Navigating the AI Turn: Framework for Responsible AI Integration in Language Education

Pham Vu Phi Ho[1], Willy Ardian Renandya[1*]

[1] Faculty of Foreign Languages, Industrial University of Ho Chi Minh City, Vietnam
[*]Corresponding author's email: willy.renandya@nie.edu.sg
[*] https://orcid.org/0000-0002-1183-0267
https://doi.org/10.54855/ijte.26626

Received: 20/04/2026 Revision: 25/05/2026 Accepted: 26/05/2026 Online: 27/05/2026

ABSTRACT

As generative artificial intelligence (GenAI) revolutionizes the teaching and learning of language, the discipline faces two challenges: to avoid the traps of superficial technological replacement and to proactively defend academic integrity and learner autonomy. This paper addresses this conflict by asking two important questions: (1) pedagogical conditions that favor the development of profound language skills in AI rather than superficial language skills; (2) organizations can promote responsible use while cultivating critical thinking. To address these questions, this paper proposes the Framework for Responsible AI Integration in Language Education (FRAILE). This unifying model incorporates recent empirical research to offer a comprehensive structural response to the AI disruption. This presentation introduces a new AI Task Typology to enable cognitive engagement within the FRAILE.

Keywords: GenAI, academic integrity, critical thinking, teacher development, AI literacy

Introduction

The integration of technology into language learning is not new. The first computer-assisted language learning (CALL) systems emerged in the 1960s in university laboratories (Li et al., 2025). Over the last few decades, technology has evolved dramatically from simple drill-and-practice software to complex networked platforms and intelligent tutoring systems that provide highly tailored, adaptive instruction (Hariyanto et al., 2025). However, artificial intelligence has quickly and profoundly revolutionized language education, particularly since the advent of ChatGPT in November 2022. There has been no other technology developed of such consequence.

The rapid rise of ChatGPT has transformed generative AI from a distant possibility into an immediate reality in language education. It is the fastest-growing online application in history, with over 1 million users in its first week and 100 million in its first two months (Reuters, 2023). Artificial intelligence (AI) was rolled out in language classrooms around the world at an unprecedented pace, not as a top-down institutional initiative, but as a grassroots, student-led movement. Instead of looking forward to a potential future, teachers were examining how

CITATION | Pham, V. P. H., & Renandya, W. A. (2026). Navigating the AI Turn: Framework for Responsible AI Integration in Language Education. *International Journal of TESOL & Education*, *6*(2), 70-85. DOI: https://doi.org/10.54855/ijte.26626

generative AI tools were affecting student writing, translation, grammar checking, vocabulary acquisition, and exam preparation, regardless of the school's approval of their use.

Wang et al. (2025) demonstrate that generative AI (GENAI) has emerged as a potent force in language education, with substantial potential for language learning, instruction, assessment, and research. The rapid expansion of research in this field is illustrated by a comprehensive review of 43 empirical studies on GENAI that were published in SSCI-indexed journals from 2022 to 2024. The review places a particular emphasis on the motivational, speaking, and writing outcomes of higher education learners in the EFL context. However, the existing literature collectively suggests that the field still lacks cohesive frameworks to support practitioners, curriculum developers, and institutional policymakers in responsibly integrating AI (Wang et al., 2025; Li et al., 2025).

This paper addresses this lacuna by engaging in a progressive inquiry that encompasses two interrelated research questions: the ethical and pedagogical challenges, the necessary responses for professional development, and the contributions of AI to language learners..

Research Questions

RQ1: To what extent and under what pedagogical conditions does the integration of AI tools really advance language skill development beyond surface-level performance outcomes in EFL/ESL contexts?

RQ2: How can language educators and institutions encourage responsible AI use without sacrificing the genuine pedagogical benefits of AI, such as the development of learner critical thinking, authentic voice, and academic integrity?

Methodological Orientation: Critical Integrative Synthesis

To address the research questions and develop the Framework for Responsible AI Integration in Language Education (FRAILE), the paper critically and integratively synthesizes current scholarship on generative artificial intelligence in language education. This is a strength in a fast-moving field, as this approach does not just attempt to summarize findings but rather seeks to identify patterns, tensions, and voids in new studies and translate them into a direct response from instructors and schools (Torraco, 2016). The synthesis is not intended as a formal systematic review, but rather to aid in the development of a framework and conceptual clarity. The objective is to explore the literature on the educational promise of AI, the threats to academic integrity and critical thinking, and the classroom conditions that make it truly beneficial.

Scope and Selection of Literature

The literature reviewed in this paper focused on the pedagogical benefits and ethical and pedagogical concerns of generative AI in language education. The evaluation was intentionally restricted to three factors. Initially, it focused on research published from the end of 2022 to the beginning of 2026, such as the post-public-release period for ChatGPT and the rise of educational research on GenAI. Secondly, the quality of the sources was of tremendous importance. Empirical studies, systematic reviews, and important conceptual contributions published in high-quality peer-reviewed academic journals (e.g., Elsevier, Springer Nature, Taylor & Francis, Wiley-Blackwell, and SAGE) were given priority. This ensured that the discussion was based on academically rigorous, reliable, and up-to-date research. This ensured that the discussion was based on reliable and up-to-date research. Third, the review has largely been limited to the EFL and ESL contexts, especially at the secondary and higher education

levels, because these are the most relevant to the issues discussed in this paper and form the major evidence base.

Analytical Logic of the Synthesis

The synthesis was carried out in three interconnected phases. In the first phase, patterns in the literature were found. Two main ideas kept coming up. AI was shown to improve pupil engagement, feedback, grammatical accuracy, and writing fluency. However, the issues discussed in the literature are rather consistent: threat to academic integrity, loss of authentic voice, cognitive outsourcing, reduction of critical engagement, and over-reliance.

