



Artificial Intelligence and Language Learning: A Systematic Literature Review of Current Trends and Future Directions

Fince Leny Sambeka^{1*}, Justien R. R. Wuisang², Jourike Jeane Runtuwarouw³, Donal Matheos Ratu⁴, Rinny Suryani Rorimpandey⁵, Tirza Annette Kumayas⁶, Fergina Lengkoan⁷

^{1,2,3} Department of Japanese Language Education, Universitas Negeri Manado, Indonesia

⁴ Department of Indonesia Language and Literature Education, Universitas Negeri Manado, Indonesia

^{5,6,7} Department of English Education, Universitas Negeri Manado, Indonesia

*Email: fincesambeka@unima.ac.id

ARTICLE INFO

Keywords:

Artificial intelligence
language learning
systematic literature review

ABSTRACT

Purpose- This study aims to synthesize current empirical evidence on the use of AI in language education by identifying commonly used AI technologies, examining reported learning and teaching benefits, and analyzing persistent challenges and future research directions.

Methodology- The study employed a Systematic Literature Review (SLR) with qualitative thematic synthesis of 50 empirical studies published between 2023 and 2024. The studies were identified using AI- and language-learning-related search terms and examined through a structured data extraction process that covered AI tools, educational contexts, language domains, reported benefits, challenges, and theoretical perspectives. The extracted data were coded and analyzed thematically to identify recurring patterns, which were interpreted through the perspectives of Computer-Assisted Language Learning (CALL), Technology Acceptance Model (TAM), and Sociocultural Theory (SCT).

Findings- The findings showed that ChatGPT and generative AI systems dominate current AI-assisted language learning research. The most frequently reported benefit was feedback, scaffolding, and personalization (44%), followed by language performance and skill development (42%) and teacher support and pedagogical productivity (36%). However, major challenges included accuracy and reliability issues, such as hallucinations (38%), AI literacy and training needs (28%), and risks of over-reliance and reduced critical thinking (26%).

Contribution- This study presents a comprehensive synthesis of recent empirical evidence and proposes a human-AI collaborative learning perspective.

Received 12 August 2026; Received in revised form 12 September 2026; Accepted 06 October 2026

Jurnal Eduscience (JES) Volume 13 No. 5 (2026)

Available online 30 October 2026

©2025 The Author(s). Published by LPPM Universitas Labuhanbatu. This is an open-access article under the [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License \(CC BY - NC - SA 4.0\)](#)

INTRODUCTION

Artificial intelligence has moved from a peripheral educational technology to a highly visible component of language learning and teaching. The recent diffusion of generative AI and large language models has intensified this shift because learners and teachers can now interact with systems that generate explanations, examples, feedback, exercises, assessments, and extended discourse in real time. In second- and foreign-language education, with this review focusing primarily on English as a Foreign/Second Language (EFL/ESL) contexts, which dominate the literature, these affordances intersect directly with core pedagogical activities such as language practice, writing development, vocabulary and grammar learning, speaking preparation, materials development, and formative assessment (Chapelle, 2001).

The emerging empirical literature illustrates both the breadth and the unevenness of this transformation. ChatGPT has been used to support academic writing and motivation (Song & Song, 2023), vocabulary development (Mugableh, 2024), grammar learning (Kucuk, 2024), feedback on EFL writing (Guo & Wang, 2024; Mahapatra, 2024), informal digital learning (Liu & Ma, 2024; Liu et al., 2024), and automated or AI-assisted evaluation (Mizumoto & Eguchi, 2023; Mizumoto et al., 2024). Teachers have also explored AI for instructional preparation, feedback, and material generation (Gao et al., 2024; Mabuan, 2024; Mohamed, 2024). These studies indicate that AI is becoming embedded across both learner-facing and teacher-facing language-education practices.

At the same time, rapid adoption has generated substantial pedagogical and ethical uncertainty. AI output may be fluent but inaccurate, incomplete, culturally biased, or pragmatically inappropriate (Algaraady & Mahyoob, 2023; Mohamed, 2024; Zheng & Stewart, 2024). Generative AI also complicates conventional assumptions about authorship, plagiarism detection, and the validity of take-home writing assessment (Alexander et al., 2023; Ibrahim, 2023). Further concerns include learner over-reliance, reduced critical engagement, privacy, teacher preparedness, and the need for explicit AI literacy (Cong-Lem et al., 2024; Derakhshan & Ghiasvand, 2024; Yuan et al., 2024). Consequently, the educational question is no longer simply whether AI can be used in language learning, but how, when, and under what conditions its use produces defensible learning outcomes.

Three theoretical perspectives are particularly useful for organizing this question. Computer-Assisted Language Learning (CALL) provides a framework for understanding technology-mediated input, interaction, feedback, practice, and language production (Chapelle, 2001). The Technology Acceptance Model (TAM) explains how perceived usefulness and perceived ease of use shape attitudes and behavioral intentions toward technology adoption (Davis, 1989). Sociocultural Theory (SCT) conceptualizes learning as mediated by tools and interaction, thereby providing a lens for interpreting AI as a potential scaffold or mediational artifact (Vygotsky, 1978; Lantolf & Thorne, 2006). Used together, these perspectives distinguish technological affordance, user adoption, and developmental mediation.

Despite the rapid growth of research, the literature remains dispersed across different language skills, educational levels, study designs, and AI functions. A synthesis is therefore needed that not only catalogs AI applications but also compares reported benefits with recurring limitations and identifies where evidence remains weak. This study addresses that need through a systematic literature review and qualitative thematic synthesis of 50 empirical studies published in 2023–2024. It is guided by three research questions: (1) What AI technologies are commonly used in language learning studies? (2) What benefits of AI-assisted language learning are reported? Moreover, (3) What challenges and research gaps remain?

