From Generative to Agentic Artificial Intelligence in Foreign Language Teaching, Learning, and Assessment

Book Series

Contemporary Studies on Digital Humanities Vol. I
Ali Tilbe, Yusuf Topaloğlu & Gökhan Dinar (Eds.)

Tuncer Can

From Generative to Agentic Artificial Intelligence in Foreign Language Teaching, Learning, and Assessment

maurer.press
Frankfurt am Main

Bibliographic Information of the German National Library

The German National Library lists this publication in the German National Bibliography; detailed bibliographic data are available online at the German National Library website: http://dnb.d-nb.de

© Published by Verlag Hans-Jürgen Maurer

under the Creative Commons Attribution 4.0 International License (CC BY 4.0).

The work, including all of its parts, is protected by copyright. Any use beyond the narrow limits of copyright law without the consent of the copyright holder is prohibited and punishable by law. This applies in particular to reproductions, translations, and processing in electronic systems.

The book has undergone a double blind peer review.

maurer.press

is an imprint of
Verlag Hans-Jürgen Maurer
Im Trierischen Hof 14
60311 Frankfurt/Germany
www.maurer.press
GPSR-Kontakt: info@maurer.press

ISBN 978-3-949550-47-8
e-ISBN 978-3-949550-48-5

Table of Contents

Preface vii

Acknowledgments viii

Abstract ix

PART I: FOUNDATION: AI & LANGUAGE EDUCATION

Chapter 1. Introduction: Why AI Matters for Language: Education Now 14

Chapter 2. Understanding Artificial Intelligence: Concepts, Technologies, and System Archetypes 34

Chapter 3. Theoretical Frameworks for AI-Mediated Language Education 52

PART II: AI IN LANGUAGE TEACHING

Chapter 4. AI as a Teaching Assistant: Planning, Materials, and Tasks 78

Chapter 5. AI-Supported Teaching of Language Skills 100

Chapter 6. Teacher Roles, Beliefs, and AI Pedagogical Literacy 133

PART III: AI IN LANGUAGE LEARNING

Chapter 7. AI-Mediated Learning Environments: Intelligent Tutors, Conversational Agents, and Adaptive Systems 168

Chapter 8. Learner Autonomy, Self-Regulation, and Agency in AI-Mediated Language Education 195

Chapter 9. Multimodality, Interaction, and Identity 225

PART IV: AI IN LANGUAGE ASSESSMENT

Chapter 10. Automated Assessment and Feedback 252

Chapter 11. Validity, Fairness, and Academic Integrity in AI-Mediated Language Assessment 300

PART V: ETHICS, EQUITY, POLICY, & FUTURE DIRECTIONS

Chapter 12. Ethics, Equity, and Policy in AI-Mediated Language Education 338

Chapter 13. Conclusion: Towards Responsible and Pedagogically Grounded AI Use 404

Chapter 14. From Generative to Agentic: AI Agents and the Next Generation of Language Learning Platforms .. 436

References 451

Preface

This book grew out of a conviction that the current transformation of language education deserves neither celebration nor alarm, but disciplined attention. Artificial intelligence has moved, in a remarkably short time, from the margins of computer-assisted language learning to the center of how languages are taught, practiced, and assessed. The thirteen chapters that follow—and a fourteen looking toward the agentic systems now emerging—attempt a comprehensive, theoretically grounded account of that movement across the three domains named in the title, held together by a single thesis: that the educational value of AI lies in augmentation, not automation, and that this is a design condition to be achieved rather than an attitude to be professed.

The book is addressed to graduate students, teacher educators, researchers, and thoughtful practitioners. Each chapter closes with a pedagogical companion—key concepts, discussion questions, and further reading—intended to support its use in graduate seminars and professional development settings. The book's synthesis culminates in the PAGE framework (Chapter 13), an integrative model for responsible AI integration, which is then applied prospectively to agentic language learning platforms (Chapter 14).

A note on method. This book is a narrative synthesis of the research literature, with particular attention to the generative and post-generative-AI period of 2023–2026. Literature was identified through structured searches of scholarly databases and citation indices—including targeted searches of the field's core journals in computer-assisted language learning and language testing—and through citation-chaining from key syntheses. The synthesis privileges peer-reviewed empirical work and systematic reviews; where preprints or grey literature are cited, this is indicated in the reference. Chapter 14 documents its own search procedure in full as a model of the transparency the book advocates.

