handling, but all eligibility and coding decisions were completed manually.

Conclusion

This scoping review maps 129 empirical studies on AI in English language teaching and learning published from 2015 to 2025. The synthesis identifies three overlapping phases, moving from system construction and automation to mobile and voice-centered interaction and affect, and then to generative co-authoring and copilot uses. Across phases, AI is most often positioned as a tutor, assessor, practice partner, co-writer, affective facilitator, and teacher support. Interpreting these roles through ICALL clarifies the instructional mechanisms behind role labels and shows that large language model tools both extend and disrupt dialogue-based ICALL traditions. Evidence remains concentrated in higher education and a small set of regions, reinforcing the need for longitudinal classroom research in underrepresented contexts. What the decade ultimately shows is that the question was never really about AI, it was always about the conditions under which learning happens. Teacher-mediated task design, reformed assessment practices, and explicit AI literacy development are not supplementary recommendations; they are the preconditions without which the tools, however capable, are unlikely to produce sustained or equitable learning gains.

Acknowledgement: This work was carried out within the framework of the Doctorate in Education Program at the Autonomous University of Barcelona, Spain.

Funding: This research is not funded.

Conflicts of interest: The author declares no conflict of interest.

Authenticity Statement: This manuscript is an original work.

Artificial Intelligence Statement: Copilot AI was used for editing and phrase refining to improve readability. Grammarly was used for proofreading and sentence rewriting for tone adjustment. Google NotebookLM, an AI-powered research tool, was used to support the thematic organization of data. All study selection, coding, and interpretation were completed manually by the author, who remains fully responsible for the content of the publication.

About the Author:

Ese Emmanuel Uwosomah is a doctoral candidate at the Autonomous University of Barcelona, Spain. He studies AI applications in edtech and language instruction.

ORCID ID <https://orcid.org/0009-0007-0482-5391>

References

- Albedah, F. (2025). Artificial intelligence in language education: A systematic review of multilingual applications, large language models, and emerging challenges. *Language Teaching Research Quarterly*, 49, 247–268. <https://doi.org/10.32038/ltrq.2025.49.13>
- Allehyani, S. H., & Algamdi, M. A. (2023). Digital competences: Early childhood teachers' beliefs and perceptions of ChatGPT application in teaching English as a Second Language (ESL). *International Journal of Learning, Teaching and Educational Research*, 22(11), 343–363. <https://doi.org/10.26803/ijlter.22.11.18>