METHODOLOGY

Research Design

This study employed a Systematic Literature Review (SLR) with qualitative thematic synthesis. The design was selected to identify, organize, compare, and interpret empirical evidence concerning AI-assisted language learning without statistically pooling heterogeneous outcomes. The review focused on technologies and functions (RQ1), reported learning and teaching benefits (RQ2), and challenges, limitations, and research

gaps (RQ3). The review logic comprised identification, screening, eligibility assessment, data extraction, quality consideration, coding, and qualitative synthesis.

The final synthesis comprised 50 empirical studies published in 2023 and 2024. The corpus contained 18 studies from 2023 and 32 from 2024. Coding was non-exclusive: one study could contribute evidence to multiple benefit, challenge, or theoretical categories. This approach was necessary because a single AI-supported activity can simultaneously involve feedback, skill development, motivation, technology acceptance, assessment, and ethical concerns.

Search Strategy and Eligibility Criteria

The review protocol used combinations of AI-related and language-learning terms. The core search logic combined ("artificial intelligence" OR "generative AI" OR chatbot* OR "AI tutor*" OR "automated feedback" OR "intelligent tutoring system*" OR "large language model*" OR ChatGPT) with ("language learning" OR "language teaching" OR "second language" OR "foreign language" OR EFL OR ESL). The methodology draft identified Scopus, Web of Science, ERIC, ScienceDirect, and Google Scholar as relevant database sources. Because the supplied materials do not preserve a completed database-by-database search log or PRISMA count, no unsupported identification or exclusion totals are reported here.

Studies were eligible when they investigated an identifiable AI-supported technology or system in a language-learning or language-teaching context, reported empirical research, were available in full text for assessment, fell within the 2023–2024 review period represented in the final dataset, and provided findings relevant to at least one research question. Non-empirical commentary, publications without a language-learning context, general digital tools without an AI component, inaccessible full texts, and records outside the defined language and publication scope were excluded.

Data Extraction and Thematic Analysis

A standardized extraction framework recorded study identification, country or context, educational level, participants or sample, research design, data source, AI tool, language skill or domain, main findings, benefits, challenges, research gaps or implications, initial codes, final themes, inclusion status, evidence source, and extraction confidence. The extraction notes explicitly avoided inventing unavailable values; missing methodological details were retained as not reported where necessary.

Thematic synthesis proceeded from initial coding to broader categories and higher-order themes. Benefit coding covered language performance and skill development; feedback, scaffolding, and personalization; motivation, engagement, and affect; autonomy, access, and usability; teacher support and pedagogical productivity; assessment and evaluation support; and creativity, multimodality, and cultural learning. Challenge coding covered accuracy, reliability, and hallucination; academic integrity and assessment fairness; over-reliance and reduced critical thinking; AI literacy and training; privacy, bias, ethics, and cultural risk; human interaction and teacher-role limitations; and technical, usability, and access constraints. The resulting patterns were then interpreted through CALL, TAM, and SCT. Formal JBI/MMAT/CASP quality scoring remained pending in the supplied extraction workbook, so the synthesis does not report a completed numerical quality appraisal result.

FINDINGS

Overview of Thematic Findings

The qualitative thematic synthesis of 50 studies published between 2023 and 2024 identified two broad dimensions of AI-assisted language learning: (1) pedagogical benefits, namely the positive learning and teaching outcomes reported in the reviewed studies, and (2) pedagogical, technological, and ethical challenges, namely the limitations, risks, and constraints associated with the use of AI. Within these dimensions, seven benefit-related subthemes and seven challenge-related subthemes were identified. The coding was non-exclusive. Therefore, a single study could contribute evidence to more than one subtheme. For example, a

study examining ChatGPT-assisted writing could simultaneously provide evidence concerning writing development, automated feedback, learner motivation, technology acceptance, and over-reliance.

The analysis showed that the most frequently identified benefit was feedback, scaffolding, and personalization (n = 22; 44%), followed by language performance and skill development (n = 21; 42%), and teacher support and pedagogical productivity (n = 18; 36%). In contrast, the most frequent challenge concerned accuracy, reliability, and hallucination (n = 19; 38%), followed by AI literacy, prompting, and training needs (n = 14; 28%) and over-reliance and reduced critical thinking (n = 13; 26%).

Table 1. Thematic Findings Across the 50 Reviewed Studies

Theme	Subtheme	n	%	Interpretation
Benefits	Language performance & skill development	21	42%	Studies report support for writing, grammar, vocabulary, speaking, reading/lexical work, or broader linguistic performance.
Benefits	Feedback, scaffolding & personalization	22	44%	AI frequently functions as iterative feedback, individualized assistance, revision support, or an adaptive scaffold.
Benefits	Motivation, engagement & affect	16	32%	Reported affective outcomes include motivation, interest, satisfaction, grit, rapport, or positive attitudes.
Benefits	Autonomy, access & usability	15	30%	AI supports self-directed and informal learning through perceived usefulness, accessibility, convenience, and ease of use.
Benefits	Teacher support & pedagogical productivity	18	36%	AI assists with feedback, materials, course design, teaching preparation, or workload-related tasks.
Benefits	Assessment & evaluation support	10	20%	AI is used for scoring, judgment, detection, accuracy assessment, or other evaluative functions.
Benefits	Creativity, multimodality & cultural learning	8	16%	AI supports creative tasks, multimodal composing, literary/cultural work, retelling, or inquiry-oriented activity.
Challenges	Accuracy, reliability & hallucination	19	38%	Concerns include inaccurate information, unstable evaluation, weak deep/pragmatic feedback, or uncertain output quality.
Challenges	Academic integrity, plagiarism & assessment fairness	9	18%	AI complicates authorship, plagiarism detection, cheating control, and the fairness of writing assessment.
Challenges	Over-reliance & reduced critical thinking	13	26%	Studies caution that excessive dependence may displace independent thinking, creativity, or learner effort.
Challenges	AI literacy, prompting & training needs	14	28%	Effective use requires prompting, verification, teacher development, digital competence, and critical AI literacy.
Challenges	Privacy, bias, ethics & cultural risk	8	16%	Privacy, algorithmic/cultural bias, ethics, authenticity, and child-safety concerns appear across contexts.
Challenges	Human interaction & teacher-role limitations	6	12%	Human pedagogical judgment, interaction, and feedback remain important where AI support is insufficient.
Challenges	Technical/usability & access constraints	2	4%	A smaller subset explicitly reports technical, access, or system-use difficulties.

