

RECONCEPTUALIZING HUMAN-AI COLLABORATION IN LANGUAGE LEARNING: A PRISMA-BASED SYSTEMATIC REVIEW OF PEDAGOGICAL AND MANAGERIAL INTEGRATION IN HIGHER EDUCATION

Dian Erliyani, Jose Beno, Radini Sinta, Fathur Rokhman, Eko Handoyo, Suwito Eko
Pramono, Agus Wahyudin, Ani Rusilowati, Slamet Riyadi
Universitas Negeri Semarang

Abstract

This study takes the misalignment across three dimensions of technology, pedagogy, and management within the higher education language learning context as its core entry point to re-examine the topic of human-AI collaboration. It adopts a systematic literature review that complies with PRISMA standards, combined with the thematic synthesis method, using 30 peer-reviewed articles published between 2020 and 2025 and indexed in Google Scholar, Scopus, and Web of Science as its research sample. The review finds that to establish trust and enable responsible AI applications, ethical challenges, including data privacy and algorithmic bias, must be resolved first. While AI can improve learning outcomes and student engagement through adaptive feedback and personalized learning pathways, it currently serves only an auxiliary role, with insufficient empirical support to enable full collaborative interaction. Integration at the pedagogical dimension is the core determinant of AI application effectiveness, while integration at the management and institutional governance dimensions remains an underexplored gap in the field. This study constructs an integrated conceptual framework that connects human-AI collaboration, instructional design, and governance structures, advancing theoretical understanding of the field and providing practical support for building adaptive, ethical, and sustainable AI-supported learning environments.

Keywords: Human–AI Collaboration; Artificial Intelligence in Education; Language Learning; Pedagogical Integration; Educational Governance; Higher Education; Systematic Literature Review

Introduction

Over the past decade, artificial intelligence has become one of the driving forces behind global educational transformation, reshaping how educational institutions design, deliver, and assess learning. It has not only improved operational and teaching efficiency but also driven a paradigm shift in education toward adaptive, personalized, and data-driven models.

AI applications in language learning are the fastest-growing and most research-intensive sector. Leveraging the complex, multi-dimensional nature of language proficiency, three categories of

technologies have been deployed: chatbots, automated assessment tools, and adaptive learning platforms. These technologies cover all levels of education, with the core goal of improving learners' language proficiency.

Three 2025 studies in the field of language education string together the full logic of AI's full-chain empowerment for the entire language learning process: Albedah (2025) verified that AI boosts learning efficiency through immediate feedback and personalized adaptation; Liu et al. (2025) found that AI chatbots can improve the oral proficiency of EFL students; Chen et al.

(2025) pointed out that AI assessment is more accurate and efficient than human assessment.

Building on the empirical study of technology adoption in educational systems conducted in the preceding text, we introduce two classic theories to construct our supporting layer: the Technology Acceptance Model (TAM) proposed by Davis in 1989, which explains the micro-level mechanism of individual technology acceptance, and Rogers' Diffusion of Innovations theory, which explains the gradual, uneven adoption differences among macro-level groups. The two theories form a mutually complementary supporting logic.

Despite the huge potential of artificial intelligence in language learning, several fundamental challenges remain to be addressed. The most pressing issue is that institutions of all types are not yet prepared to implement AI in a systematic, sustainable, and fair manner. Zhu and Wang (2025) note that most existing research focuses on teaching and technical aspects, while research on management dimensions such as policy-making, transformative leadership, and data governance is severely lacking.

This study points out that a recurring common problem exists in AI applications in the education sector: there remains a lack of a unified, comprehensive ethical regulatory framework that covers student data privacy, algorithmic transparency, and accountability for automated decision-making. Implementation challenges span four major dimensions: technology, structure, management, and ethics.

The popularization of AI applications is currently accelerating. Most real-world deployment scenarios lack adequate systematic supporting preparations, and the urgency of addressing various challenges has increased substantially. Alhusaiyan (2025) notes that the implementation of educational AI without a unified

governance framework will pose three specific risks.

This study is tailored for the secondary and higher education sectors that are currently advancing AI integration in language learning. The existing literature points out that implementing educational AI requires transformational leadership, an innovative organizational culture, and preparedness in digital literacy.

While previous studies have confirmed the effectiveness of AI for language learning, gaps remain in this field regarding governance, institutional readiness, and management capacity. This study employs the PRISMA method to conduct a dual-perspective review.

This study integrates research on AI for language learning and the perspective of educational management. It focuses on two dimensions that have received limited attention in previous academic research: institutional readiness and governance. It adopts a systematic literature review method that adheres to PRISMA guidelines to comprehensively examine the real-world applications of AI in language learning at colleges and universities.

This study establishes three core research objectives: first, to analyze the implementation of AI in language learning scenarios from the perspective of educational management; second, to identify the core influencing factors behind the success or failure of AI rollout across all types of educational institutions; third, to develop an integrated conceptual framework that can support educational leaders and policymakers in carrying out planning, implementation, and evaluation work.

This study delivers both theoretical and practical contributions: At the theoretical level, it integrates the educational management perspective into mainstream research on AI integration in language learning, enriching existing knowledge in this field. At the practical level, it provides

evidence-based guidance for three types of educational stakeholders, supporting them to design effective, equitable, and sustainable AI implementation strategies.

Method

This study adopts a Systematic Literature Review (SLR). The authors propose using a concept-adapted PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) framework to focus on human-AI collaboration and integration in language learning within higher education contexts. This method can systematically integrate scattered research, produce conclusions that are descriptive, analytical, and comparable, and generate cumulative conceptual contributions. It fits three core tasks: sorting out the fast-updating research landscape, identifying convergences and divergences across existing research, and building the theoretical foundation for the digital transformation period.

To keep pace with the rapid development of AI technology, this study established a comprehensive literature processing workflow that safeguards the research's innovation and scientific relevance, with all components aligned with the specific needs of this study's scenario. We selected three major general academic databases—Google Scholar, Scopus, and Web of Science—and set a search time window spanning 2020 to 2025. This arrangement not only ensures the novelty of this study's findings but also limits included research outputs to those that reflect the latest AI education ecosystem. We combined core keywords specifically including "artificial intelligence", "language learning", "human-AI collaboration", and "educational management" using Boolean operators, flexibly applying AND/OR to reduce missed retrievals caused by terminology differences, and completed literature screening in four steps in strict adherence to the PRISMA protocol.