The next stage was to interpret these patterns in the light of existing theories of education. Vygotsky's Zone of Proximal Development helped us understand the difference between AI as a tool to help us (Vygotsky, 1978) and AI as a replacement for human thinking. Sweller's Cognitive Load Theory provides a framework for understanding how AI can facilitate learning while also removing the need for cognitive effort (Sweller, 1988). The SAMR model was then applied to distinguish simple AI use that merely substitutes for the learner's work from more advanced uses that enable deep revision, critique, and transformation (Puentedura, 2006).

The third phase focused on transforming these empirical and theoretical insights into a pragmatic framework for language education. This has led to the development of the AI Task Typology, which classifies AI-supported activities according to the degree of cognitive engagement involved, and the more extensive FRAILE framework, which encompasses classroom practice, assessment design, teacher development, and institutional accountability. In the final phase, the UNESCO AI Competency Framework for Teachers enabled the integration of pedagogical approaches with institutional capacity building (UNESCO, 2024).

Purpose of This Approach

This necessary integration is not meant to be a complete review of all existing research. Instead, it seeks to synthesize the most relevant recent evidence that is both pedagogically useful and theoretically informed. The benefit of this method is that it links empirical evidence to practical educational decision-making, thus offering language educators and institutions a more solid foundation for responsible practice in the age of AI (Torraco, 2016; UNESCO, 2024).

Review of Literature

This section reviews the literature on two related themes: (1) the opportunities and challenges of AI for the development of language skills, and (2) the threats to academic integrity, critical thinking, and learner autonomy. The reviewed studies show a common pattern in fast-moving fields of research, namely a descriptive accumulation of research, mostly in EFL higher education contexts in East Asia, with most of the studies based on self-reported data from small samples (Wang et al., 2025; Li et al., 2025; Liu et al., 2025). This paper is based on the convergence of the inequalities.

Theme 1: AI for Language Skill Development — Promise and Limitations

An increasing corpus of empirical studies validates that AI tools—when utilized intentionally and under the guidance of instructors—can significantly enhance language skill development, especially in writing. Nguyen and Pham (2025) examined the use of ChatGPT in an IELTS writing course in Can Tho, Vietnam. Employing a quasi-experimental pre-test/post-test design with 32 students, the experimental group significantly outperformed the control group across all IELTS writing components. Mahapatra (2024), in a mixed-methods intervention study at an

Indian university, similarly reported significant positive effects of ChatGPT as a formative feedback tool on the academic writing skills of undergraduate ESL students. Polakova and Ivenz (2024) found improvements in writing conciseness and passive-voice use following ChatGPT-mediated feedback sessions involving 110 EFL students. Asadi et al. (2025), in Thinking Skills and Creativity (Elsevier), documented that integrating ChatGPT with teacher feedback yielded favorable results for EFL writing skills, especially when the feedback modalities were intentionally sequenced.

Yuan and Liu (2025), in Computers in Human Behavior (Elsevier), found that Chinese EFL learners using AI tools exhibited markedly greater engagement, enjoyment, and motivation than their non-AI counterparts. Pham and Huynh (2025) reported substantial improvements in motivation, goal-setting, and self-regulated learning behaviors among Vietnamese high school students who used AI tools. Lo et al. (2024), in a systematic review of 70 ChatGPT studies in Smart Learning Environments (Springer Nature), found that 41.4% of the studies focused on writing and only 7.1% on speaking, with no studies specifically addressing listening, indicating a concerning pedagogical constraint.

A major gap remains: The focus on surface-level writing parts does not examine how AI can support deep argumentative reasoning, disciplinary voice, or long-term writing development. In a review of AI-powered writing instruction in the ECNU Review of Education (SAGE), Xiao et al. (2025) found that while ChatGPT improved sentence-level quality and structural clarity, its capacity for fostering original, culturally embedded argumentation is limited and understudied.

Theme 2: Over-Reliance, Academic Integrity, and Critical Thinking Erosion

While evidence for AI's instructional benefits is growing, a concurrent body of research underscores the risks posed by uncritical use of AI to learner autonomy, intellectual development, and academic integrity. Werdiningsih et al. (2024), in a qualitative case study conducted at an Indonesian university and published in Cogent Arts & Humanities (Taylor & Francis), found that ChatGPT was appreciated for alleviating writing uncertainties and clarifying vocabulary; however, it also elicited apprehensions about the authenticity of student work. AI suggestions were sometimes too complex or insensitive to different cultures, and students knew that passing off AI-generated content as their own voice was risky.

Khan et al. (2025), in a qualitative study of 43 Indian EFL learners published in Cogent Education (Taylor & Francis), identified substantial issues related to excessive dependence, the decline of critical thinking, and risks to academic integrity. Students said that AI-generated answers didn't always fit the way Indians normally write. Gerlich (2025), in Societies (MDPI), documented that excessive reliance on AI tools results in "cognitive offloading," a phenomenon characterized by a significant decline in students' analytical reasoning abilities. Critical thinking was found to be the primary moderator between AI use and over-reliance, as students with higher levels of critical thinking used AI outputs more judiciously and did not accept them uncritically (Hou et al., 2025).

Wang et al. (2025) conducted a scoping review of 43 empirical studies published in SSCI journals between 2022 and 2024 and found that the main issues identified in the literature were overreliance, academic integrity, and critical thinking, with 46.5% of studies lacking well-defined theoretical frameworks. Nazim and Alzubi (2025) in PLOS ONE reported that 278 EFL teachers in Saudi Arabia considered institutional policies and ethical frameworks important but lacking safeguards.