Declaration on the use of artificial intelligence. Artificial intelligence tools were used in the preparation of this manuscript—in drafting assistance, language editing, and the compilation of literature—under the author's direction and with the author's full responsibility for all content. Following the initial draft, the manuscript underwent a comprehensive human-directed revision in which every bibliographic entry was individually

verified against scholarly databases (Scite, publisher records, and citation indices); entries that could not be verified were removed or corrected, and all cited claims were checked against their sources. No AI system is credited with authorship, and all interpretations, arguments, and remaining errors are the author's own. This declaration is made in accordance with current publisher policies on AI-assisted manuscript preparation.

Acknowledgments

On a personal note, I would like to extend my deepest gratitude and love to my wife, Betim Neyran Can, for her unwavering support, patience, encouragement, and understanding throughout this journey. Her presence and belief in my work have been a constant source of strength and motivation.

I am also deeply grateful to my son, Tuncer Ada Can, and my daughter, Neyran Lara Can, for their love, patience, and inspiration. Their presence has brought joy and meaning to every stage of this endeavor.

I would like to express my sincere gratitude to Prof. Dr. İrfan Şimşek for his invaluable academic and intellectual support, insightful guidance, and encouragement throughout the development of this book. His expertise and thoughtful contributions have been greatly appreciated.

To all of them, I offer my heartfelt thanks for being an invaluable part of this journey and for making the completion of this book possible.

Abstract

Artificial intelligence is rapidly reshaping foreign language education, transforming not only the technologies available to teachers and learners but also the pedagogical, professional, and ethical conditions under which languages are taught, learned, and assessed. *From Generative to Agentic Artificial Intelligence in Foreign Language Teaching, Learning, and Assessment* provides a comprehensive and theoretically grounded examination of this transformation, tracing the evolution from computer- and mobile-assisted language learning to generative AI, large language models, multimodal systems, and emerging agentic AI architectures.

Drawing on recent empirical research, systematic reviews, and developments in AI-mediated language education, the book examines the implications of artificial intelligence across three interconnected domains: teaching, learning, and assessment. It explores AI-supported instructional design, materials development, language-skill instruction, personalized and adaptive learning, conversational agents, learner autonomy and self-regulation, multimodality, automated feedback and scoring, and changing teacher and learner roles. Particular attention is given to the Teacher–AI–Student relationship and to the growing need for AI pedagogical literacy as educators move from using AI as a tool toward orchestrating increasingly autonomous systems.

Alongside these opportunities, the book critically addresses questions of validity, fairness, academic integrity, algorithmic bias, privacy, cognitive offloading, unequal access, and overreliance on automated systems. It argues that the educational value of AI lies in *augmentation rather than automation* and that technological capability should remain subordinate to pedagogical purpose and human judgment. This position is consolidated through the *PAGE framework—Pedagogy-first, Agency-preserving, Governance-bound, and Equity-oriented*—as an integrative model for evaluating and designing responsible AI use in language education.

Looking beyond generative AI, the book concludes by examining the emergence of agentic AI and multi-agent language-learning environments capable of planning, adapting, evaluating, and coordinating learning processes with increasing autonomy. It argues that this transition intensifies rather than diminishes the need for human oversight, transparent governance, assessment validity, and pedagogical accountability. Consequently, the

volume proposes a human-centered vision of AI-mediated language education in which innovation is guided by pedagogy, learner and teacher agency, ethical responsibility, and equitable access.

Keywords: Artificial Intelligence; Generative AI; Agentic AI; Foreign Language Education; Language Teaching and Learning; Language Assessment; Large Language Models; AI-Mediated Language Learning; Learner Agency; Teacher–AI–Student Triad; AI Pedagogical Literacy; PAGE Framework; Human-Centered AI; Automated Feedback and Assessment; Responsible AI

PART I

Foundation: AI & Language Education

Chapter 1. Introduction: Why AI Matters for Language. Education Now

1.1 A Moment of Convergence in Language Education

The integration of Artificial Intelligence (AI) into educational contexts—particularly in foreign language (FL) and second language (L2) education and, most visibly, in English Language Teaching (ELT)—has become a transformative force in recent years. AI has moved beyond the role of a supplementary tool and is increasingly positioned as a *collaborative partner* that can amplify pedagogical capacity: generating language, responding interactively, and supporting decision-making across teaching, learning, and assessment. This shift has accelerated with Generative AI (GenAI) and large language models (LLMs) such as ChatGPT, whose core capability—processing and generating human language—directly intersects with the fundamental objective of language education (Mateos-Blanco et al., 2024; Wang, Zhang, Yao, & Seedhouse, 2025).