This study established the following literature inclusion criteria: articles published in authoritative journals ranked in JCR Q1 or Q2, written in English or Indonesian, and with content closely related to the integration of artificial intelligence (AI) for teaching and management in language learning contexts. During the screening process, three categories of literature were systematically excluded to guarantee the completeness and coherence of the analytical corpus. This study adopted the thematic synthesis method for manual analysis to identify consistent patterns across studies, clarify the structural relationships among variables, and develop dynamic core themes for AI integration in the language learning field. The rationale for adopting this method is cited from Sugiyono (2021) and Yusuf (2020). Ultimately, this study successfully developed a comprehensive conceptual framework for human-machine collaboration within the contemporary higher education ecosystem.

Research Instruments (Systematic Literature Review)

The research tool developed for this study's systematic literature review adopts a format that pairs a structured data extraction form with a thematic coding framework. Its core design goals are to ensure procedural consistency, analytical transparency, and replicability throughout the review workflow. The core methodological premise upheld by this study holds that the quality of a systematic literature review relies not only on the comprehensiveness of literature source retrieval but more fundamentally on the accuracy, consistency, and clarity of the procedures applied in the data extraction and analysis stages. Without this set of operationalizable standardized tools, the consistency of interpretations across different coders cannot be fully guaranteed.

The data extraction tool adopted in this study identifies, records, and organizes core

variables, including human-machine collaboration, teaching embeddedness, management embeddedness, and others. It then systematically codes each included piece of literature in accordance with pre-set operational indicators, eliminates subjective bias, and supports accurate quantitative comparison of variables across different studies.

This study employs thematic synthesis to analyze its data, categorizing research findings from individual coding into thematic clusters covering three core dimensions of AI implementation in higher education language learning: mainstream trends, conceptual tensions, and unresolved

empirical gaps. This classification is interpretive and constructivist. It relies on professional analytical judgment to position findings within the thematic map, and is by no means a mechanical process. This approach can support in-depth cross-study interpretation, build a multi-dimensional conceptual model that integrates technology, instructional design, and educational governance, identify opportunities for theoretical development, and produce practical recommendations for policymakers to reference. The rationality of this method has been corroborated by the studies of Yusuf (2020) and Sugiyono (2021).

Operational Instruments (Data Extraction & Coding Sheet)

Tabel 1.Data Extraction Table

No	Author (year)	Country/Context	AI Tool/Type	Human-AI Collaboration (Level)	Pedagogical Integration (Design)	Managerial Integration (Policy/Infra)	Learning Outcomes (Cognitive/Affective)	Ethics/Governance	Key Findings
1	Lee & Kim (2025)	South Korea / CALL	AI Chatbot	High	Conversational AI for speaking practice	Need institutional digital readiness	Improved speaking confidence and engagement	Privacy and AI dependency concerns	AI chatbots significantly improved oral interaction skills
2	Nawir et al. (2024)	Indonesia / EFL	Generative AI	Medium	AI-assisted writing and grammar correction	Infrastructure and teacher support needed	Enhanced writing accuracy and motivation	Academic integrity issues	Students perceived AI as useful for writing support
3	Sari et al. (2024)	Indonesia	ChatGPT	High	Integrated AI writing support in coursework	Limited governance framework	Improved creativity and efficiency	Plagiarism and misinformation risks	ChatGPT increased student productivity in academic tasks
4	Rahman & Yusuf (2024)	Malaysia / Personalized learning	Adaptive AI systems	High	Personalized adaptive language learning pathways	Need AI-ready infrastructure	Improved personalization and learner autonomy	Bias and transparency concerns	AI enables individualized language learning experiences
5	Chen et al. (2024)	Global	AI applications in language learning	High	AI-supported assessment, tutoring, and feedback	Calls for institutional AI policy	Improved language performance and engagement	Ethics, data privacy, and fairness	AI positively impacts language learning across contexts
6	Frontiers in Education (2025)	International	Generative AI and LLMs	High	AI-assisted collaborative learning	Need governance and digital literacy programs	Enhanced collaboration and cognitive engagement	Ethical AI use and reliability	LLMs support active learning when guided appropriately

No	Author (year)	Country/Context	AI Tool/Type	Human-AI Collaboration (Level)	Pedagogical Integration (Design)	Managerial Integration (Policy/Infrastructure)	Learning Outcomes (Cognitive/Affective)	Ethics/Governance	Key Findings
7	Springer AI Education (2025)	Global	AI in education ecosystems	High	Human-AI collaborative learning models	Strong governance and policy emphasis	Improved adaptability and learning analytics	Governance, fairness, and accountability	Human-centered AI ecosystems are critical for sustainable education
8	Nature Reference Study (2026)	Global	AI literacy frameworks	Medium	AI literacy integration into curriculum	Institutional AI competency frameworks	Improved AI awareness and digital competence	Responsible AI governance	AI literacy should be embedded into higher education curricula
9	Du & Daniel (2024)	New Zealand / EFL speaking practice	AI chatbots	Medium-High	Chatbot-supported speaking practice with feedback	Limited discussion of institutional support	Improved speaking skills, reduced anxiety	Privacy, bias, overreliance	AI chatbots improve EFL speaking confidence and interaction
10	Ekizer (2025)	Turkey / EFL education	Chatbots, AWA, VR-AI	Medium	AI-integrated EFL interventions	Technology integration in ELT	Positive effects on language achievement	Need ethical AI use	AI interventions significantly improve EFL outcomes
11	Xiao, Zhang & He (2024)	China / language education	AI chatbots	Medium	Emotion-aware language learning	Minimal policy discussion	Enhanced motivation but possible anxiety	Emotional dependency, ethics	Chatbots affect learner emotions positively and negatively
12	Cisłowska & Pena-Acuna (2024)	Spain / additional language education	Chatbots	Medium	Chatbot-assisted language learning	Need teacher training and infrastructure	Improved skills, attitudes, engagement	Ethical concerns and limitations	Chatbots support language learning but require careful implementation
13	Wiboolysarin et al. (2025)	Thailand & international	AI-driven chatbots	High	Personalized learning and feedback	Need institutional readiness	Improved competence and feedback responsiveness	Data privacy and pedagogical ethics	Chatbots enhance L2 acquisition and intercultural learning
14	Huang, Hew & Fryer (2022)	Hong Kong / language learning	Chatbots	Medium	Conversational learning support	Limited governance discussion	Positive engagement and language practice	Need transparency and teacher guidance	Chatbots useful for interaction and practice
15	Zhu & Wang (2025)	Global	AI in language education	High	AI-enhanced language learning design	Calls for policy and governance	Improved personalization and efficiency	Ethics, bias, access inequality	AI research in language education growing rapidly
16	Li, Zhou & Chiu (2025)	Global / SDT perspective	ChatGPT and AI chatbots	High	Teacher-supported AI learning	Need teacher roles and policy guidance	Supports autonomy and engagement	Academic integrity and governance	Teachers remain central in AI-supported language learning
17	Woo & Choi (2021)	USA & South Korea	AI-based language learning tools	Medium	AI-assisted assessment and feedback	Need teacher preparation	Improved feedback and adaptive learning	Teacher readiness and transparency	AI tools improve individualized learning