Research gaps

The literature reviewed under the above themes describes a domain of rapid empirical accumulation and considerable structural disintegration. There are two main gaps that directly shape the reasoning of this paper.

First, the documented Surface-Depth Paradox in language acquisition assisted by AI. Current empirical evidence indicates that generative AI tools provide consistent, statistically significant benefits in surface-level language performance, particularly in grammar accuracy, writing fluency, and vocabulary range (Nguyen & Pham, 2025; Mahapatra, 2024; Polakova & Ivenz, 2024). But the field lacks a systematic understanding of the pedagogical conditions necessary to foster deeper cognitive engagement. Research reveals that attention remains focused on quantifiable, discrete elements of writing, and the role of AI in supporting original, culturally relevant argumentation and higher-order reasoning remains poorly understood (Xiao et al., 2025; Li et al., 2025). This gap calls for scrutiny of the carefully designed conditions that raise AI from a mere expedient to a transformative educational paradigm.

Second, there is an important safeguard in academic integrity and critical thinking. Institutional responses are often defensive and reactive, although the risks of "cognitive offloading" and identity erosion in AI-driven writing are well documented (Gerlich, 2025; Khan et al., 2025). Students are using AI extensively, with as many as 86% using it in some contexts, but few are aware of AI policies at their institutions (Digital Education Council, 2024). Moreover, teachers often mention the lack of clear ethical guidelines and institutional safeguards as a major challenge for responsible implementation (Nazim & Alzubi, 2025). This gap of high usage and low awareness of the policies calls for a shift from "AI policing" to "AI aware" pedagogical and assessment design.

Take all these holes, and you get that the value of AI as a tool for teaching languages is not in the technology itself. Rather, it is created or compromised by the quality of human decisions over task design and ethical oversight.

RQ1: AI-Enhanced Language Learning — Affordances, Conditions, and the Surface-Depth Problem

What AI Consistently Delivers

In a variety of EFL and ESL contexts, AI tools have demonstrated consistent, reproducible benefits across quantifiable aspects of language performance. Of particular importance is the sequencing finding of Asadi et al. (2025) that AI feedback is not equivalent to teacher feedback but can be improved with careful planning. This leads to an important pedagogical point: AI should be a spur to reflection, not the ultimate arbiter of knowledge.

Li et al. (2025), in a comprehensive scoping review of 144 peer-reviewed articles from the Web of Science, Scopus, and ERIC databases published in Computers & Education: Artificial Intelligence (Elsevier), show that the evidence base in this area is mainly short-term, single-institution studies. This concentration is not only a limitation on research but also a worrying pedagogical constraint, in which the superficial advantages of AI in writing are equated with greater communicative language proficiency.

The Theoretical Lens: Scaffolding Versus Substitution

You need to work through three theoretical frameworks to understand why AI is so much better at making shallow progress than deep learning progress. Vygotsky's (1978) Zone of Proximal

Development (ZPD) suggests that effective learning occurs when learners receive mediated assistance on tasks slightly beyond their independent abilities. AI at its best is a dynamic scaffold. In a systematic review published in Smart Learning Environments (Springer Nature), Lee et al. (2025) have confirmed that generative AI can provide structured support for writing, speaking, and reasoning, helping learners refine ideas and evaluate arguments. On the other hand, when AI is used to do the cognitive work for the learner, rather than support the learner, it circumvents the Zone of Proximal Development (ZPD) and creates cognitive overload: measurable output without productive struggle (Gerlich, 2025).

This analysis is supported by Sweller's (1988) Cognitive Load Theory (CLT). "When it produces content, selects words, and organizes arguments simultaneously, AI eliminates much of the work learners have to do. The best way to teach is to preserve the germane load, which occurs when AI detects an error and asks the student to fix it on their own. The SAMR model (Puentedura, 2006) operationalizes these differences at the task level. In a systematic review published in Discover Computing (Springer Nature), most AI integrations were at the Substitution or Augmentation level, whereas those at the Modification and Redefinition levels were significantly less prevalent (Bao et al., 2025). Redefinition is illustrated in a Vietnamese EFL study in which 130 university students participated in AI-mediated inquiry-based reading projects. The students developed multimedia infographics and podcasts, resulting in notable improvements in higher-order thinking skills and enhanced learner confidence (Duc, 2026).

AI's Structural Limitations: Culture, Discipline, and Rhetoric

Most of the time, generative AI is trained on English corpora from Western native speakers. These corpora hold ideas of what constitutes good writing that do not always align with the rhetorical traditions of learners from non-Western backgrounds. Xiao and colleagues (2025) found that students repeatedly noted concern about culturally insensitive comments. Werdiningsih et al. (2024) found that Indonesian EFL students believed AI suggestions were not rhetorically consistent with their academic and cultural contexts, resulting in grammatically correct but culturally inappropriate text. According to Wang et al. (2025), the field's focus on ChatGPT may lead to the privileging of certain cultural-linguistic norms at the expense of the diversity that real language education ought to celebrate.

A Framework: AI Task Typology by Cognitive Engagement

Drawing on ZPD (Vygotsky, 1978), CLT (Sweller, 1988), and SAMR (Puentedura, 2006), the following framework organizes AI-assisted language-learning tasks by cognitive engagement level.