What makes the current moment distinctive is not simply technological novelty, but the *convergence* of (a) increasing global multilingual communication demands, (b) the datafication of educational processes, and (c) AI systems that can participate in learning interactions and produce high-quality linguistic output. Recent bibliometric mappings indicate rapid growth in AI-related language education research, with thematic clusters emphasizing AI-supported instruction, learner engagement, and evolving AI frameworks—alongside persistent concerns about ethics, equity, and sustainable implementation (Yang et al., 2025; Zhang, Hussin, & Abd Majid, 2025). In other words, AI is no longer peripheral to language education; it is shaping the field's research agendas and practical debates.

This growth in attention should not be mistaken for uniform evidence. A multilevel meta-analysis of 117 effect sizes drawn from 46 EFL studies conducted between 2022 and 2025 found that AI-based instruction produced a medium-to-large overall effect on language learning—but the effect was strongest for vocabulary and weakest for speaking (Torres & Kahveci, 2025). The chapters that follow return to this unevenness repeatedly: some claims made on AI's behalf are backed by convergent, well-powered findings; others rest on smaller samples, shorter interventions, or outcome measures that capture performance rather than learning. Distinguishing, which is which, skill by skill and claim by claim, is one of this book's recurring tasks.

1.2 Purpose and Orientation of This Book

This book is written in response to this pivotal moment. It does not treat AI as a passing trend, nor does it assume that technological capability automatically yields educational benefit. Instead, it positions AI as a powerful but malleable force whose impact depends on pedagogy, institutional policy, and ethical commitments.

By synthesizing research on FL/L2 teaching, learning, and assessment, this volume aims to clarify why AI matters now and how language educators can respond with both ambition and caution. The book's synthesis culminates in an integrative contribution of its own: the PAGE framework (Pedagogy-first, Agency-preserving, Governance-bound, Equity-oriented), developed in Chapter 13, which consolidates the volume's principles into a diagnostic model for responsible AI integration and is applied prospectively to emerging agentic platforms in Chapter 14. The chapters that follow examine AI as a paradigm shift in language education, explore practical and research-based integrations, and address assessment innovation and integrity in an era where machine-generated language is increasingly ubiquitous (Wang et al., 2025; Yang et al., 2025; Zhang et al., 2025).

Ultimately, the question for language education is not whether AI will be present—because it already is—but how it will be shaped to enhance learning, protect equity, sustain academic integrity, and preserve the human dimensions of language, culture, and meaning.

1.3 The Global Context: English as a Lingua Franca and Scalable Language Provision

AI matters now partly because language education faces intensified global pressures. English, in particular, continues to operate as a lingua franca across academia, commerce, and international mobility, generating strong institutional and learner demand for scalable, effective language provision. As higher education institutions increasingly integrate language education with employability and disciplinary training, AI-driven resources offer an efficient pathway to expand practice opportunities and feedback capacity at scale—especially in contexts where class sizes, time constraints, and instructor workload limit individualized support (Wang et al., 2025; Zhang et al., 2025).

This does not imply that AI is a simple solution to long-standing structural challenges; rather, it foregrounds why AI adoption has become urgent. The demand for more language learning opportunities—more interaction, more feedback, more personalization—has collided with the practical limits of human-only provision. AI enters precisely at this pressure point.

The unevenness of this provision is documented, not merely assumed. A PRISMA/TPACK review of 125 empirical studies on generative AI in higher education (2021–2025) found a negligible correlation ($r = 0.12$) between institutional investment in faculty development and actual pedagogical transformation, alongside persistent Global North–South divides in who benefits from that investment (Altamirano et al., 2026). Within EFL specifically, a review of 22 empirical studies drawn from 189 screened records found significant learning gains from AI tools in school settings, but reported that digital-divide and teacher-readiness challenges persisted alongside those gains (Kundu & Bej, 2025). Even the scholarship studying these inequities is unevenly distributed: a decade-spanning bibliometric analysis of AI-in-education-ethics research found the field expanding rapidly in volume while remaining concentrated in the United States, China, and the European Union (Li, 2025)—the same handful of contexts that already enjoy the greatest instructional capacity.