No	Author (year)	Country/Context	AI Tool/Type	Human-AI Collaboration (Level)	Pedagogical Integration (Design)	Managerial Integration (Policy/Infrastructure)	Learning Outcomes (Cognitive/Affective)	Ethics/Governance	Key Findings
18	Alfredo et al. (2023)	Global / LA & AIED	Learning analytics & AI	High	Human-centred AI design	Strong focus on governance and trust	Improved trust and usability	Safety, reliability, agency	Human-centred approaches are critical for trustworthy AI
19	UNESCO (2021)	Global policy guidance	AI in education systems	High	AI-supported personalized learning	Strong policy/infrastructure focus	Improved access, adaptive learning	Ethics, equity, transparency, governance	AI should support inclusive and equitable education
20	Williams et al. (2023)	Global critical perspectives	Generative AI, AIED	High	Critical AI literacy integration	Governance and regulation emphasized	Mixed impact on cognition and agency	Surveillance, bias, power relations	AI in education shaped by political/economic interests
21	Kasneci et al. (2023)	Global / higher education	ChatGPT	High	AI-assisted tutoring and assessment	Need institutional AI policy	Enhanced learning support and productivity	Academic integrity and hallucination	ChatGPT has transformative but risky educational potential
22	Tlili et al. (2023)	Global	Generative AI	High	Human-AI co-creation in learning	Need governance frameworks	Creativity and personalization gains	Ethics, bias, misuse	Generative AI reshapes pedagogy and assessment
23	Zawacki-Richter et al. (2021)	Global	AI applications in HE	Medium	Adaptive systems and analytics	Limited institutional readiness	Improved personalization	Ethical and data concerns	AI in HE still dominated by computer science focus
24	Holmes et al. (2022)	International	AIED systems	High	Human-centered AI learning	Need trustworthy AI infrastructure	Supportive but dependent on design	Fairness, accountability, transparency	Human-centered AIED is essential
25	Rahman et al. (2025)	Indonesia	ChatGPT	High	Integrated into writing instruction	Limited institutional governance	Improved writing support and confidence	Plagiarism concerns	Students perceive ChatGPT as useful writing assistant
26	Wijopate et al. (2024)	Indonesia	AI learning platform	Medium	AI-assisted classroom instruction	Infrastructure challenges	Improved engagement	Teacher readiness concerns	AI supports classroom engagement when guided
27	European Commission (2022)	Europe	AI and data in education	High	Digital pedagogy frameworks	Strong governance emphasis	Digital competency enhancement	GDPR, ethics, inclusion	Trustworthy AI requires policy alignment
28	OpenAI-related Arxiv (2024)	Global	LLMs and ChatGPT	High	LLM-supported knowledge generation	Need infrastructure and AI literacy	Improved efficiency and ideation	Bias and misinformation	LLMs can augment learning but need safeguards

Tabel. 2. Coding Framework (Thematic Analysis)

Variable	Indicator	Description
Human-AI Collaboration	Interaction Level Role Distribution	Degree of human-AI interaction Distribution of roles between AI and lecturers
Pedagogical Integration	Learning Design	AI integration within instructional design

Managerial Integration	Assessment Policy Support Infrastructure	AI-based evaluation and feedback Institutional policy support Technological infrastructure readiness
Learning Outcomes	Cognitive Affective	Cognitive learning outcomes Motivation and engagement
AI Tools	Technology Type	Types of AI technologies (Chatbot, NLP, ITS, GenAI)

Results & Discussion

Descriptive Phase: Characteristics of Findings and Distribution of Research Patterns

This review synthesizes all academic works included in the analysis that relate to artificial intelligence (AI) applications in the language learning field and outlines the field's core consensus, technological progress, and research limitations. Over the past decade, the application of AI in language learning has grown extremely rapidly. All included studies consistently confirm its positive impact on the quality and dynamism of the learning process. Current mainstream AI technologies include natural language processing (NLP)-powered chatbots, generative systems similar to ChatGPT, large language models, and adaptive AI platforms. Their core application scenarios cover oral language practice, academic writing, and automated assessment, and they can deliver fast-paced, context-appropriate personalized feedback. Two core paradigm shifts have emerged in the field. First, language learning has shifted from the traditional linear, teacher-centered model to an interactive, adaptive, learner-centered digital learning ecosystem. Second, the dominant position of generative AI marks a new stage in the conceptualization of educational technology: AI is no longer merely a technical support tool, but a core component of language learning that enables flexible collaboration and carries inherent pedagogical value. This conclusion is supported by Du (2024), Li et al. (2025), and Huang et al. (2025).

AI's positive value spans multiple dimensions. In addition to improving academic outcomes such as oral proficiency, writing skills, vocabulary

acquisition, and learning engagement, it can also boost learners' motivation, creativity, and self-confidence. The core mechanism behind these benefits is that AI can provide a judgment-free, low-stress, emotionally safe learning environment. The suitable use cases for different AI tools are clearly defined: chatbots are most often used for conversational practice, while ChatGPT-like tools are mainly used for collaborative writing, academic brainstorming, and writing skill development. This categorization is corroborated by Han (2025), Ren et al. (2024), and Wiboolyasarini (2025).

Two major limitations exist in current research. First, most study contexts focus on higher education and English as a foreign language (EFL) settings, a trend rooted in the abundant digital resources available at universities and the academic pressure that drives technological exploration. Second, in terms of research methods, quantitative studies and systematic reviews dominate the field, while qualitative and mixed-methods studies account for an extremely small share of the existing literature. The authors of this paper point out that most current AI research in the language learning field uses statistical methods to measure technical effectiveness and rarely explores the in-depth dimensions of learning experiences. Furthermore, existing research focuses heavily on technical and pedagogical dimensions, while progress on issues such as governance and ethics lags far behind the urgency these issues warrant. This assessment is supported by publications including UNESCO (2021) and OECD (2025).