Table 1

AI task typology by cognitive engagement

Level	SAMR	Learner Role	Example	Educational Value
1 — Substitutive	Substitution	Passive recipient	Submitting essay to AI for grammar correction without engaging with explanations	Low; high dependency risk
2 — Augmentative	Augmentation	Active selector	Using AI vocabulary suggestions, then independently selecting and contextualizing options	Moderate; appropriate as supplement
3 — Scaffolded Revision	Modification	Critical evaluator	Using AI to generate counterarguments, then revising one's own position in response; comparing AI and teacher feedback	High; promotes metacognitive engagement
4 — Transformative	Redefinition	Inquiry co-creator	Designing AI-mediated cross-cultural dialogue projects; using AI to generate competing perspectives for critical synthesis	Very high; requires teacher design expertise

The point is that Levels 1 and 2, where most of the AI is currently being used, only offer the small improvements that have been written about. Levels 3 and 4 are the real places AI could make a difference. This is not essentially a problem of technology. Levels 1 and 4 differ in the teacher's decisions about how to teach, not in the AI tool.

RQ2: Academic Integrity, Critical Thinking, and the Ethics of AI Use

Reframing the Problem: From Detection to Design

The main reaction of schools to AI in language education has been defensive, with schools using AI detection tools, changing their plagiarism policies, and warning students against misusing AI. These measures help with the symptoms, but they do not get to the root of the problem. Academic integrity in the age of AI isn't just about catching the cheaters. It's about how to teach. Assessment tasks that rely exclusively on outputs readily produced by generative AI are insufficient for measuring the competencies that language educators seek to assess.

This distinction is precisely articulated by Corbin et al. (2025) in Assessment & Evaluation in Higher Education (Taylor & Francis) as "discursive changes" that do not alter the mechanics of assessment, but instead focus on communicating policies on AI use. Instead, what is needed are "structural changes". These changes would change both what is being tested and how. Recent survey data show that students are already using AI extensively, but institutions are still

not providing enough guidance. In the Digital Education Council Global AI Student Survey 2024, 86% of students reported using AI in their studies, 54% reported using it at least weekly, and only 5% reported being fully aware of institutional AI policies (Digital Education Council, 2024).

AI Literacy as a Language Learning Competency

One of the key points of this part is that AI literacy, or the ability to critically evaluate AI outputs, create effective prompts, understand the limits of AI, and make decisions based on ethics, should be considered a language learning skill in itself. Wang et al. (2025) explicitly argue that prompt engineering is an important but under-studied component of AI literacy. They say that students who can write good prompts get qualitatively different, more useful AI outputs for learning than students who just ask for vague requests.

An autoethnographic study published by Hsu (2025) proposed the SUPER framework for the ethical and effective use of ChatGPT in academic writing. The framework consists of 5 key ideas: Support Not a Substitute (AI should help brainstorm, not do the whole thing); Unique Perspective (the writer's own voice should be the main focus of the work); Prompt Engineering (writing clear, iterative, and context-specific prompts); Ethical Use (following institutional rules and revealing AI's contributions); and Reflection (regularly asking whether AI is a help or a hindrance to one's intellectual effort). The framework is particularly useful because it considers the psychological aspects of using AI, which students often turn to as a means of escaping writing anxiety, and it offers clear instructions for converting this motivation into an effective learning strategy.

A systematic review by Lee et al. (2025) found that pedagogical framing was the primary determinant of AI's impact on critical thinking in EFL environments. Integrating AI into tasks, such as asking learners to analyze AI-generated content to identify inaccuracies, assess cultural relevance, and compare it with human perspectives, led to significant improvements in critical thinking.

From AI Policing to AI-Aware Assessment Design

The best answer to integrity and critical thinking problems is to change the way assessments are structured to require personal, contextualized, iterative, and reflective engagement that AI will not easily replicate. In a qualitative study, Khlaif et al. (2025) surveyed 61 faculty members in Education Sciences and identified four major reasons for the re-designing of assessments in the artificial intelligence era: to maintain academic integrity, to prepare learners to work in the AI-mediated professional contexts, to adapt to technological developments, and to conform to institutional policy.

There is substantial empirical evidence supporting many task design principles for language teachers. The best strategy is probably the oral defense of written work, so that it will be very difficult for AI to replace students' work for a long time. This is because students are required to explain and answer questions about their written work (Corbin et al., 2025). The nature of process portfolios, which involves documenting brainstorming notes, successive drafts, feedback logs, and reflective commentaries, makes it almost impossible for AI alone to fabricate the entire learning journey. Text that is culturally and personally situated cannot be authentically produced by generic AI tools. Werdiningsih et al. (2024) found that Indonesian EFL learners instructed to use knowledge derived from their own culture were much more likely to critically evaluate AI suggestions than to accept them as they were.

Five Design Principles for AI-Resilient Assessment

Based on the evidence examined, five design principles are suggested:

- Principle 1: Make sure the tasks fit with personal and cultural information. Tests should require knowledge, experience, or cultural positioning that AI cannot draw on, such as personal narrative, community-based inquiry, or locally situated rhetorical conventions.
- Principle 2: Make the process and the product accessible. Process portfolios, revision logs, and brainstorming records all demonstrate what you've learned that can't be faked with just one AI.
- Principle 3: Include parts that are spoken. An oral defense or structured discussion of written work necessitates real-time, spontaneous demonstration of comprehension—a type of communicative competence that generative AI cannot replicate on a learner's behalf (Corbin et al., 2025).
- Principle 4: Use comparative critique as a way to teach. When students compare their own writing to AI-generated writing, explain why they made the choices they did, and look for differences, they are learning how to use AI and language in real ways.
- Principle 5: Make sure that institutional policy matches how tasks are set up and how teachers are trained. When institutional policies are vague, harsh, or disconnected from professional development support, it becomes harder to redesign assessments.