1.4 From Educational Technology to Educational Infrastructure

Earlier technologies in language education (e.g., CALL and MALL) typically supported learning through content delivery, drill-and-practice, multimedia input, or online communication. Contemporary AI, by contrast, increasingly functions as *educational infrastructure*: it shapes how tasks are designed, how feedback is generated, how progress is tracked, and how performance is evaluated. Systematic reviews and field mappings show that AI tools in language education now span automated writing feedback, chatbots, speech recognition, adaptive platforms, and analytics-driven instructional supports (Mohebbi, 2025; Wang et al., 2025; Zhang et al., 2025).

This infrastructural shift changes the nature of pedagogical work. When AI systems can generate examples, explanations, reformulations, comprehension questions, and simulated dialogue, the teacher's role is less about producing all input and more about *orchestrating learning conditions*—

deciding what AI should do, when, and under what constraints, and ensuring that learning goals, quality standards, and ethical boundaries are met.

1.5 Personalization and Adaptive Learning: Addressing the "One-Size-Fits-All" Constraint

A central reason AI matters now is its potential to challenge one of the most persistent limitations of conventional instruction: the "one-size-fits-all" structure. Traditional pedagogies—often constrained by time, curriculum pacing, and class size—struggle to address individual differences in proficiency, motivation, learning pace, and linguistic background. AI-powered systems, particularly those leveraging NLP and machine learning, are widely framed as enabling a more personalized language learning landscape (Sangkala & Mardonovna, 2024; Mohebbi, 2025).

Across the literature, personalization typically appears in three interrelated forms:

- *Tailored learning paths*: AI can support differentiated practice by generating tasks and texts at an appropriate level, targeting recurring errors, and adjusting difficulty over time.
- *Adaptive scaffolding:* AI-driven tutoring and conversational systems can act as responsive partners, offering prompts, rephrasing, and incremental support aligned with learner performance.
- *Accessibility and inclusivity:* AI-mediated supports (e.g., text-to-speech, speech-to-text, captioning, multimodal input) can reduce barriers for diverse learners and expand participation opportunities.

At the same time, the research cautions that personalization is pedagogically meaningful only when aligned with sound instructional principles. Without clear learning objectives and teacher guidance, personalization can become mere convenience—lots of output, but limited development.

The aggregate evidence supports this caution rather than undercutting it. The same multilevel meta-analysis cited above found effects concentrated unevenly by skill, strongest for vocabulary learning and weakest for speaking, within the same 46-study sample (Torres & Kahveci, 2025). A separate PRISMA-guided review of 22 studies in school-level EFL settings similarly found significant gains across reading, writing, listening, speaking, and vocabulary, but reported that digital-divide and teacher-readiness challenges persisted alongside those gains (Kundu & Bej, 2025). Personalization, that is, is empirically supported in aggregate and highly conditional in practice.

1.6 Efficiency and Immediate Feedback: More Practice, Faster Correction, Better Timing

Another major reason AI matters now is the provision of *immediate feedback*, which has long been a scarce resource in language classrooms. In large classes, teachers cannot continuously provide individualized corrective feedback on grammar, vocabulary, discourse organization, pronunciation, and fluency. AI tools—such as Automated Writing Evaluation (AWE), grammar and style assistants, and speech recognition systems—extend feedback frequency and reduce the delay between production and response.

Reviews of AI in spoken English learning emphasize that AI-powered tools can support pronunciation and fluency development by enabling repeated practice and timely correction, while also lowering anxiety through private rehearsal spaces (Xu & Ismail, 2025). Similarly, systematic reviews of AI-assisted learning highlight immediate feedback as a driver of engagement and self-regulation—provided that learners understand how to interpret feedback and avoid uncritical dependence on automated corrections (Mohebbi, 2025). In short, AI's importance is not just that it gives feedback, but that it changes the *timing* and *availability* of feedback—two variables that strongly shape learning trajectories.

This picture, too, is partial rather than uniform. A PRISMA-guided review of 15 studies found that AI-supported feedback and automated scoring measurably improved writing quality, scoring accuracy, and learner engagement—but found dialogic competence and direct teacher intervention conspicuously unimproved by the same tools (Alhusaiyan, 2024). The efficiency gain, in other words, is clearest for feedback that can be rendered as a label or a score, and far less certain for feedback that depends on negotiating meaning.