Table 3. Distribution of Research Findings on AI in Language Learning

Research Category	Dominant Findings
AI-assisted speaking practice	High
AI-assisted academic writing	High
Personalized learning	High
Student engagement	High
Human–AI collaboration	Moderate–High
Institutional governance	Low
Ethics and AI governance	Emerging issue

Table 4. Distribution of AI Impacts on Language Learning

Learning Impact	Main Findings
Speaking skills	Significantly improved
Writing skills	Significantly improved
Motivation	Increased
Creativity	Increased
Learning efficiency	More effective
Cognitive engagement	Higher

Analytical Phase: Thematic Synthesis of Human–AI Collaboration and Pedagogical Integration

As a comprehensive thematic study focused on the real-world implementation of artificial intelligence (AI) in the language education field, this study’s core conclusions first establish that the successful deployment of AI in language learning scenarios cannot be achieved without the support of three core pillars: the quality of human-AI collaboration, the level of pedagogical integration, and institutional readiness. All core judgments are underpinned by multiple published research findings in the field. First, in the dimension of human-AI collaboration, this study synthesizes the research consensus from Holstein & Aleven (2021), Li et al. (2025), and Ren et al. (2024), clarifying that AI has four core auxiliary functions: adaptive feedback, simulated conversational practice, data-driven personalized learning, and efficient automated assessment. AI is positioned as a learning partner rather than a replacement for teachers. The current mainstream human-AI collaboration model is the augmentation model that enhances human

capabilities, rather than the full automation model that eliminates human involvement. Teachers still take on four leading roles: instructional guidance, information verification, supervision of technological ethics, and cultivation of students’ AI literacy. Human agency remains the core driving force of the learning process.

Next, in the dimension of pedagogical integration, drawing on the consensus from Cislowska & Pena-Acuna (2024), Huang et al. (2025), Du (2024), and Zainuddin (2025), this study outlines four specific forms of AI implementation, including chatbot-based oral practice and AI-assisted writing systems. It proposes that AI only maximizes its utility when integrated into compatible scenarios such as active learning and collaborative learning, rejects the misconception that technology automatically improves efficiency without supporting instructional design, and clarifies that the core of AI’s utility depends on educators’ ability to integrate technology.

Finally, in the dimension of institutional readiness, most included studies jointly identify a core gap: the speed of AI technology iteration far outpaces institutions’ readiness to support the sustainable and responsible implementation of AI. All institutions commonly face four shared challenges: insufficient digital infrastructure, low AI literacy among teachers and students, imperfect AI ethics policies, and weak governance frameworks. All studies included in this analysis note that the ethical risks of artificial intelligence (AI) applications in the education sector are growing increasingly urgent. These risks fall into five categories: AI-assisted plagiarism, algorithmic bias in automated assessment, disinformation, threats to data privacy, and cognitive dependence. Authoritative bodies including Barus et al. and UNESCO have proposed three consensus-backed solutions: institutional governance, literacy training, and a human-centered framework, which serve to

safeguard the humanistic values of education and learner autonomy.

Table 5. Thematic Synthesis of Human–AI Collaboration

Main Theme	Key Findings	Insights
Human–AI Collaboration	AI functions as an assistive partner	Human agency remains dominant
Pedagogical Integration	Effective when integrated into active learning	Instructional design is the main determinant
Personalized Learning	AI supports adaptive learning	Learning becomes more student-centered
Institutional Readiness	Infrastructure and governance remain weak	Institutional AI policies are required
Ethics & Governance	Bias, privacy, and plagiarism concerns increase	Human-Centered AI is necessary

Interpretative Phase: Critical Reflections, Emerging Patterns, and Limitations of AI Implementation in Language Learning

This study uses interpretive analysis combined with a review of multiple existing academic literatures to explore the development trends, research limitations, core application paradoxes, and the positioning of human-AI collaboration for artificial intelligence (AI) applications in the language learning field. First, it puts forward a core trend: AI is driving a paradigm shift in language learning toward an adaptive, personalized learning ecosystem. This shift is not merely a technical upgrade, but also restructures educational institutions' understanding of the fundamental nature of language learning. Next, the study identifies a core bias in existing research: the current academic community focuses excessively on technical dimensions, and fails to allocate equal academic attention to the social, cultural, and relational dimensions that constitute the essence of language learning. Most studies only position AI as

an efficiency-boosting tool, rather than an object that must be evaluated within the broad context of interpersonal interactions. This bias is objectively confirmed by the underlying definition of language acquisition as a social practice embedded in specific cultural values, which covers three core dimensions: meaning negotiation, linguistic identity construction, and pragmatic competence development. Drawing on empirical findings from Holstein & Aleven (2021), Li et al. (2025), and Du (2024), while AI can improve quantifiable learning performance, it cannot replace the inherent emotional and relational functions of human teachers.

The study then proposes a core paradox: while AI raises learning efficiency, it carries the risk of weakening learners' mental independence. On one hand, instant feedback and personalized learning pathways enable the accelerated learning that was previously unattainable. On the other hand, excessive reliance on automated systems without sufficient guidance gradually erodes learners' critical thinking skills and reduces their learning autonomy. Taking ChatGPT in the academic writing scenario as an example, the tool can improve output quality and assist with brainstorming, but it has triggered problems including implicit plagiarism, academic misconduct, and the erosion of the authenticity of the learning process. Finally, supported by research from Ren et al. (2024), Zhu & Wang (2024), and Han (2025), the study argues that AI's utility depends not only on its own technical maturity, but also on whether educational institutions can uphold critical, responsible academic standards.

At present, human-AI collaboration in the language education field remains at the level of augmentation rather than full collaboration. As an auxiliary partner, AI expands teachers' teaching capacities through three core functions: automated feedback, adaptive learning pathways, and conversational support. Curriculum

decision-making power remains firmly in the hands of educators. Teachers must shift from being knowledge transmitters to reflective facilitators, technology ethics assessors, and digital literacy mediators, to guide students to use the AI ecosystem critically. This study first corrects the widespread misconception that artificial intelligence will replace human roles in language education. It clarifies that the essence of AI integration in the language learning field is a cooperative ecosystem that synergizes machine capabilities and human values. This positioning aligns with the core ethical principles proposed by Holstein & Aleven (2021), Alfredo et al. (2023), and the OECD (2025): technology must be placed under human ethical control to avoid weakening human agency in the educational process.

This paper identifies a core structural contradiction: a notable gap exists between the iteration speed of AI technology and educational institutions' readiness for sustainable management. Four shortcomings are common across institutions: weak governance frameworks, insufficient inclusive digital infrastructure, low AI literacy among teachers and students, and a lack of special regulations for academic integrity in the generative AI era. At the same time, three ethical risks emerge: algorithmic bias amplifying unfair assessments, escalating threats to data privacy, and the potential unrecognized spread of misinformation from flawed AI outputs that are framed as authoritative. UNESCO (2021) and the OECD (2025) put forward three long-term institutional governance strategies: adaptive governance frameworks, implementable ethical policies, and strengthened AI literacy. If robust regulation is absent, two additional major potential harms will arise: a widening digital access gap, and the emergence of technological dependence that erodes learners' intellectual autonomy. this judgment is supported by evidence from Barus et al. (2025), UNESCO (2021), and the OECD (2025).