Implications for Teacher Development and Professional Practice

The Most Critical, Most Underserved Actor

The classroom teacher is the most important and least supported person in the AI-in-language-education ecosystem. Every argument made in the previous sections comes down to one practical truth: the real benefits of AI can only be realized, and the risks to its integrity and critical thinking can only be reduced, by teachers who have the professional skills, institutional support, and confidence in their teaching methods to make principled decisions about how to use AI. UNESCO's AI Competency Framework for Teachers notes that only 7 countries worldwide had developed AI frameworks or programs for teachers by 2022. This shows how unprepared the system is on a large scale.

Teacher AI Literacy: A Multi-Dimensional Professional Competency

Ng et al. (2021), in a seminal framework published in Computers and Education: Artificial Intelligence, delineate AI literacy as comprising four interconnected domains: foundational conceptual knowledge of AI; practical pedagogical integration of AI; critical evaluation of AI tools' pedagogical utility, accuracy, cultural appropriateness, and ethical alignment; and awareness of algorithmic bias, student privacy, equity, and broader implications. These four dimensions are interdependent; a teacher proficient in operating AI tools yet incapable of critically assessing their cultural biases lacks AI literacy in an educationally significant manner.

Du et al. (2025), in a qualitative case study of Chinese university EFL teachers published in Empowering Educators (Springer Nature), identified that effective teachers possess what the researchers term Intelligent-TPACK—an evolving knowledge base that integrates technological, pedagogical, and content knowledge with a focused emphasis on AI ethics. Bahari and Liu (2025), in a comprehensive pretest-posttest experimental study involving 184 EFL teachers, offered unique experimental evidence that structured, theoretically informed

professional development aligned with a multidimensional AI literacy framework yielded substantial, quantifiable improvements in digital competency and professional engagement. Pan and Wang (2025), in the European Journal of Education (Wiley-Blackwell), found that higher AI literacy is significantly associated with ethical, critically engaged, and pedagogically intentional AI integration.

The Global Competency Gap

There is a global structural problem: the gap between what teachers need to know about AI and what they already know. Nazim and Alzubi (2025), in PLOS ONE, found that 278 university EFL teachers in Saudi Arabia were most worried about overreliance. They also found that institutional policies and ethical frameworks were important but lacked safeguards. Ilma and Rohmah (2025), in Cogent Education (Taylor & Francis), found that a gap persists between the Indonesian government's push for digital integration and teachers' actual AI skills. Many teachers said they were aware of AI tools but had little technical knowledge and almost no training in using them in the classroom. Babanoğlu et al. (2025), discovered that although prospective EFL teachers acknowledged AI's potential, they concurrently expressed a sense of unpreparedness for its pedagogical and ethical integration.

Teacher Identity Under Pressure

Du et al. (2025) identified three clusters of identity tension experienced by Chinese EFL teachers in adapting to AI integration: conflicting I-positions (contested pedagogical beliefs on student-centered teaching); new I-positions (redefined responsibilities in lesson design, assessment, and cultivation of AI literacy); and constrained I-positions (obstacles such as the absence of institutional policy, students' uncritical engagement with AI, and limited Intelligent-TPACK). The traditional two-way interaction between the teacher and the student has been converted into a three-way interaction between the teacher, AI, and the student. And this changed the power dynamic and the way grades were handed out. These tensions of identity are not just psychological problems for individuals but structural ones, resulting from the disjuncture between the pedagogical demands posed by AI and the institutional resources available to meet them.

A Reform Roadmap for Language Teacher AI Education

A three-tier roadmap is suggested based on UNESCO (2024) and the requirements of language education:

Tier 1: Pre-Service Education: Programs should focus on teaching pedagogical reasoning about AI rather than tool use. This requires basic knowledge of conceptual AI, domain-specific AI literacy for language education (e.g., the cultural appropriateness of AI feedback, development of AI-integrated tasks at appropriate SAMR levels), ethics-informed design practice, and supervised practicum experiences in teaching AI-integrated lessons.

Tier 2: In-Service Professional Development: Programs must be ongoing (not one-time events) and grounded in the four-dimensional AI literacy framework of Ng et al. (2021). They should also be members of communities of practice where teachers exchange task designs integrated with AI, critically evaluate AI tools, and formulate ethical guidelines that are agreed upon by all. It needs administrative support (clear guidance on the curriculum, protected time, etc.)

Tier 3: Institutional Policy: Schools should develop AI use policies focused on how AI can support learning, rather than simply banning it. They also need to develop assessment frameworks that account for AI and ensure that all teachers and students have the technology they need to use AI responsibly.

Conclusion: Implications for Research, Practice, and Policy

The four arguments presented in previous Sections concerning pedagogical scaffolding, academic integrity, assessment validity, and teacher development are interrelated issues. They are different parts of the same basic problem: how to use AI in ways that really help language teaching and learning. The Framework for Responsible AI Integration in Language Education (FRAILE), shown in Figure 1, brings these four areas together into a clear, useful model that operates at three levels that support one another.

Figure 1.