Faster correction can also come apart from deeper learning. In a randomized experiment with 117 university students, the group receiving ChatGPT feedback produced the highest-scoring essays of any condition, yet showed no corresponding advantage in knowledge gain or transfer to new tasks—a pattern the researchers attribute to metacognitive laziness, in which fast, competent feedback lets learners bypass the effortful self-monitoring that produces durable learning (Fan et al., 2024). Immediacy, that is, is not itself a guarantee of learning; it is a variable that can cut either way depending on how learners use it.

1.7 Redefining Teacher and Student Roles: The Teacher–AI–Student (TAS) Triad

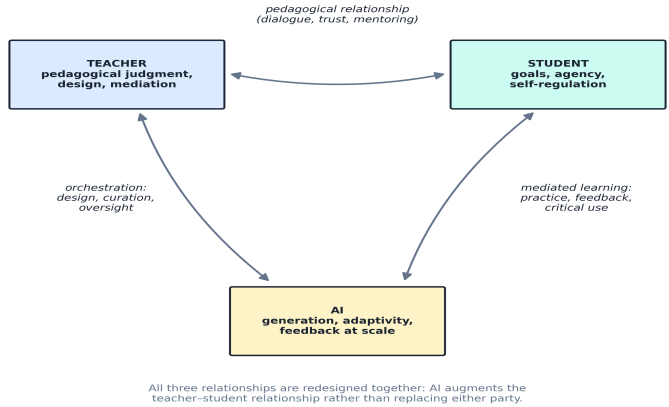


Figure 1. The Teacher–AI–Student (TAS) triad: three relationships redesigned together

The AI era forces a reconsideration of classroom roles and authority structures. Increasingly, AI is conceptualized not as a replacement for humans but as an augmentative partner within a *Teacher–AI–Student (TAS) triad* (Zhu & Li, 2025)**,** where pedagogical agency is distributed across participants.

- *Teachers* may be relieved from routine tasks (e.g., mechanical correction, repetitive drills, basic scoring), enabling greater focus on higher-order pedagogy: designing tasks, teaching strategy use, cultivating intercultural competence, and guiding critical language awareness. Research on the future of language teaching emphasizes the need for sustained professional development so teachers can integrate AI in ways that remain pedagogically coherent and ethically grounded (Hubbard & Schulze, 2025).
- *Students* shift from passive recipients to active agents who collaborate with AI tools, make decisions about AI support, and learn to evaluate AI output critically.
- *Institutions* face the responsibility of developing AI literacy policies and support systems—ensuring ethical use, academic integrity, and equitable access.

This triadic framing is crucial because it avoids two extremes: AI as a magical solution on one hand, or AI as a purely disruptive threat on the other. It instead emphasizes that outcomes depend on how AI is pedagogically positioned and governed.

1.8 Navigating Challenges: Ethics, Integrity, Bias, and Educational Values

The rapid adoption of AI also intensifies complex challenges. Concerns about data privacy, algorithmic bias, and academic dishonesty (including AI-assisted plagiarism) now occupy a central place in language education discussions. Research on teacher identity negotiation shows that AI adoption can produce professional uncertainty and "role tension," particularly when institutions adopt AI faster than teachers are supported to understand it (Zaman, Hussain, & Tabassam, 2024).

Equally significant is the pedagogical risk of overreliance: if learners outsource planning, drafting, revising, or reasoning to AI, they may produce fluent text without developing underlying competence. This is why many scholars increasingly argue for *AI literacy* as an educational priority—not only technical literacy, but the ability to evaluate outputs, recognize limitations, and use AI ethically and strategically (Li & Ding, 2025; Mateos-Blanco et al., 2024). Put simply, AI forces the field to ask: *What counts as learning? What counts as evidence? What values should language education protect?*

These are not speculative concerns. An umbrella review of 46 reviews on generative AI in education identifies ChatGPT as the most widely adopted tool for tasks including foreign-language learning and writing support, and finds that academic-dishonesty concerns run through the reviewed literature (Namoun et al., 2024). A SWOT analysis of ChatGPT in education reaches a similar conclusion from a different angle, naming over-reliance, bias, and privacy risk as the technology's principal threats even as it identifies personalized tutoring as its principal strength (Rane et al., 2024). Nor are learners uniformly credulous: in a survey of 443 undergraduate students, generally positive attitudes toward ChatGPT for academic use coexisted with explicit concern about accuracy, ethics, and dependency (Pallivathukal et al., 2024). A broader thematic review of 27 articles on generative AI in higher education reaches the same conclusion at the level of institutions, finding that benefits for personalized learning have not resolved parallel concerns about integrity, privacy, and equitable governance (Coman et al., 2026)—questions Chapters 11 and 12 of this book treat at length.