Feedback, Scaffolding, and Personalization as the Dominant Benefit

The most prominent theme across the corpus was feedback, scaffolding, and personalization, identified in 22 of the 50 studies (44%). The finding suggests that the principal pedagogical value of generative AI lies not merely in its ability to produce complete texts but in its capacity to provide immediate and individualized assistance during learning. In writing contexts, ChatGPT was employed to identify linguistic errors, propose revisions, generate ideas, restructure text, and provide formative feedback. Algaraady and Mahyoob (2023), for example, showed that ChatGPT was relatively effective at identifying surface-level writing problems, although deeper structural and pragmatic issues remained difficult for the system to address. Guo and Wang (2024) similarly found that ChatGPT generated extensive feedback across content, organization, and language dimensions. Evidence from intervention research further indicates that AI feedback can contribute to writing development. Mahapatra (2024) reported improvement in ESL academic writing following ChatGPT-supported formative feedback, while Ghafouri et al. (2024) demonstrated that a structured ChatGPT-based writing protocol supported both learners' writing development and teachers' self-efficacy.

Learners also actively determine the quality of AI scaffolding. Yan (2024) found substantial differences in how L2 writers sought and processed ChatGPT feedback, suggesting that learner agency and metacognitive regulation influence whether AI feedback becomes educationally productive. Thus, AI feedback should not be conceptualized simply as automatic correction. Its educational value depends on an iterative process in which learners request assistance, evaluate the response, make revisions, and assume responsibility for the final linguistic decision.

Language Performance and Skill Development

The second most frequent benefit was language performance and skill development, identified in 21 studies (42%). The strongest evidence was concentrated in writing, although improvements were also reported in vocabulary, grammar, speaking, and literary interpretation. Boudouaia et al. (2024) reported that learners receiving ChatGPT-4-assisted writing instruction outperformed comparison learners in several dimensions, including task achievement, coherence, grammatical accuracy, and lexical performance. Similarly, Song and Song (2023) found improvements in both academic writing and learning motivation among EFL learners exposed to AI-assisted instruction. Positive results were not restricted to writing. Mugableh (2024) identified stronger vocabulary outcomes among learners using ChatGPT-generated vocabulary activities, while Kucuk (2024) reported improvements in grammar learning following ChatGPT-supported instruction.

Regarding speaking, Muniandy and Selvanathan (2024) found that learners perceived ChatGPT as a useful partner for learning to speak. However, the speaking evidence remains much smaller than the writing evidence. This imbalance indicates that the current literature provides stronger support for AI-assisted writing than for oral-language development. Rees and Lew (2024) provide an important counterpoint. Their comparison between GPT-generated lexical definitions and dictionary definitions did not demonstrate a clear superiority of AI-generated definitions. This finding reinforces the need to avoid interpreting AI availability as automatic evidence of superior learning effectiveness. The findings, therefore, indicate that generative AI can support language performance, but the strength of evidence varies substantially across language domains.

Teacher Support and Pedagogical Productivity

Teacher-related benefits appeared in 18 studies (36%). AI was described as potentially useful for material generation, instructional preparation, feedback, lesson design, assessment support, and personalized learning. Teachers in several studies recognized ChatGPT's usefulness while simultaneously expressing reservations regarding its classroom implementation (Allehyani & Algamdi, 2023; Gao et al., 2024; Mabuan, 2024; Mohamed, 2024; Ulla et al., 2023).

Guo and Wang (2024) provide a particularly relevant example. ChatGPT produced more feedback than human teachers and distributed comments across several writing dimensions, creating opportunities to reduce teachers' feedback burden. However, teachers also identified weaknesses involving relevance, readability, and excessive feedback length. Yeh (2024) further demonstrates that generative AI may support pedagogical innovation when embedded within inquiry-oriented instructional design. These findings suggest that AI's

contribution to teachers is strongest when the technology is integrated into a pedagogical framework rather than used as an autonomous replacement for instructional decision-making.

Motivation, Engagement, and Learner Autonomy

Affective benefits were identified in 16 studies (32%), while autonomy, access, and usability appeared in 15 studies (30%). Song and Song (2023) found that AI-assisted writing contributed not only to writing outcomes but also to learners' motivation. Ghafouri (2024) reported positive effects involving L2 grit when ChatGPT was incorporated into a structured teacher-student rapport protocol. Similarly, Kim and Park (2023) found that young learners considered AI-generated role-play materials more interesting than conventional textbook scripts. Klimova et al. (2024) reported generally favorable student attitudes toward ChatGPT in university foreign-language education.