The FRAILE Framework: Framework for Responsible AI Integration in Language Education

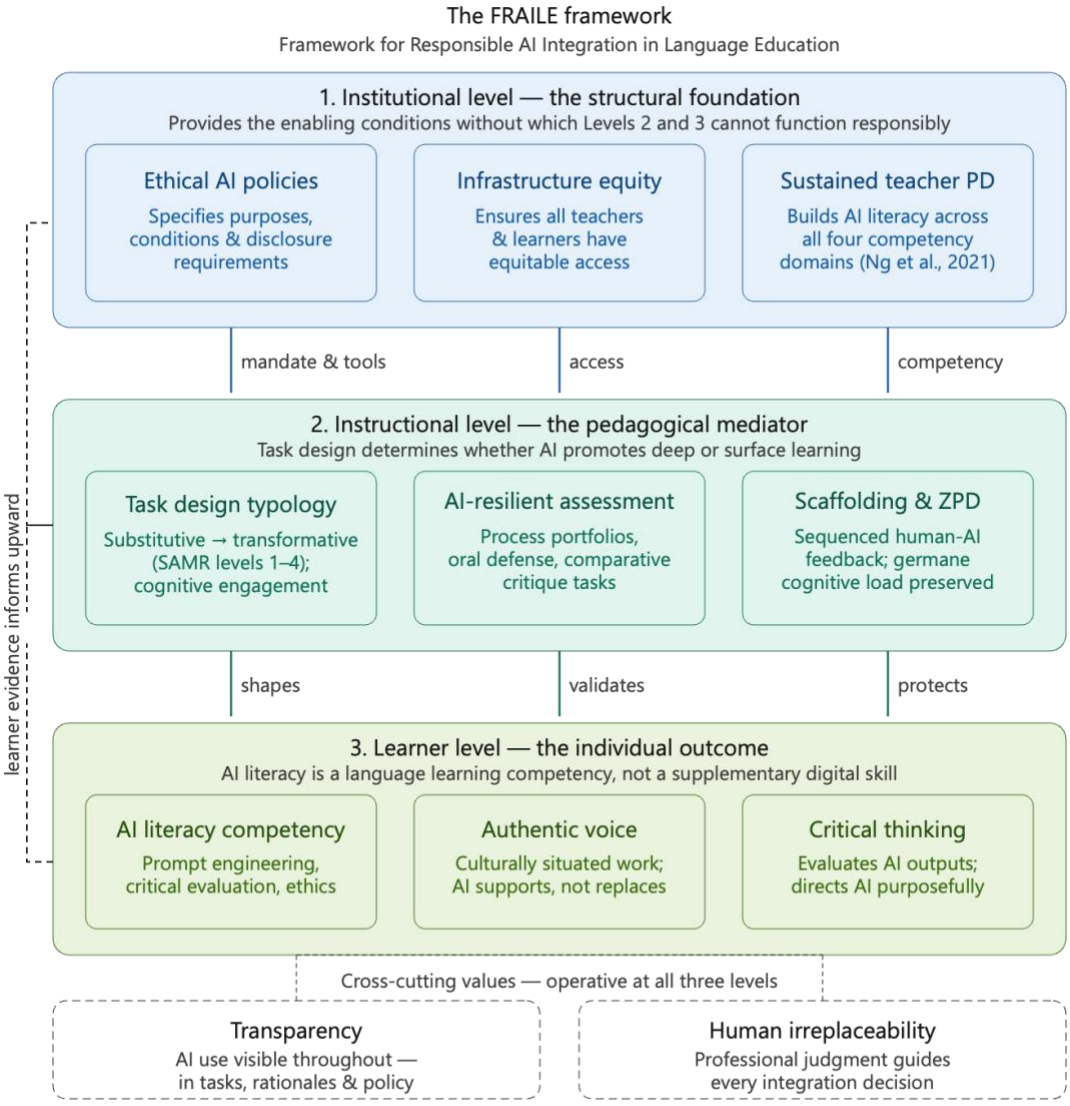

Figure 1 shows that FRAILE's three levels are neither hierarchical nor sequential; they must grow simultaneously. At the institutional level, the authority, access, and skills are available to design at the instructional level. The instructional level creates, monitors, and protects learner outcomes. Evidence of those outcomes is fed back to inform both institutional policy and

instructional practice. At all three levels, two general principles apply: transparency and irreplaceability of humans. These principles guarantee that AI integration is human-centric and guided by critical analysis.

This paper has addressed two central questions of AI in language education. The findings for RQ1 suggest that AI consistently enhances surface-level language performance (e.g., grammar accuracy, writing fluency, vocabulary range) across a range of EFL and ESL contexts. However, these benefits are largely realized at the Enhancement levels of the SAMR model and are highly dependent upon instructor scaffolding. For the moment, the promise of AI for deep, critical, and culturally authentic language development continues to rely on pedagogical design conditions that are not yet widely adopted in the field. The AI Task Typology introduced here identifies four levels of cognitive engagement, from Substitutive to Transformative, and offers language educators a functional design lexicon for moving AI use towards achieving authentic educational transformation.

In relation to RQ2, the field's focus on AI detection tools is symptomatic rather than root-cause oriented. This evidence suggests a better path: constructing AI-resistant and AI-informed assessments, instilling AI literacy in learners as a pedagogical duty, and preemptively cultivating teachers' professional skills to ethically incorporate AI. The SUPER framework of Hsu (2025), the strategy-graded reliance model of Hou et al. (2025), and the five AI-resilient assessment design principles presented here together provide a practically applicable response.

Implications for Researchers

The field requires a distinct move toward longitudinal, mixed-methods, and cross-contextual research methodologies. There is insufficient evidence to support the causal, developmental, or generalizable claims required for responsible pedagogical guidance (Wang et al., 2025; Li et al., 2025). The current evidence base is largely composed of short-term, single-institution studies of higher-education EFL in East Asia. Research on the integration of AI across contexts in K-12, vocational, heritage-language, and multilingual contexts would increase the generalizability of the evidence base. What is needed urgently is an integrative framework allowing for a constructive dialogue between research on language learning and teacher education.