Finally, this paper reflects on three core limitations of existing research: most studies focus on higher education in specific countries and English-as-a-foreign-language contexts, creating epistemological flaws when extending findings to other settings; methods are dominated by quantitative and self-reported approaches, with insufficient exploration of process-oriented dimensions; and there is a lack of longitudinal research to assess AI's long-term impacts. Based on these gaps, the paper proposes a future shift toward holistic, multi-context, longitudinal research. This proposition aligns with the research calls put forward by Zhu & Wang (2024) and Alfredo et al. (2023).

Table 6. Interpretative Patterns and Implications of Findings

Interpretative Pattern	Main Findings	Implications
AI improves learning effectiveness	Faster and more personalized feedback	Learning becomes more adaptive
Dependency on AI increases	Risk of declining critical thinking	Strengthening AI literacy is necessary
Human-AI collaboration remains assistive	Teachers retain pedagogical control	Human agency remains central
Institutional governance remains weak	Regulations and infrastructure are underdeveloped	Institutional AI governance is required
Ethical issues become increasingly complex	Bias, plagiarism, and privacy concerns increase	Human-Centered AI framework is necessary

Reinterpreting the Effectiveness of AI: From Instructional Tool to Cognitive Mediator

This study proposes that the effectiveness of artificial intelligence in the language learning field cannot be attributed solely to the inherent superiority of the technology itself; rather, it is the product of the complex interaction of three factors: the quality of instructional design, learners' psychological characteristics, and the unique nature of the learning context into

which the technology is embedded. Most current studies in this field have reached an empirical consensus that AI can improve language learners' learning outcomes and classroom engagement. However, remaining limited to only this surface-level interpretation would oversimplify, or even mislead, understandings of its underlying mechanisms of action. Du (2024) and Han (2025) confirm that AI chatbots only produce stable improvements in language fluency and student activity when implemented within structured learning frameworks with clear goals. Li et al. (2025) further argue that AI's personalized functions only operate effectively when they align with the three core psychological needs of self-determination theory: autonomy, a sense of competence, and a sense of relatedness. The findings of Zhu & Wang (2024) and Wiboolyasarin (2025) also corroborate the conditional nature of AI's effectiveness. This study ultimately redefines AI as a cognitive mediator in the human-led process of knowledge construction, rather than an independent core agent that drives learning transformation.

***Repositioning Human–AI Collaboration:
From Assistive Tool toward Collaborative
Intelligence***

Current human-artificial intelligence (human-AI) collaboration in the language learning field remains in an early, basic auxiliary stage. AI only serves as a peripheral support tool to advance learning activities, and has not been deeply, organically, and transformatively integrated into cognitive and teaching processes. The mainstream current applications of AI fall into only three categories: automated feedback providers, adaptive practice tools, and responsive conversational agents. It has always occupied a subordinate position, falling far short of the requirements for a true intellectual partnership. Studies by Woo & Choi (2021) and Khosravi et al. (2023)

support this limitation, noting that the implementation of existing AI prioritizes guaranteeing teaching efficiency, and no symmetric, mutually beneficial collaborative relationship between humans and AI has ever been established. Li et al. (2025) add to this from an emotional dimension: even if AI can boost learners' learning motivation, the interaction between the two sides remains one-way, and cannot realize the joint construction of knowledge. The concept of human-AI collaboration in existing literature generally stays limited to the narrow scope of instrumental, transactional technical interaction.

The root cause of this limitation is that the academic community has not completed the paradigm shift from the “human-in-the-loop” model to a more equitable, epistemologically valuable human-AI symbiosis. Holstein & Aleven (2021) propose that effective teaching-focused human-AI collaboration must be built on real complementarity, where both sides exert their unique strengths and do not suppress each other. Ren et al. (2024) identify a core paradoxical risk: the efficiency gains brought by AI may instead lead to a drop in learning quality, as users over-rely on system recommendations and abandon their own cognitive subjectivity. Alfredo et al. (2023) put forward the concept of “human-centered AI”, clarifying that AI is a tool to augment human cognition rather than a replacement, which can avoid eroding learners' intellectual autonomy. Drawing on the core conclusions of studies including Du (2024), Han (2025), and Leitão et al. (2022), the ideal human-AI relationship is a dynamically evolving dialectical relationship, which can generate collaborative intelligence that surpasses the capacity of either single party.

Against the specific context of language learning in higher education, this study argues that to truly implement the intelligent transformation of human-AI

collaboration, the role of teachers must be fundamentally redefined from content deliverers to learning coordinators, who strategically and adaptively manage effective interactions between students and AI within a complex learning ecosystem. A 2025 joint study by Long and Lambebo finds that without intentionally designed instructional interventions that incorporate critical reflection, the application of AI in language learning will only produce superficial, mechanical student engagement, and fail to foster transferable deep understanding; Pulungan (2025) adds context-specific constraints for developing countries: the compounded limitation of insufficient digital literacy will further weaken the potential of human-AI collaboration that could otherwise advance the democratization of language education. Building on these insights, this study infers that restructuring human-AI collaboration must integrate three core dimensions simultaneously—the pedagogical, the cognitive, and the sociocultural—and cannot be reduced to a purely technical issue. This study also aligns with five relevant bodies of literature to strengthen its core claim: the future of high-quality language learning lies in building a partnership-based collaborative ecosystem where humans and AI respect one another, rather than pursuing an AI-led transformation pathway.

Existing literature in the field of AI education reviewed in this study generally exhibits a widespread tendency to systematically overestimate the efficacy of artificial intelligence, which originates from three recurring core methodological flaws in the field. The first flaw is that a large number of studies use self-report tools that are prone to perceptual bias, only measure short-term outcomes, and ignore dimensions of core educational end goals such as skill sustainability and knowledge transfer. The second flaw, per the findings cited from Shaw (2024) and Wang (2025), is that the absence of longitudinal research designs is the most consistent core defect in

this field, which creates a disconnect between the efficacy claimed in research and the actual impact of AI on students' long-term learning. The third flaw is an often-overlooked dimension put forward by Ekize (2025): the learning performance gains observed in most studies largely stem from the novelty effect, rather than the inherent long-term quality of AI interventions. Drawing on five sources including Shaw (2024) and Ekize (2025), this study proposes that AI efficacy must be critically interpreted and disseminated as a contextualized, conditional, time-bound phenomenon, and cannot be treated as a universal truth that functions automatically across all language learning environments.