The autonomy dimension was particularly evident in research on informal digital learning. Liu et al. (2024) and Liu and Ma (2024) demonstrated that learners use generative AI outside formal classrooms for individualized English-language practice. Xiao and Zhi (2023) likewise found that learners valued immediate feedback, idea generation, and personalized assistance. However, increased learner autonomy through AI creates a paradox. Continuous availability can strengthen self-directed learning but can also facilitate dependence. Consequently, productive AI-supported autonomy requires the ability to decide when assistance is appropriate and when independent performance is necessary.

AI for Assessment and Evaluation

Ten studies (20%) contributed evidence concerning assessment or evaluation. Mizumoto and Eguchi (2023) demonstrated the potential of AI language models for automated essay scoring, while Mizumoto et al. (2024) found substantial correspondence between GPT-based writing-accuracy assessment and human evaluation. Cho (2023) similarly investigated ChatGPT's ability to judge the nativelikeness of sentences produced by native and Korean EFL writers. Such studies demonstrate that generative AI can perform evaluative functions beyond text generation.

However, assessment applications also expose major limitations. Ibrahim (2023) found that AI detectors used to identify generated writing produced inconsistent results, while Alexander et al. (2023) demonstrated difficulties experienced by teachers attempting to distinguish AI-generated from human-generated writing. Tsai et al. (2024) further showed that AI-assisted revision can improve essay scores, but this improvement simultaneously generates questions concerning assessment fairness. Therefore, AI may be valuable as an assessment-support technology, but the findings do not support its use as an autonomous high-stakes decision-maker.

Accuracy and Reliability as the Primary Challenge

The most common challenge was accuracy, reliability, and hallucination, appearing in 19 studies (38%). Algaraady and Mahyoob (2023) demonstrated that ChatGPT was substantially more successful with surface-level errors than with deeper pragmatic or structural problems. Mohamed (2024) similarly reported teacher concerns involving inaccurate information and bias. Nugroho et al. (2024) found that learners themselves recognized the possibility of receiving inaccurate information from ChatGPT. Gao et al. (2024) likewise identified reliability concerns among university EFL teachers considering integration of LLMs.

Even where AI performed strongly, verification remained necessary. Mizumoto et al. (2024), for instance, reported strong correspondence between AI and human judgments of writing accuracy, yet generative variability and contextual interpretation remain relevant limitations. Consequently, the findings support the principle that fluency should not be equated with accuracy. Generative AI can produce highly plausible language even when the underlying content or interpretation is questionable.

AI Literacy, Over-Reliance, and Critical Thinking

AI literacy emerged in 14 studies (28%), while 13 studies (26%) highlighted risks associated with overreliance or reduced critical thinking. Teachers and learners require competencies that extend beyond the

basic operation of an AI platform. Allehyani and Algamdi (2023), Alrishan (2023), Gao et al. (2024), Klimova et al. (2024), and Xiao and Zhi (2023) collectively indicate the importance of understanding AI capabilities, prompting strategies, verification, and responsible educational use.

The findings also indicate that inappropriate reliance on AI may interfere with cognitive development. Cong-Lem et al. (2024) reported teacher concerns that AI-assisted plagiarism and excessive dependence could undermine original thinking and language development. Similar concerns appeared in Derakhshan and Ghiasvand (2024), Mohamed (2024), Ulla et al. (2023), and Yuan et al. (2024). Accordingly, AI literacy should be understood as critical literacy rather than operational literacy. Learners need to question AI-generated information, compare alternative sources, detect potential bias, acknowledge AI assistance, and retain responsibility for linguistic and academic decisions.

Academic Integrity, Ethics, and Cultural Risk

Nine studies (18%) addressed academic integrity, plagiarism, or assessment fairness, while eight (16%) addressed privacy, ethics, bias, or cultural concerns. Alexander et al. (2023), Cong-Lem et al. (2024), and Ibrahim (2023) demonstrate that AI-generated writing disrupts conventional assumptions regarding authorship and plagiarism detection. Detection technology provides only a partial solution because both human detection and automated detectors are imperfect. Therefore, relying solely on AI-detection systems may create additional fairness problems.

The cultural dimension also deserves attention. Zheng and Stewart (2024) found that AI-generated moral dilemma materials could reproduce culturally dominant value patterns, demonstrating that apparently neutral materials may encode particular ideological or cultural assumptions. Similarly, Gao et al. (2024), Mohamed (2024), and Yuan et al. (2024) raised concerns involving privacy, misinformation, bias, authenticity, and ethical use. The evidence, therefore, supports a shift from AI prohibition to AI governance, involving transparent use policies, learner disclosure, process-oriented assessment, verification, and explicit AI literacy education.

AI Technologies Represented in the Corpus

The corpus was dominated by ChatGPT and closely related generative-AI systems. In the 50-study candidate dataset, 37 studies were coded specifically as ChatGPT. In contrast, the remaining studies involved ChatGPT-4, broader generative AI, generative AI with detectors, AI detectors, multiple AI tools, AI writing tools, AI language models, OpenAI GPT, or combined ChatGPT/LLM categories. Functionally, these technologies were used as conversational agents, feedback and revision systems, content and exercise generators, personalized learning supports, assessment and scoring tools, and AI detection technologies. The distribution indicates that the 2023–2024 evidence base is primarily a generative-AI/LLM literature rather than a balanced survey of all forms of educational AI.

DISCUSSION

Interpretation Through Computer-Assisted Language Learning

The coding analysis showed that 41 of the 50 studies (82%) were analytically relevant to CALL-related processes. This figure represents a theoretical mapping rather than the number of original studies that explicitly use CALL as their theoretical framework. The findings strongly align with the principles of Computer-Assisted Language Learning because AI was employed for language practice, writing, feedback, vocabulary, grammar, speaking, assessment, materials development, and informal language learning. Traditional CALL emphasizes the pedagogical value of computer-mediated input, interaction, feedback, and opportunities for language production (Chapelle, 2001). Generative AI extends these affordances by allowing learners to initiate open-ended interaction and receive dynamically generated responses.