- Alshumaimeri, Y. A., & Alshememry, A. K. (2023). The extent of AI applications in EFL learning and teaching. *IEEE Transactions on Learning Technologies*, 17, 653–663. <https://doi.org/10.1109/TLT.2023.3322128>
- Baker, R. S., & Hawn, A. (2022). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education*, 32(4), 1052–1092. <https://doi.org/10.1007/s40593-021-00285-9>
- Balraj, B. M. (2025). Exploring the use of ChatGPT in academic writing: A systematic literature review on undergraduates' perceptions. *Arab World English Journal, Special Issue on Artificial Intelligence*, 348–357. <https://doi.org/10.24093/awej/AI.20>
- Bibauw, S., François, T., & Desmet, P. (2022). Dialogue systems for language learning: Chatbots and beyond. In N. Ziegler & M. González-Lloret (Eds.), *The Routledge handbook of second language acquisition and technology* (pp. 121–134). Routledge. <https://doi.org/10.4324/9781351117586-12>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Daher, R. (2025). Integrating AI literacy into teacher education: A critical perspective paper. *Discover Artificial Intelligence*, 5, Article 217. <https://doi.org/10.1007/s44163-025-00475-7>
- Dakakni, D., & Safa, N. (2023). Artificial intelligence in the L2 classroom: Implications and challenges on ethics and equity in higher education: A twenty-first century Pandora's box. *Computers and Education: Artificial Intelligence*, 5, Article 100179. <https://doi.org/10.1016/j.caeai.2023.100179>
- Daud, A., Aulia, A. F., Muryanti, Harfal, Z., Nabilla, O., & Ali, H. S. (2025). Integrating artificial intelligence into English language teaching: A systematic review. *European Journal of Educational Research*, 14(2), 677–691. <https://doi.org/10.12973/eu-jer.14.2.677>
- Duong, T.-N.-A., & Chen, H.-L. (2025). An AI chatbot for EFL writing: Students' usage tendencies, writing performance, and perceptions. *Journal of Educational Computing Research*, 63(2), 406–430. <https://doi.org/10.1177/07356331241312363>
- Dwivedi, Y. K. et al. (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, Article 102642. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>
- Elahi Shirvan, M., & Taherian, T. (2025). ICALL and individual differences. In L. McCallum & D. Tafazoli (Eds.), *The Palgrave encyclopedia of computer-assisted language learning*. Palgrave Macmillan. https://doi.org/10.1007/978-3-031-51447-0_289-1
- Huang, X. et al. (2023). Trends, research issues and applications of artificial intelligence in language education. *Educational Technology & Society*, 26(1), 112–131. [https://doi.org/10.30191/ETS.202301_26\(1\).0009](https://doi.org/10.30191/ETS.202301_26(1).0009)
- Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). Exploring generative artificial intelligence preparedness among university language instructors: A case study. *Computers and Education: Artificial Intelligence*, 4, Article 100156. <https://doi.org/10.1016/j.caeai.2023.100156>
- Kreijkes, P. et al. (2026). Effects of LLM use and note-taking on reading comprehension and memory: A randomised experiment in secondary schools. *Computers & Education*, 243, Article 105514. <https://doi.org/10.1016/j.compedu.2025.105514>

- Kundu, A., & Bej, T. (2025). Transforming EFL teaching with AI: A systematic review of empirical studies. *International Journal of Artificial Intelligence in Education*, 35, 2281–2314. <https://doi.org/10.1007/s40593-025-00470-0>
- Li, B. et al. (2025). Two years of innovation: A systematic review of empirical generative AI research in language learning and teaching. *Computers and Education: Artificial Intelligence*, 9, 100445. <https://doi.org/10.1016/j.caeai.2025.100445>
- Memarian, B., & Doleck, T. (2023). Fairness, accountability, transparency, and ethics (FATE) in artificial intelligence (AI) and higher education: A systematic review. *Computers and Education: Artificial Intelligence*, 4, Article 100152. <https://doi.org/10.1016/j.caeai.2023.100152>
- Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*, 18(1), Article 143. <https://doi.org/10.1186/s12874-018-0611-x>
- Obari, H., Lambacher, S., & Kikuchi, H. (2020). The impact of using AI and VR with blended learning on English as a foreign language teaching. In K.-M. Frederiksen, S. Larsen, L. Bradley, & S. Thouësny (Eds.), *CALL for widening participation: Short papers from EUROCALL 2020* (pp. 253–258). Research-publishing.net. <https://doi.org/10.14705/rpnet.2020.48.1197>
- Page, M. J. et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Pollock, D. et al. (2023). Recommendations for the extraction, analysis, and presentation of results in scoping reviews. *JBIM Evidence Synthesis*, 21(3), 520–532. <https://doi.org/10.11124/JBIES-22-00123>
- Sapuan, W. N. M., Sulaiman, N. A., & Mohamad, M. (2025). The trends in the integration of AI technologies in English language teaching: Systematic literature review. *Arab World English Journal*, 16(3), 259–269. <https://doi.org/10.24093/awej/vol16no3.15>
- Schulze, M., & Heift, T. (2012). Intelligent CALL. In M. Thomas, H. Reinders, & M. Warschauer (Eds.), *Contemporary computer-assisted language learning* (pp. 67–81). Bloomsbury Academic. https://www.researchgate.net/publication/275642465_Intelligent_CALL
- Sharadgah, T. A., & Sa'di, R. A. (2022). A systematic review of research on the use of artificial intelligence in English language teaching and learning (2015–2021): What are the current effects? *Journal of Information Technology Education: Research*, 21, 337–377. <https://doi.org/10.28945/4999>
- Sharma, V. K., & George, R. (2026). Teachers' perceptions and experiences of ChatGPT integration in language learning and academia: A mixed-method study. *Arab World English Journal*, 17(1), 103–115. <https://doi.org/10.24093/awej/vol17no1.7>
- Shuliar, V., Shkurko, V., Polukhtovych, T., Semeniako, Y., Shanaieva-Tsymbal, L., & Koltok, L. (2023). Using artificial intelligence in education. *BRAIN: Broad Research in Artificial Intelligence and Neuroscience*, 14(3), 516–529. <https://doi.org/10.18662/brain/14.3/488>
- Sun, Y. (2018). A research on classroom teaching ability system construction of English teachers combined with artificial intelligence. In J. Mizera-Pietraszko & P. Pichappan (Eds.), *Lecture notes in real-time intelligent systems: Advances in Intelligent Systems and Computing* (Vol. 613, pp. 1–10). Springer. https://doi.org/10.1007/978-3-319-60744-3_1