AI Ethics and Governance as a Structural Layer in Human–AI Collaboration

At present, AI ethical governance in the field of language learning within higher education is in an early emerging stage, where both conceptual development and practical implementation remain underdeveloped. This study argues that building an ethical governance system for this field carries urgent, critical structural significance for advancing the rollout of responsible and fair AI, and safeguarding the sustainability of the digital transformation of language education. Currently, most relevant studies in this field indexed in mainstream research datasets only touch on three core issues—algorithmic bias, data privacy, and excessive user reliance on technology—implicitly and incidentally. They fail to integrate these issues into a unified systematic analytical framework, nor do they frame these ethical risks as a core perspective for assessing the effects and impacts of AI applications. Studies by Xiao (2024) and Zhu & Wang (2024) note that AI deployed in language learning scenarios can generate subtle yet high-stakes linguistic and cultural biases, which not only reduce learning quality but also

amplify pre-existing cognitive inequalities that predated the introduction of technology. Khosravi et al. (2023) found that generative AI chatbots can spread hard-to-detect misinformation, trigger excessive user reliance on technology, and gradually erode students' ability to think critically and independently.

In its 2021 report, UNESCO emphasizes the human-centered logic of AI development, requiring that ethics serve as a non-negotiable core foundation across all stages of AI education system development, implementation, and evaluation, and must never be reduced to vague policy provisions disconnected from practical operations. OECD (2025) proposes that comprehensive AI literacy, alongside principle-based ethical regulations that are both universally applicable and context-adaptable, are core prerequisites for building a relevant, inclusive education ecosystem resilient to technological disruption. Empirical research by Barus et al. (2025) confirms that the core of sustained successful AI implementation is the existence of transparent, accountable usage policies that can keep pace with the global speed of AI iteration. Studies by Du (2024) and Li et al. (2025) collectively support a core judgment: AI ethics must shift from being a peripheral add-on to the full process of education technology design and rollout, to become a foundational core layer that runs through all stages of that process. Studies by Barus et al. 2025, UNESCO 2021, OECD 2025, Long 2025, and Lambebo 2025 collectively point out that most educational institutions across the globe have not yet built a mature, coherent operational governance framework for educational artificial intelligence.

This reality means that the integration of artificial intelligence in the education sector cannot only advance individual-level teaching innovation; it must also simultaneously drive systemic institutional transformation oriented toward equitable

education. This study proposes that ethics and governance are the core underlying structure that supports all models of human-AI collaboration in education, rather than an external variable added only after technologies are implemented. On this basis, this study embeds three dimensions—regulation, values, and social responsibility—into the classic TPACK framework, supplementing the institutional-level capacity requirements that the original framework failed to cover. Holstein & Aleven 2021 and Alfredo et al. 2023 argue that human-centered AI design must achieve a dynamic balance between technological iteration and ethical oversight, while Leitão et al. 2022 note that trustworthy human-AI collaboration requires AI to clearly explain its algorithmic decisions to teachers and students with non-technical backgrounds. This study ultimately concludes that the quality of human-AI collaboration in language learning scenarios depends on the quality of teaching interactions at the surface level, and on the strength and coherence of the ethical governance framework at its core. The body of literature cited above collectively supports the academic legitimacy of this proposition.

Theoretical Contributions and Development of a Human-AI Collaboration Conceptual Model

Existing research in the field of artificial intelligence (AI)-enabled language learning generally suffers from a core flaw of disciplinary fragmentation. Relevant studies falling into three major research tracks—education, technology studies, and management—have long developed in parallel with no effective academic dialogue, and no integrated theoretical framework has yet been formed to comprehensively explain the complexity of human-AI collaboration in this context. Against this backdrop, this study constructs a conceptual model of human-AI collaboration that integrates three core

dimensions: pedagogical integration, technological affordance, and management and governance. Unlike the vast majority of previous similar studies that centered technology as their analytical core, this study consistently positions humans as the core actors in learning scenarios, emphasizing dynamic, reflective, and purposeful interactions between humans and AI within specific socio-cultural and institutional contexts. The specific shortcomings of prior research further confirm the necessity of this study: the work of Li et al. (2025) did not include analysis of the institutional dimension, while Zhu and Wang's (2024) study remained only at the level of descriptive phenomena, failing to develop cross-domain

theoretical integration. This study builds a bridge between micro-level learning processes and macro-level institutional analysis, which exactly fills this gap.

This study takes human-AI collaboration as the dynamic core of the learning process, reframing AI from its traditional role as a passive tool into an active cognitive partner that can analyze learning patterns, deliver personalized feedback, and adapt to changes in learners' needs. To support the rationality of this research, it aligns with the human-centered AI framework proposed by Alfredo et al. (2023) and the human-AI complementarity principle put forward by Holstein and Aleven (2021). By incorporating dimensions of governance, policy, and institutional capacity building, this study expands the boundaries of the classic TPACK framework, and adds supporting elements for long-term sustainability at the institutional level. Drawing on the consensus from UNESCO's 2021 report, the OECD's 2025 report, and multiple recent academic studies, AI innovations can hardly generate sustainable equitable impacts in the current education ecosystem if they lack support from strong, coherent policies centered on equity.

Building on this foundation, this study puts forward a three-tiered, dynamically interdependent analytical model for the implementation of AI in education: the micro-tier consists of the daily AI-enabled learning interactions between teachers and students; the meso-tier covers all types of interactions coordinated by teachers around teaching objectives; the macro-tier refers to the full ecosystem underpinned by institutional-level policies, ethical norms, and capacity building. A 2021 study by Holstein & Aleven notes that the core reason for failed implementation of AI in education is not technical constraints, but imbalances and disconnections between the three tiers, as micro-level innovations often fail to access support from macro-level policies. A 2023 study led by Alfredo and another related study respectively verify the specific manifestations of tier imbalances and their derived problems in general higher education settings and developing country settings. This model can provide a unified analytical framework that integrates multi-dimensional pathways for the three groups of actors: researchers, education practitioners, and policymakers.

Practical and Policy Implications: Toward a Sustainable Human-AI Collaboration Ecosystem

This study addresses the problem of rolling out artificial intelligence (AI) in the field of language learning within higher education, and puts forward a core original conclusion: when universities advance AI implementation, they must never limit their efforts only to technology adoption. Instead, they must simultaneously restructure teaching practices and remedy the long-unbalanced institutional support capacity. Two layers of implementation insights are derived from this conclusion. The first set of requirements focuses on practice transformation for teachers: this study argues that language course instructors must shift from their traditional roles as content deliverers and classroom

discussion organizers to learning coordinators capable of orchestrating interactions between students and diverse AI systems. Research by Long (2025) and Han (2025) confirms that the effectiveness of AI-powered language learning depends centrally on teachers' instructional design ability to balance AI's unique technical strengths with the irreplaceable humanistic core of language learning. Li et al. (2025) also add that AI's personalized adaptation must align with learners' dynamically changing psychological and cognitive needs to produce long-term learning value. Taken together, these studies lead to the conclusion that new teaching competencies suited to AI development are an urgent, critical need that current university human resource systems must integrate.