Implications for Practitioners and Institutions

Develop design tasks for language teachers at AI Typology Levels 3 and 4. Make AI literacy development a primary teaching goal. Use portfolios and oral components to clarify and measure learning. Push for structured professional development. For institutions: Develop AI policies that are guiding, not just restrictive; invest in long-term, theory-based professional development aligned with UNESCO (2024); and prioritize infrastructure equity so that responsible AI integration benefits all learners and teachers, not just those in well-resourced contexts.

The educational value of AI for language learning does not stem from the technology itself, but from the quality of human decisions about when, how, and why to use it. Language teachers are the best people to make those decisions because they know about culture, communication, and the deeply human aspects of making meaning. What they need now from researchers, institutions, and policymakers is evidence, support, and training to become better.

References

Asadi, M., Ebadi, S., & Mohammadi, L. (2025). The impact of integrating ChatGPT with teachers' feedback on EFL writing skills. *Thinking Skills and Creativity, 56*, 101766. https://doi.org/10.1016/j.tsc.2025.101766

Babanoğlu, M. P., Öztürk Karataş, T., & Dündar, E. (2025). Envisioning the future of AI-assisted EFL teaching and learning: Conceptual representations of prospective teachers. *SAGE Open*. https://doi.org/10.1177/21582440251341590

Bahari, A., & Liu, Y. (2025). AI integration in EFL teacher development: a mixed-methods evaluation of digital competency, professional trajectories, and pedagogical innovation within adaptive learning ecosystems. *Interactive Learning Environments*, 1-17. https://doi.org/10.1080/10494820.2025.2591251

Bao, W., Wang, T., Zhang, L., Yusop, F. D., & Ruan, X. (2025). A systematic review of AI in second language acquisition using the expanded SAMR model (2015–2024). *Discover Computing, 28*(1), 292. https://doi.org/10.1007/s10791-025-09833-6

Corbin, T., Dawson, P., & Liu, D. (2025). Talk is cheap: Why structural assessment changes are needed for a time of GenAI. *Assessment & Evaluation in Higher Education, 50*(7), 1087–1097. https://doi.org/10.1080/02602938.2025.2503964

Digital Education Council. (2024, August 7). *What students want: Key results from DEC global AI student survey 2024*. Digital Education Council. Retrieved from https://www.digitaleducationcouncil.com/post/what-students-want-key-results-from-dec-global-ai-student-survey-2024

Du, C., Peng, Y., binti Ahmad, N. K., & binti Jamil, A. H. (2025). EFL teachers' identity tensions and transformation in AI-driven teaching in China: A dialogical approach. In V. P. H. Pham et al. (Eds.), *Empowering educators: Integrating AI tools for personalized language instruction*. Springer. https://doi.org/10.1007/978-3-032-01348-4_7

Duc, D. H. (2026). Operationalizing SAMR Redefinition in EFL reading: AI as a mediating tool for literacy innovation. *Digital Technologies Research and Applications, 5*(1), 66–82. https://doi.org/10.54963/dtra.v5i1.1900

Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies, 15*(1), Article 6. https://doi.org/10.3390/soc15010006

Hariyanto, Kristianingsih, F.X.D. & Maharani, R. (2025). Artificial intelligence in adaptive education: a systematic review of techniques for personalized learning. *Discover Education, 4*, 458. https://doi.org/10.1007/s44217-025-00908-6

Hou, C., Zhu, G., & Sudarshan, V. (2025). The role of critical thinking on undergraduates' reliance behaviours on generative AI in problem-solving. *British Journal of Educational Technology, 56*(5), 1919–1941. https://doi.org/10.1111/bjet.13613

Hsu, H.-P. (2025). An autoethnographic study of ESL academic writing with ChatGPT: From psychological insights to the SUPER framework. *Cogent Education, 12*(1), 2543113. https://doi.org/10.1080/2331186X.2025.2543113

Khan, R., Qamar, M. T., Ansari, M. S., & Yasmeen, J. (2025). Enhancing or impairing? Exploring Indian EFL learners' academic writing narratives with ChatGPT. *Cogent Education, 12*(1), 2514329. https://doi.org/10.1080/2331186X.2025.2514329

Lee, S., Choe, H., Zou, D., & Jeon, J. (2025). Generative AI (GenAI) in the language classroom: A systematic review. *Interactive Learning Environments, 34*(1), 335–359. https://doi.org/10.1080/10494820.2025.2498537

Li, B., Tan, Y. L., Wang, C., & Lowell, V. (2025). Two years of innovation: A systematic review of empirical generative AI research in language learning and teaching. *Computers and Education: Artificial Intelligence*, 100445. https://doi.org/10.1016/j.caeai.2025.100445

Liu, J., Sihes, A. J. B., & Lu, Y. (2025). How do generative artificial intelligence (AI) tools and large language models (LLMs) influence language learners' critical thinking in EFL education? A systematic review. *Smart Learning Environments, 12*(1), 48. https://doi.org/10.1186/s40561-025-00406-0

Lo, C.K., Yu, P.L.H., Xu, S. *et al.* (2024). Exploring the application of ChatGPT in ESL/EFL education and related research issues: A systematic review of empirical studies. *Smart Learning Environments, 11*, 50. https://doi.org/10.1186/s40561-024-00342-5

Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments, 11*(1), 9. https://doi.org/10.1186/s40561-024-00295-9

Khlaif, Z. N., Alkouk, W. A., Salama, N., & Abu Eideh, B. (2025). Redesigning assessments for AI-enhanced learning: A framework for educators in the generative AI era. *Education Sciences*, *15*(2), 174. https://doi.org/10.3390/educsci15020174