For example, ChatGPT can generate vocabulary exercises (Mugableh, 2024), support grammar learning (Kucuk, 2024), provide writing feedback (Guo & Wang, 2024; Mahapatra, 2024), generate speaking materials (Kim & Park, 2023), support academic writing (Song & Song, 2023), and create adaptive instructional materials

(Yeh, 2024). The findings, therefore, suggest a transition from relatively static CALL toward what may be conceptualized as generative and adaptive CALL, in which learning resources are produced dynamically in response to learner input. Nevertheless, CALL theory also emphasizes pedagogical appropriateness. Technology does not become educationally valuable merely because it is technologically sophisticated. The inconsistent findings involving lexical definitions (Rees & Lew, 2024), automated feedback (Algaraady & Mahyoob, 2023), and assessment reliability (Ibrahim, 2023) demonstrate that AI output must still be evaluated in relation to specific language-learning objectives.

Interpretation Through the Technology Acceptance Model

Sixteen studies (32%) were analytically associated with TAM variables, while six studies in the extracted dataset explicitly incorporated TAM-related constructs. Davis's (1989) Technology Acceptance Model proposes that perceived usefulness and perceived ease of use influence attitudes and behavioral intentions toward technology adoption. This pattern was clearly reflected in the reviewed evidence. Alrishan (2023) found that usefulness and ease of use significantly contributed to pre-service teachers' intention to use ChatGPT. Dehghani and Mashhadi (2024) similarly identified usefulness, ease of use, self-efficacy, enjoyment, system quality, and subjective norms as important elements of teacher acceptance. Among learners, Liu and Ma (2024), Liu et al. (2024), Shaikh et al. (2023), and Vo and Nguyen (2024) reported generally favorable perceptions concerning accessibility, usefulness, usability, and AI-supported learning.

However, the synthesis identifies an important limitation of TAM when applied to educational AI: Technology acceptance is not equivalent to language-learning effectiveness. A learner may perceive ChatGPT as useful and easy to use without demonstrating measurable gains in language proficiency. Therefore, studies examining user acceptance should not be interpreted as equivalent to experimental evidence of learning improvement. Future AI-assisted language-learning research should combine TAM variables with objective outcomes such as writing performance, vocabulary development, speaking proficiency, delayed retention, or independent performance without AI assistance.

Interpretation Through Sociocultural Theory

The strongest analytical linkage emerged with Sociocultural Theory, with 42 studies (84%) containing findings relevant to mediation, scaffolding, interaction, learner agency, feedback, personalization, or teacher support. Again, this percentage represents a theoretical interpretation rather than evidence that 42 original studies explicitly adopted SCT. According to Vygotsky (1978), learning and cognitive development are mediated through culturally constructed tools and social interaction. Lantolf and Thorne (2006) extended this perspective to second-language development, emphasizing mediation and assisted performance. From this perspective, generative AI can be conceptualized as a digital mediational artifact. The dominance of feedback and scaffolding in the present synthesis supports this interpretation. Learners use AI to receive explanations, alternatives, examples, corrections, prompts, and feedback before revising their own performance (Ghafouri et al., 2024; Guo & Wang, 2024; Mahapatra, 2024; Yan, 2024).

However, SCT also provides a theoretical explanation for the risks identified in the challenge analysis. Scaffolding should ultimately support movement toward more independent performance. If the learner continually transfers problem-solving to AI, assistance may become substitution rather than mediation. This distinction explains why studies simultaneously report both personalization and learner autonomy alongside concerns about overreliance and declining critical thinking (Cong-Lem et al., 2024; Derakhshan & Ghiasvand, 2024; Mohamed, 2024; Yuan et al., 2024). Thus, the central pedagogical issue is not simply whether learners use AI, but whether its use progressively strengthens or weakens learner agency.

Integration of CALL, TAM, and Sociocultural Theory

The three theoretical perspectives explain different dimensions of the findings and should therefore be understood as complementary. CALL explains what AI enables pedagogically. It accounts for technology-mediated feedback, practice, language production, assessment, and materials development. TAM explains whether teachers and learners are willing to adopt the technology. Perceived usefulness, ease of use, self-

efficacy, and attitudes influence adoption. Sociocultural Theory explains how AI-supported interaction may contribute to development. AI becomes educationally meaningful when it operates as a mediational and scaffolding resource without displacing learner agency and human interaction.

The integration can be represented conceptually as: AI technological affordances (CALL) → user acceptance and adoption (TAM) → mediated learning and scaffolding processes (SCT) → language-learning outcomes. However, this process is moderated by AI literacy, teacher mediation, accuracy, ethics, task design, and learner agency. Consequently, the synthesis supports a human-AI collaborative language-learning model rather than an AI-substitution model.

Human-AI Collaboration as the Central Pedagogical Model

The findings consistently indicate that AI performs particularly well in activities requiring speed, repetition, language generation, preliminary feedback, personalization, and information retrieval. Teachers remain essential where instruction requires contextual judgment, interpersonal sensitivity, motivational support, curricular decisions, ethical supervision, and nuanced assessment. Ahmed (2023), for example, showed that learners were more satisfied with teacher-mediated interaction than with ChatGPT-assisted learning on several dimensions. Algaraady and Mahyoob (2023) demonstrated that human instructors were more effective at detecting deeper pragmatic errors. Guo and Wang (2024) showed that AI feedback can complement rather than eliminate teacher feedback.