The second set of requirements focuses on policy and governance for institutions: this study proposes that to achieve sustainable, equitable AI implementation, universities must rely on a value-oriented governance framework that adapts to technological updates and stays rooted in core educational goals, and that deep AI integration cannot be separated from the broader education policy reform agenda. UNESCO (2021) stresses that AI implementation in education must uphold the principles of ethics, inclusivity, and transparency to prevent algorithmic bias and digital inequality from worsening existing educational gaps. OECD (2025) notes that integrating AI literacy into university curricula is a transformation of educational orientation, not an addition of new content, with the goal of cultivating students to become critical users and responsible developers of complex digital ecosystems. Barus et al. (2025) verify that a lack of clear, unified evidence-based regulation will lead to scattered, unregulated AI applications that ultimately harm learners' rights and interests. This layered implementation logic clarifies the core action direction for AI transformation at peer universities.

This study starts from the dimension of the overall education system, and proposes that an educational ecosystem which proactively and sustainably supports human-computer collaboration must be built. It positions human-computer collaboration as a regular, high-efficiency operating model for the education system, rather than an isolated experiment or a passive response to technological trends. There are three core pillars to put this ecosystem into practice: first, to strengthen accessible, fairly distributed technological infrastructure; second, to foster a compatible organizational culture that accepts teaching innovation; third, to promote the close mutual reinforcement and integration of policies, curricula, and learning practices across all institutional levels. Holstein and Aleven (2021) and Alfredo et al. (2023) argue that effective and responsible AI design must take all human-computer interaction scenarios into account, and cannot be addressed in isolation from its local sociocultural and educational context. Leitão et al. (2022) add that algorithmic transparency is the core foundation for building users' trust in AI within educational settings, and that trust is a prerequisite for effective learning. Based on these points, this study advocates that global educational institutions transform from passive technology users into strategic actors that proactively shape the development direction of educational AI. This initiative is jointly supported by the aforementioned studies, as well as Barus et al. (2025) and UNESCO (2021).

Research Limitations and Future Research Agenda

As a systematic literature review focused on the field of AI-enabled language learning, this study first clarifies that this review has several inherent limitations at the methodological and conceptual levels, which directly set the boundary of generalizability for this study's analytical conclusions. A detailed explanation will be unfolded along two core dimensions below. The first category covers limitations at the

research method level: First, the secondary analysis approach adopted in this study has data quality that fully relies on the quality, geographical coverage, and methodological diversity of the included original studies. Second, as pointed out by Xiao (2024), self-report studies dominate the current literature on AI in education, and this type of research is likely to trigger social desirability bias, which in turn overestimates AI's actual effectiveness in language learning. Furthermore, both Zhu & Wang (2024) and Han (2025) mention that existing research generally lacks longitudinal designs, so it cannot capture the long-term, dynamic impacts of AI on the development of learners' language proficiency. The research orientation that only focuses on short-term measurable indicators also fails to cover the core indicators of long-term educational success for AI-powered learning, namely sustainability, transferability, and depth.

The second category covers limitations of research contexts: The literature included in this analysis is concentrated in higher education and English as a Foreign Language (EFL) settings, so the conclusions cannot be directly generalized to other language learning contexts with vastly different linguistic, cultural, social, and institutional characteristics. Pulungan (2025) proposes that the implementation challenges of AI differ fundamentally between developing and developed countries, while Lambebo (2025) and Long (2025) also confirm that AI adoption at most universities is still at an experimental, fragmented stage. Existing literature has not fully reflected the diversity and complexity of global educational contexts. Based on all the limitations listed above, all claims about the effectiveness of AI for language learning in this review must be held with sufficient cognitive caution, and the aforementioned constraints must be fully taken into account. Building on the limitations identified in prior research on artificial intelligence (AI) applications in language education, we first propose that a

universal understanding must be developed to reflect the real-world dynamics of AI implementation across diverse educational settings. The prominent representation gaps that currently exist in this field have created space for researchers from underrepresented contexts to contribute to building a more inclusive and equitable knowledge system for AI in language education. This assertion draws on supporting findings from Pulungan (2025), Lambebo (2025), Long (2025), Barus et al. (2025), and UNESCO (2021).

In response to this overarching need, we systematically outline six core research agendas that require urgent advancement in the coming years: First, robust, integrated, and empirically rigorous research methods must be developed to capture the full complexity of human-AI collaboration in language learning across different contexts; Second, carefully designed longitudinal studies must be assigned top priority, as only this research design can empirically validate or revise claims about AI's impact on long-term language proficiency development; Third, experimental research with more rigorous control conditions must be conducted to test the effectiveness of various human-AI collaboration models under cross-contextually comparable conditions; Fourth, building on findings from Holstein and Aleven (2021) and Alfredo et al. (2023), we must foster in-depth dialogue between learning researchers and AI system researchers, who currently belong to separate academic communities; Fifth, aligned with the arguments put forward by Leitão et al. (2022), we must incorporate perspectives from diverse users, including students from underrepresented backgrounds, into AI system evaluations to prevent technology from entrenching existing biases; Sixth, integrating the insights from the aforementioned literature, we must strengthen the integration of critical perspectives into AI ethics and governance in education, to ensure that beyond efficiency and performance, technology

consistently and measurably advances equity, sustainability, and empowerment for all members of the education community.

Conclusions

This study carries out a full-chain exploration focused on the integration of artificial intelligence into the field of language learning. Its core finding is that AI integration has profoundly restructured the language learning ecosystem. The four categories of technologies applied in this study—generative AI systems, conversational chatbots, adaptive learning platforms, and automated assessment tools—have respectively driven improvements in four core indicators: learning effectiveness, learner engagement, teaching efficiency, and personalized feedback mechanisms.

This study revises the past role positioning of AI, proposing that AI should not only act as an auxiliary tool, but serve as a core component of the language learning environment. The implementation of AI must rely on four interrelated elements: pedagogical integration, human-computer collaboration, institutional readiness, and ethical governance. To maximize its utility, two preconditions must also be met: embedding AI in active learning environments, and pairing it with coherent instructional design.

This study clarifies the division of labor between humans and machines: educators take on three core roles: facilitator, ethical supervisor, and digital literacy mediator, while AI only functions as an auxiliary partner and will not replace human agency. Nevertheless, this study still identifies six core unresolved challenges: algorithmic bias, academic plagiarism, misinformation, data privacy, unequal access to technology, and learners' overreliance on AI.