- Tan, X., Cheng, G., & Ling, M. H. (2025). Artificial intelligence in teaching and teacher professional development: A systematic review. *Computers and Education: Artificial Intelligence*, 8, Article 100355. <https://doi.org/10.1016/j.caeai.2024.100355>
- Tricco, A. C. et al. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, 169(7), 467–473. <https://doi.org/10.7326/M18-0850>
- Van Horn, K. R. (2024). ChatGPT in English language learning: Exploring perceptions and promoting autonomy in a university EFL context. *TESL-EJ*, 28(1). <https://doi.org/10.55593/ej.28109a8>
- Ward, M. (2017). ICALL's relevance to CALL. In K. Borthwick, L. Bradley, & S. Thouësny (Eds.), *CALL in a climate of change: Adapting to turbulent global conditions* (pp. 328–332). Research-publishing.net. <https://doi.org/10.14705/rpnet.2017.eurocall2017.735>
- Yassin, B., Almousa, N., & Almousa, A. A. (2026). Artificial intelligence in EFL higher education: Effects on academic performance and social competence. *Frontiers in Education*, 11, Article 1736757. <https://doi.org/10.3389/feduc.2026.1736757>
- Zainuddin, N. et al. (2024). Responsible and ethical use of artificial intelligence in language education: A systematic review. *Forum for Linguistic Studies*, 6(5), 316–325. <https://doi.org/10.30564/fls.v6i5.7092>
- Zhang, Q., Nie, H., Fan, J., & Liu, H. (2025). Exploring the dynamics of artificial intelligence literacy on English as a foreign language learners' willingness to communicate: The critical mediating roles of artificial intelligence learning self-efficacy and classroom anxiety. *Behavioral Sciences*, 15(4), Article 523. <https://doi.org/10.3390/bs15040523>
- Zhu, M. & Wang, C. (2025). A systematic review of AI in language education: Current status and future implications. *Language Learning & Technology*, 29(1), 1–29. <https://doi.org/10.64152/10125/73606>

Appendices

Appendix A

The complete list of the 129 publications included in this review is available as a separate open-access dataset. The file can be accessed through Zenodo using the following citation: Research File. (2026). Data set. Zenodo. <https://doi.org/10.5281/zenodo.18733807>

Appendix B

This appendix provides a visual excerpt of the ICALL charting and coding sheet used to systematically extract data from the included studies