The findings therefore reject two extreme positions: that AI should completely replace language teachers, and that AI has no educational value. Instead, the evidence supports teacher-AI complementarity.

Implications for Language Education

Several implications emerge from the synthesis. First, AI should be integrated primarily as a scaffolding and feedback technology, because these functions received the strongest support across the corpus. Second, institutions should explicitly teach AI literacy, including prompt formulation, output verification, awareness of bias, privacy protection, ethical attribution, and appropriate disclosure of AI use.

Third, as the assessment needs to be redesigned, traditional take-home writing tasks are increasingly vulnerable to AI-assisted cheating. Process-based assessment, oral defense, iterative drafts, reflection logs, and transparent AI-use declarations may provide more defensible evidence of learner competence. Fourth, teachers require professional development not only on how to operate AI applications but also on when AI should and should not be used pedagogically. Finally, future research should move beyond student perceptions and short-duration interventions. More longitudinal and controlled studies are needed to determine whether AI-supported gains remain when learners perform independently.

Research Gaps

The thematic synthesis indicates several priorities for future research. The current evidence remains disproportionately concentrated on writing. Although several studies have investigated grammar, vocabulary, speaking, reading, and cultural learning, these areas remain underrepresented compared with writing-related applications. More research is therefore required on speaking, listening, pronunciation, pragmatics, interactional competence, and intercultural communication. Research is also heavily concentrated in university and adult EFL settings. Younger learners, primary and secondary school learners, and learners with diverse educational needs remain comparatively under-researched.

Future studies should also distinguish between different forms of AI assistance. Brainstorming, translation, automated feedback, full-text generation, conversation practice, assessment, and vocabulary support involve distinct cognitive processes and should not be grouped under a single label such as “ChatGPT use.” Finally, longitudinal research is needed to examine an especially important question: Does repeated AI-assisted performance eventually produce stronger independent language ability, or merely stronger AI-assisted performance? This distinction should become a central concern in the next generation of AI-assisted language-learning research.

Methodological Considerations

The findings should be interpreted in light of the methodological logic of the review. A Systematic Literature Review with qualitative thematic synthesis was selected to identify, organize, compare, and interpret empirical evidence without statistically pooling heterogeneous outcomes. This choice is appropriate for a literature in which studies differ in learner levels, language skills, research designs, data sources, and AI functions. However, it also means that the percentages reported in Table 1 describe how frequently a theme appeared across the corpus rather than the size or strength of any intervention effect.

Coding was non-exclusive, so that a single study could contribute evidence to several benefit, challenge, or theoretical categories; therefore, the subtheme percentages overlap and should not be summed. The same applies to the CALL, TAM, and SCT figures, which represent analytical mapping of the findings onto each framework rather than the number of original studies that explicitly adopted it. The standardized extraction framework retained missing methodological details as not reported rather than inferring them, which protects the integrity of the synthesis but also means that some study characteristics could not be compared systematically.

Limitations of the Review

This review has several limitations. First, the corpus covers only studies published in 2023–2024 and therefore does not capture later developments in AI-assisted language learning. Second, the supplied materials do not preserve a completed database-by-database search log or PRISMA count, so identification and exclusion totals are not reported. Formal JBI/MMAT/CASP quality scoring also remained pending, so the synthesis does not present a completed quality appraisal, and the included studies were not weighted by methodological rigor.

Third, the evidence base is dominated by ChatGPT and closely related generative AI systems, so the findings reflect the generative AI/LLM literature rather than a balanced survey of all forms of educational AI. Fourth, the evidence is concentrated on writing and on university and adult EFL settings, which limits the generalizability of the findings to other language skills, younger learners, and other language contexts. These limitations indicate that the conclusions should be read as a synthesis of reported patterns in the reviewed corpus rather than as definitive causal evidence.

CONCLUSION

The thematic synthesis of 50 studies demonstrates that generative AI has substantial potential for language education, particularly as a tool for feedback, scaffolding, personalization, skill development, learner engagement, autonomy, and teacher support. The most frequently identified benefit was feedback, scaffolding, and personalization (44%), while language performance and skill development appeared in 42% of the studies. In contrast, the most frequent concern was AI accuracy and reliability (38%), followed by the need for AI literacy and training (28%) and risks of over-reliance and reduced critical thinking (26%). The theoretical synthesis demonstrates that CALL, TAM, and Sociocultural Theory address complementary dimensions of AI-assisted language learning. CALL explains technological and pedagogical affordances; TAM explains user acceptance; and Sociocultural Theory explains mediation, scaffolding, interaction, and learner agency. Taken together, the findings indicate that the most pedagogically defensible future for AI in language education is not autonomous AI instruction but human–AI collaborative learning, in which AI expands learning opportunities. At the same time, teachers and learners retain responsibility for judgment, meaning, ethics, and educational decisions.

In terms of practical and policy implications, the synthesis suggests that institutions and teachers should integrate AI primarily as a scaffolding and feedback technology, because these functions received the strongest support across the corpus. Institutions should explicitly teach AI literacy, including prompt formulation, output verification, bias awareness, privacy protection, ethical attribution, and disclosure of AI use. They should provide teachers with professional development on when AI should and should not be used pedagogically. Assessment policies also need to be redesigned to use process-based assessment, iterative

drafts, oral defense, reflection logs, and transparent AI-use declarations in place of take-home tasks that are vulnerable to undetected AI assistance.

For future research, more studies are needed on speaking, listening, pronunciation, pragmatics, interactional competence, and intercultural communication, as well as on younger learners and learners with diverse educational needs. Future studies should distinguish between different forms of AI assistance rather than grouping them as “ChatGPT use,” and longitudinal and controlled designs are needed to determine whether AI-supported gains persist when learners perform independently.