Nazim, M., & Alzubi, A. A. F. (2025). Empowering EFL teachers' perceptions of generative AI-mediated self-professionalism. *PLOS ONE, 20*(6), e0326735. https://doi.org/10.1371/journal.pone.0326735

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence, 2*, 100041. https://doi.org/10.1016/j.caeai.2021.100041

Nguyen, L. N. A., & Pham, T. D. (2025). Integrating AI tools in IELTS writing: A case study from teachers' perspectives on personalized instruction. In V. P. H. Pham et al. (Eds.), *Empowering educators: Integrating AI tools for personalized language instruction*. Springer Nature. https://doi.org/10.1007/978-3-032-01348-4_4

Pan, Z., & Wang, Y. (2025). From technology-challenged teachers to empowered digitalized citizens: Exploring the profiles and antecedents of teacher AI literacy in the Chinese EFL context. *European Journal of Education, 60*(1), e70020. https://doi.org/10.1111/ejed.70020

Pham, V. P. H., & Huynh, Q. Q. (2025). Employing AI tools for vocabulary acquisition and autonomous learning. In V. P. H. Pham et al. (Eds.), *Empowering educators: Integrating AI tools for personalized language instruction*. Springer Nature. https://doi.org/10.1007/978-3-032-01348-4_10

Polakova, P., & Ivenz, P. (2024). The impact of ChatGPT feedback on the development of EFL students' writing skills. *Cogent Education, 11*(1), 2410101. https://doi.org/10.1080/2331186X.2024.2410101

Puentedura, R. R. (2006). *Transformation, technology, and education*. Hippasus. Retrieved from http://hippasus.com/resources/tte/

Reuters. (2023) *ChatGPT sets record for fastest-growing user base - analyst note.* Reuters. Retrieved on 16 April 2026 from https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Torraco, R. J. (2016). Writing integrative literature reviews: Using the past and present to explore the future. *Human Resource Development Review, 15*(4), 404–428. https://doi.org/10.1177/1534484316671606

UNESCO. (2024). *AI competency framework for teachers*. United Nations Educational, Scientific and Cultural Organization. https://unesdoc.unesco.org/ark:/48223/pf0000391104

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wang, Y., Zhang, T., Yao, L., & Seedhouse, P. (2025). A scoping review of empirical studies on generative artificial intelligence in language education. *Innovation in Language Learning and Teaching*, 1–28. https://doi.org/10.1080/17501229.2025.2509759

Werdiningsih, I., Marzuki, & Rusdin, D. (2024). Balancing AI and authenticity: EFL students' experiences with ChatGPT in academic writing. *Cogent Arts & Humanities, 11*(1), 2392388. https://doi.org/10.1080/23311983.2024.2392388

Ilma, A., & Rohmah, Z. (2025). AI in EFL education: teachers' competence and the roadblocks to teaching material development. *Cogent Education*, *12*(1), 2588471. https://doi.org/10.1080/2331186X.2025.2588471

Xiao, F., Zhu, S., & Xin, W. (2025). Exploring the landscape of generative AI (ChatGPT)-powered writing instruction in English as a foreign language education: A scoping review. *ECNU Review of Education, 9(1), 1-19*. https://doi.org/10.1177/20965311241310881

Yuan, L., & Liu, X. (2025). The effect of artificial intelligence tools on EFL learners' engagement, enjoyment, and motivation. *Computers in Human Behavior, 162*, Article 108474. https://doi.org/10.1016/j.chb.2024.108474

Biodata

Assoc. Prof. Dr. Pham Vu Phi Ho, Faculty of Foreign Languages, Industrial University of Ho Chi Minh City, Vietnam, was previously Vice President of Ba Ria – Vung Tau University and Vice President of Van Hien University, Vietnam. He created and designed ICTE Robots for language learners and IELTs practices. Pham has published more than 80 research articles in both local and International Journals, such as Computer-Assisted language Learning, SAGE Open, System, and Journal of Psycholinguist Research (23 Scopus/14 ISI), and 10 books and course-books, 3 edited books (2 Scopus-indexed) published by Springer Nature (2) and IGI Global (1) of which 3-course books were used for both the undergraduate and graduate levels at Van Lang University, HCMC Open University, Vietnam, and Lourdes College, Higher Education Department, Cagayan de Oro City, Philippines. He has international *teaching experience at Suranaree University of Technology in Thailand and Gyeongju University in* South Korea. He is the President of AsiaCALL. He is now the Editor-in-Chief of the

International Journal of TESOL & Education. He is a managing editor of CALL-EJ (Scopus Q1), an editorial staff member for the *Suranaree Journal of Social Science* (Scopus Q3), and a peer reviewer for some international Journals indexed in ISI/Scopus such as Computer Assisted Language Learning, Open Sage... (*Scopus*), (*ISI*), (*Google Scholar*).

Dr. Willy A Renandya is a language teacher educator with extensive teaching experience in Asia. He currently teaches language education courses in Singapore, including at NIE-NTU, SUSS, and RELC. He is also a visiting professor at Chulalongkorn University and Wuhan University. He has given numerous keynote presentations at international ELT conferences, including AsiaTEFL, IATEFL, KOTESOL, SEAMEO RELC, and ThaiTESOL. He has published extensively in the area of second language education and second language acquisition. He maintains a large teacher professional development forum called Teacher Voices: https://www.facebook.com/groups/teachervoices/.

Website: Willy's ELT Corner https://www.willyrenandya.com
Google Scholar: https://scholar.google.com.sg/citations?user=gHW1fVIAAAAJ