111									
	A	B	C	D	E	F	G	H	I
1	Study / Citation	Geographic Location & Ed. Lev.	Research Design & Participants	Target Skill	AI Tool Type (and Name)	1st Cycle: Aims Modes & Functions	1st Cycle: Reported Outcomes	2nd Cycle: Functional Role	2nd Cycle: ICALL Type
2	Alhabbash et al. (2016)	[Extract from PDF]	[Extract from PDF]	Grammar (English Tenses)	Intelligent tutoring system	Aims: Grammar instruction. Mode: Rule-based feedback. Function: Automated evaluation.	Accuracy and efficiency.	Tutor	Tutorial ICALL
3	Dizon & Tang (2019)	[Extract from PDF]	Pilot study; [Participant N]	Speaking / Autonomous learning	Voice assistant (Alexa)	Aims: Conversational practice. Mode: Voice interaction. Function: Providing speaking opportunities.	Increased practice time, autonomous learning.	Practice partner	Dialogue-based ICALL
4	Duong & Chen (2025)	[Extract from PDF]	[Extract from PDF]	EFL Writing	AI Chatbot	Aims: Writing support. Mode: Open-ended prompting. Function: Drafting and revision assistance.	Changes in writing performance and student perceptions.	Co-writer and revision partner	Generative dialogue systems
5									
6									
7									

Appendix C

Table 4. *Information sources and example search logic. Final search date: 31 August 2025*

Source	Coverage	Example search logic
ERIC, Google Scholar, Semantic Scholar	2015 to 2025	(artificial intelligence OR AI OR chatbot OR intelligent tutoring OR speech recognition OR social robot OR large language model OR ChatGPT OR generative) AND (English OR EFL OR ESL OR ELT OR ELL)
Publisher platforms (ScienceDirect, SpringerLink, Wiley, Taylor and Francis Online)	2015 to 2025	(AI OR artificial intelligence OR intelligent) AND (English language teaching OR English language learning)
Open access collections (MDPI, Frontiers)	2015 to 2025	(ChatGPT OR generative AI OR large language model) AND (EFL OR ESL OR academic writing OR speaking)

Appendix D

Table 5. *Functional roles of AI in ELT/L and their ICALL alignment*

Functional role label	Role definition in ELT/L	Typical tools in the corpus	ICALL application type
Tutor	Provides structured practice and adaptive sequences for grammar, vocabulary, or reading	Intelligent tutoring systems, adaptive platforms	Tutorial ICALL
Assessor and feedback system	Generate scores, diagnostic feedback, or automated error detection	Automated writing evaluation, speech scoring, NLP feedback tools	NLP feedback systems

Practice partner	Offers conversational or speaking practice with repeated opportunities for output	Chatbots, voice assistants, social robots	Dialogue-based ICALL
Affective facilitator	Targets motivation, anxiety reduction, or willingness to communicate through supportive interaction	Conversational agents, mobile speaking tools	Dialogue-based ICALL
Co-writer and revision partner	Supports drafting, idea generation, paraphrasing, and revision under task constraints	ChatGPT and similar large language model tools	Generative dialogue systems
Co-designer and teacher support	Assist teachers with materials, task design, or professional learning activities	Teacher facing generative tools, planning assistants	Generative dialogue systems and teacher mediation

Appendix E

Table 6. *Tool and technology categories in the corpus*

Tool category	Common functions	Most visible phase
Intelligent tutoring and adaptive platforms	Grammar and vocabulary practice, sequencing, analytics	2015 to 2018, continuing into 2022
Automated assessment and scoring models	Writing and speaking evaluation, prediction of test outcomes	2015 to 2022
Speech recognition and pronunciation applications	Pronunciation feedback, fluency practice, speaking confidence	2019 to 2022
Conversational agents and social robots	Dialogue practice, engagement, affect support	2019 to 2022
Large language model tools	Drafting, revision, feedback, AI literacy training	2023 to 2025
Hybrid ecologies combining AI with VR or gamification	Immersive speaking tasks, engagement, resource integration	2019 to 2022, sporadic in 2023 to 2025