REFERENCES

- Ahmed, M. A. (2023). ChatGPT and the EFL classroom: Supplement or substitute in Saudi Arabia's eastern region. *Information Sciences Letters*, 12(7), 2727–2734. <https://doi.org/10.18576/isl/120704>
- Alenizi, M. A. K., Mohamed, A. M., & Shaaban, T. S. (2023). Revolutionizing EFL special education: How ChatGPT is transforming the way teachers approach language learning. *Innoeduca: International Journal of Technology and Educational Innovation*, 9(2), 5–23. <https://doi.org/10.24310/innoeduca.2023.v9i2.16774>
- Alexander, K., Savvidou, C., & Alexander, C. (2023). Who wrote this essay? Detecting AI-generated writing in second language education in higher education. *Teaching English with Technology*, 23(2), 25–43. <https://doi.org/10.56297/BUKA4060/XHLD5365>
- Algaraady, J., & Mahyoob, M. (2023). ChatGPT's capabilities in spotting and analyzing writing errors experienced by EFL learners. *Arab World English Journal, Special Issue on CALL*, 9, 3–17. <https://doi.org/10.24093/awej/call9.1>
- Allehyani, S. H., & Algamdi, M. A. (2023). Digital competences: Early childhood teachers' beliefs and perceptions of the application of ChatGPT in teaching English as a second language. *International Journal of Learning, Teaching and Educational Research*, 22(11), 343–363. <https://doi.org/10.26803/ijlter.22.11.18>
- Alhammad, A. I. (2024). The impact of ChatGPT on developing Saudi EFL learners' literature appreciation. *World Journal of English Language*, 14(2), 331–341. <https://doi.org/10.5430/wjel.v14n2p331>
- Alrishan, A. M. H. (2023). Determinants of intention to use ChatGPT for professional development among Omani EFL pre-service teachers. *International Journal of Learning, Teaching and Educational Research*, 22(12), 187–209. <https://doi.org/10.26803/ijlter.22.12.10>
- Boudouaia, A., Mouas, S., & Kouider, B. (2024). A study on ChatGPT-4 as an innovative approach to enhancing English as a foreign language writing learning. *Journal of Educational Computing Research*. <https://doi.org/10.1177/07356331241247465>
- Chapelle, C. A. (2001). *Computer applications in second language acquisition: Foundations for teaching, testing, and research*. Cambridge University Press.
- Cho, H. (2023). Analyzing ChatGPT's judgments on the nativelikeness of sentences written by English native speakers and Korean EFL learners. *Multimedia-Assisted Language Learning*, 26(2), 9–32. <https://doi.org/10.15702/mall.2023.26.2.9>
- Cong-Lem, N., Tran, T. N., & Nguyen, T. T. (2024). Academic integrity in the age of generative AI: Perceptions and responses of Vietnamese EFL teachers. *Teaching English with Technology*, 24(1), 28–47. <https://doi.org/10.56297/FSYB3031/MXNB7567>
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340. <https://doi.org/10.2307/249008>
- Dehghani, H., & Mashhadi, A. (2024). Exploring Iranian English-as-a-foreign-language teachers' acceptance of ChatGPT in English language teaching: Extending the technology acceptance model. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-024-12660-9>
- Derakhshan, A., & Ghiasvand, F. (2024). Is ChatGPT an evil or an angel for second language education and research? A phenomenographic study of research-active EFL teachers' perceptions. *International Journal of Applied Linguistics*. <https://doi.org/10.1111/ijal.12561>
- Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, 20, 57. <https://doi.org/10.1186/s41239-023-00425-2>