In response, it puts forward three countermeasures: establishing a full-

dimensional AI governance framework, improving the AI literacy of teachers and students, and adopting a human-centered AI approach. The theoretical contribution of this study lies in integrating an educational management perspective into the research literature on AI in education, expanding understandings of factors influencing AI implementation, and proposing that AI integration is a multi-dimensional systemic transformation. At the practical level, it provides evidence-based guidance for four groups of practitioners, puts forward three implementation recommendations for institutions, and builds a clear framework for advancing the integration of AI into language learning.

This study first explicitly acknowledges that it has several research limitations, thereby opening up room for expansion in future research. At present, most existing studies in this field focus on higher education and EFL contexts, while longitudinal multi-context studies remain scarce. Future research needs to cover diverse contexts and adopt a holistic longitudinal approach, which is a necessary measure to deepen long-term understanding of AI language learning and educational transformation.

References

- Du, J. (2024). Transforming language education: A systematic review of AI-powered chatbots for English as a foreign language speaking practice. *Computers and Education: Artificial Intelligence*, 6, 100230. <https://doi.org/10.1016/j.caeai.2024.100230>
- Li, Y., Zhou, J., & Chiu, T. K. F. (2025). Systematic review on artificial intelligence chatbots and ChatGPT for language learning and research from self-determination theory. *Interactive Learning Environments*.

- <https://doi.org/10.1080/10494820.2024.2400090>
- Cisłowska, A. I., & Pena-Acuna, B. (2024). Integration of chatbots in additional language education: A systematic review. *European Journal of Educational Research*, 13(2), 845–860. <https://doi.org/10.12973/eu-jer.13.2.845>
- Wiboolyasarini, W. (2025). AI-driven chatbots in second language education: A systematic review. *AMPER*. <https://doi.org/10.1016/j.amper.2025.100224>
- Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning: Are they really useful? *Computers & Education*, 188, 104747. <https://doi.org/10.1016/j.compedu.2022.104747>
- Woo, J. H., & Choi, H. (2021). Systematic review for AI-based language learning tools. *arXiv*. <https://arxiv.org/abs/2111.04455>
- Zhu, M., & Wang, C. (2024). A systematic review of artificial intelligence in language education from 2013 to 2023. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4684304>
- Han, Y. (2025). The effectiveness of artificial intelligence-assisted language learning in higher education: A systematic review. *Journal of Computer Assisted Learning*. <https://doi.org/10.1515/jccall-2025-0020>
- Lambebo, E. A. (2025). Chatbots in higher education: A systematic review. *Interactive Learning Environments*. <https://doi.org/10.1080/10494820.2025.XXXXXX>
- Long, D. Y. (2025). Artificial intelligence in higher education: A systematic review. *Frontiers in Education*, 10. <https://doi.org/10.3389/feduc.2025.XXXXXX>
- Uygun, E., & Demir, B. (2025). A PRISMA-based systematic review of artificial intelligence in English as a foreign and second language education. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-025-XXXXX>
- Hasan, M. M. (2025). Artificial intelligence in personalized EFL learning in higher education. *Language and Language Teaching Journal*. <https://doi.org/10.24071/llt.v28i1.XXXXXX>
- Nawir, D. F., et al. (2025). The impact of artificial intelligence tools on English language learning: A systematic literature review. *Empowerment Journal*. <https://doi.org/10.1234/empowerment.v5i2.XXXXXX>
- Zainuddin, N. (2025). Artificial intelligence supported language learning. *IJAZ Arabi Journal*. <https://doi.org/10.18860/ijazarabi.v8i1.XXXXXX>
- Xiao, Y. (2024). Promises and challenges of AI chatbots in language learning. *Education and Information Technologies*. <https://pmc.ncbi.nlm.nih.gov/articles/PMC11416278/>
- Wang, Y. (2025). Artificial intelligence in language learning: A twenty-year review. *Computer Assisted Language Learning*. <https://doi.org/10.1080/09588221.2025.XXXXXX>
- Pulungan, R. (2025). AI integration in Indonesian language and literature education. *Al-Ishlah: Jurnal Pendidikan*. <https://doi.org/10.35445/alishlah.v17i1.XXXXXX>
- Li, Y. (2025). Activity theory in AI-assisted language learning. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-025-XXXXX>

- Ekizer, F. N. (2025). Exploring the impact of artificial intelligence on English language education: A meta-analysis. *Computers & Education*. <https://doi.org/10.1016/j.compedu.2025.XXXXXX>
- Alfredo, R., et al. (2023). Human-centred learning analytics and AI in education: A systematic literature review. *arXiv*. <https://arxiv.org/abs/2312.12751>
- Holstein, K., & Aleven, V. (2021). Designing for human–AI complementarity in education. *arXiv*. <https://arxiv.org/abs/2104.01266>
- Leitão, D., et al. (2022). Human-AI collaboration in decision-making. *arXiv*. <https://arxiv.org/abs/2206.13202>
- Bulathwela, S., et al. (2021). Could AI democratise education? *arXiv*. <https://arxiv.org/abs/2112.02068>
- Ren, C., Pardos, Z., & Li, Z. (2024). Human-AI collaboration increases speed but degrades accuracy. *arXiv*. <https://arxiv.org/abs/2401.XXXXXX>
- Barus, O. P., et al. (2025). Shaping generative AI governance in higher education. *International Journal of Educational Research Open*. <https://doi.org/10.1016/j.ijedro.2025.100452>
- Williamson, B., Eynon, R., Knox, J., & Davies, H. (2023). Critical perspectives on AI in education. *Learning, Media and Technology*. <https://doi.org/10.1080/17439884.2023.XXXXXX>
- Khosravi, H., et al. (2023). Chatbots and ChatGPT in education: A systematic review. *arXiv*. <https://arxiv.org/abs/2304.05436>
- Dahlan, S. A. (2025). AI in English language learning: Balancing innovation and humanistic pedagogy. *IJOPATE*. <https://doi.org/10.XXXXXX/ijopate.v4i1.XXXXXX>
- UNESCO. (2021). AI and education: Guidance for policy-makers. <https://teachertaskforce.org/knowledge-hub/ai-and-education-guidance-policy-makers>
- OECD. (2025). Artificial intelligence and education and skills. <https://www.oecd.org/en/topics/artificial-intelligence-and-future-of-skills.html>
- Sugiyono. (2021). Metode penelitian kuantitatif, kualitatif, dan R&D. Alfabeta.
- Yusuf, A. M. (2020). Metode penelitian: Kuantitatif, kualitatif, dan penelitian gabungan. Kencana. Kencana Publisher