- Gao, Y., Wang, Q., & Wang, X. (2024). Exploring EFL university teachers' beliefs in integrating ChatGPT and other large language models in language education: A study in China. *Asia Pacific Journal of Education*, 44(1), 29–44. <https://doi.org/10.1080/02188791.2024.2305173>
- Ghafouri, M. (2024). ChatGPT: The catalyst for teacher-student rapport and grit development in L2 class. *System*, 120, 103209. <https://doi.org/10.1016/j.system.2023.103209>
- Ghafouri, M., Hassaskhah, J., & Mahdavi-Zafarghandi, A. (2024). From virtual assistant to writing mentor: Exploring the impact of a ChatGPT-based writing instruction protocol on EFL teachers' self-efficacy and learners' writing skill. *Language Teaching Research*. <https://doi.org/10.1177/13621688241239764>
- Guo, K., & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies*, 29, 8435–8463. <https://doi.org/10.1007/s10639-023-12146-0>
- Ibrahim, K. (2023). Using AI-based detectors to control AI-assisted plagiarism in ESL writing: "The terminator versus the machines." *Language Testing in Asia*, 13, 46. <https://doi.org/10.1186/s40468-023-00260-2>
- Kartal, G. (2024). The influence of ChatGPT on the thinking skills and creativity of EFL student teachers: A narrative inquiry. *Journal of Education for Teaching*. <https://doi.org/10.1080/02607476.2024.2326502>
- Kim, S., & Park, S. H. (2023). Young Korean EFL learners' perception of role-playing scripts: ChatGPT vs. textbooks. *Korean Journal of English Language and Linguistics*, 23, 1136–1153. <https://doi.org/10.15738/kjell.23.202312.1136>
- Klimova, B., Pikhart, M., & Al-Obaydi, L. H. (2024). Exploring the potential of ChatGPT for foreign language education at the university level. *Frontiers in Psychology*, 15, 1269319. <https://doi.org/10.3389/fpsyg.2024.1269319>
- Kucuk, T. (2024). ChatGPT-integrated grammar teaching and learning in EFL classes: A study of Tishk International University students in Erbil, Iraq. *Arab World English Journal*, Special Issue on ChatGPT, 100–111. <https://doi.org/10.24093/awej/ChatGPT.6>
- Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of second language development*. Oxford University Press.
- Liu, G. L., Darvin, R., & Ma, C. (2024). Exploring AI-mediated informal digital learning of English (AI-IDLE): A mixed-method investigation of Chinese EFL learners' AI adoption and experiences. *Computer-Assisted Language Learning*. <https://doi.org/10.1080/09588221.2024.2310288>
- Liu, G., & Ma, C. (2024). Measuring EFL learners' use of ChatGPT in informal digital learning of English based on the technology acceptance model. *Innovation in Language Learning and Teaching*, 18(2), 125–138. <https://doi.org/10.1080/17501229.2023.2240316>
- Liu, M., Zhang, L. J., & Biebricher, C. (2024). Investigating students' cognitive processes in generative AI-assisted digital multimodal composing and traditional writing. *Computers & Education*, 211, 104977. <https://doi.org/10.1016/j.compedu.2023.104977>
- Mabuan, R. A. (2024). ChatGPT and ELT: Exploring teachers' voices. *International Journal of Technology in Education*, 7(1), 128–153. <https://doi.org/10.46328/ijte.523>
- Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed-methods intervention study. *Smart Learning Environments*, 11, 9. <https://doi.org/10.1186/s40561-024-00295-9>
- Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics*, 2(2), 100050. <https://doi.org/10.1016/j.rmal.2023.100050>
- Mizumoto, A., Shintani, N., Sasaki, M., & Teng, M. F. (2024). Testing the viability of ChatGPT as a companion in L2 writing accuracy assessment. *Research Methods in Applied Linguistics*, 3(2), 100116. <https://doi.org/10.1016/j.rmal.2024.100116>
- Mohamed, A. M. (2024). Exploring the potential of an AI-based chatbot (ChatGPT) in enhancing English as a foreign language teaching: Perceptions of EFL faculty members. *Education and Information Technologies*, 29(3), 3195–3217. <https://doi.org/10.1007/s10639-023-11917-z>
- Mugableh, A. I. (2024). The impact of ChatGPT on the development of vocabulary knowledge of Saudi EFL students. *Arab World English Journal*, Special Issue on ChatGPT, 265–281. <https://doi.org/10.24093/awej/ChatGPT.18>

- Muniandy, J., & Selvanathan, M. (2024). ChatGPT, a partnering tool to improve ESL learners' speaking skills: Case study in a public university, Malaysia. *Teaching Public Administration*. <https://doi.org/10.1177/01447394241230152>
- Nugroho, A., Andriyanti, E., Widodo, P., & Mutiaraningrum, I. (2024). Students' appraisals post-ChatGPT use: Students' narrative after using ChatGPT for writing. *Innovations in Education and Teaching International*. <https://doi.org/10.1080/14703297.2024.2319184>
- Rees, G. P., & Lew, R. (2024). The effectiveness of OpenAI GPT-generated definitions versus definitions from an English learners' dictionary in a lexically orientated reading task. *International Journal of Lexicography*, 37(1), 50–74. <https://doi.org/10.1093/ijl/ecad030>
- Shaikh, S., Yayilgan, S. Y., Klimova, B., & Pikhart, M. (2023). Assessing the usability of ChatGPT for formal English language learning. *European Journal of Investigation in Health, Psychology and Education*, 13(9), 1937–1960. <https://doi.org/10.3390/ejihpe13090140>
- Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, 14, 1260843. <https://doi.org/10.3389/fpsyg.2023.1260843>
- Tsai, C. Y., Lin, Y. T., & Brown, I. K. (2024). Impacts of ChatGPT-assisted writing for EFL English majors: Feasibility and challenges. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-024-12722-y>
- Ulla, M. B., Perales, W. F., & Busbus, S. O. (2023). 'To generate or stop generating response': Exploring EFL teachers' perspectives on ChatGPT in English language teaching in Thailand. *Learning: Research and Practice*, 9(2), 168–182. <https://doi.org/10.1080/23735082.2023.2257252>
- Vo, T. K. A., & Nguyen, H. (2024). Generative artificial intelligence and ChatGPT in language learning: EFL students' perceptions of technology acceptance. *Journal of University Teaching and Learning Practice*, 21, 6. <https://doi.org/10.53761/fr1rkj58>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Xiao, Y., & Zhi, Y. (2023). An exploratory study of EFL learners' use of ChatGPT for language learning tasks: Experience and perceptions. *Languages*, 8(3), 212. <https://doi.org/10.3390/languages8030212>
- Yan, D. (2024). Feedback-seeking abilities of L2 writers using ChatGPT: A mixed-method multiple case study. *Kybernetes*. <https://doi.org/10.1108/K-09-2023-1933>
- Yeh, H. C. (2024). The synergy of generative AI and inquiry-based learning: Transforming the landscape of English teaching and learning. *Interactive Learning Environments*. <https://doi.org/10.1080/10494820.2024.2335491>
- Yuan, Y., Li, H., & Sawaengdist, A. (2024). The impact of ChatGPT on learners in English academic writing: Opportunities and challenges in education. *Language Learning in Higher Education*, 14(1), 41–56. <https://doi.org/10.1515/cercles-2023-0006>
- Zheng, Y. D., & Stewart, N. (2024). Improving EFL students' cultural awareness: Reframing moral dilemmatic stories with ChatGPT. *Computers and Education: Artificial Intelligence*, 6, 100223. <https://doi.org/10.1016/j.caeai.2024.100223>