1.9 From CALL and MALL to AI-Mediated Language Education

The trajectory of technology in foreign language education is not a simple linear progression from older to newer tools, but rather a sequence of paradigm shifts that have continuously redefined the roles of teachers, learners, and technologies themselves. Across these shifts, technology has evolved from static delivery systems to mobile learning companions and, most recently, to intelligent, adaptive agents capable of mediating learning processes. The transition from Computer-Assisted Language Learning (CALL) and Mobile-Assisted Language Learning (MALL) to AI-mediated language education thus reflects a deeper epistemological and pedagogical transformation rather than mere technological innovation.

This periodization is not merely narrative convenience; it is traceable in the research record itself. A bibliometric analysis of AI-related research in L2 teaching and applied linguistics spanning 1995 to 2022 traces the field's development across exactly this period (Kartal & Yeşilyurt, 2024), while a longitudinal bibliometric analysis of 774 peer-reviewed publications between 2005 and 2025 maps the growth and thematic shifts of research on intelligent-agent-supported language learning, including chatbots and virtual tutors, across almost exactly the same decades (Liu, 2026). The eras this section describes, in short, leave a visible footprint in the literature itself, not only in retrospective argument.

1.9.1 The Foundational Era: Computer-Assisted Language Learning (CALL)

Computer-Assisted Language Learning (CALL) emerged as an interdisciplinary field at the intersection of applied linguistics, educational psychology, and computer science. Its development is commonly described through three major phases, each corresponding to dominant learning theories and technological affordances of its time (Bax, 2003).

Structural / Behavioristic CALL (1960s–1970s)

The earliest phase of CALL, often referred to as Computer-Assisted Instruction (CAI), was strongly influenced by behaviorist learning theory. Language learning was conceptualized as habit formation, and computers were used primarily to deliver repetitive drills, grammar exercises, and vocabulary practice with immediate feedback.

A prominent example is the PLATO (Programmed Logic for Automatic Teaching Operations) system, which provided self-paced language drills

and controlled practice activities. While limited in interactivity and flexibility, these early systems introduced two enduring principles: individualized pacing and automated feedback. However, learner agency and communicative use of language remained minimal, as the computer functioned largely as a tutor delivering predetermined content.

Communicative CALL (1980s–1990s)

As communicative language teaching gained prominence, CALL moved away from drill-based instruction toward more open-ended, meaning-focused activities. In this phase, computers were increasingly viewed as tools for stimulating interaction, hypothesis testing, and critical thinking rather than enforcing accuracy alone (Warschauer & Healey, 1998).

Simulations, text reconstruction programs, and problem-solving tasks allowed learners to generate original utterances and engage with language in more meaningful ways. Nevertheless, feedback mechanisms remained largely rule-based, and adaptation to individual learner needs was limited.

Integrative CALL and Normalization

The integrative phase of CALL sought to combine language skills, multimedia resources, and networked communication into coherent learning environments (Chapelle, 2001; Levy & Stockwell, 2006). This phase is closely associated with the concept of *normalization*, whereby technology becomes fully embedded in everyday teaching practice and fades into the background (Bax, 2003). Despite these advances, CALL systems still relied on predefined pathways and static content, restricting their capacity for deep personalization or real-time pedagogical decision-making.

1.9.2 The Mobility Shift: Mobile-Assisted Language Learning (MALL)

The emergence of mobile technologies marked a significant expansion of CALL's scope. Mobile-Assisted Language Learning (MALL) capitalized on smartphones, tablets, and wireless connectivity to enable learning beyond institutional settings. This shift supported the "anytime, anywhere" principle, allowing learners to exploit small pockets of time and space for language practice (Stockwell, 2022).

One of MALL's key contributions lies in its support for contextualized and situated learning. For instance, SMS-based systems such as MOLT (MObile Learning Tool) delivered vocabulary reminders at spaced intervals,

enhancing retention through principles of distributed practice. Similarly, QR-code–based activities in Japanese EFL contexts encouraged learners to interact physically with their surroundings, linking linguistic input to real-world contexts.

MALL strengthened learner autonomy, increased exposure to target languages, and blurred the boundary between formal and informal learning. However, despite its flexibility and motivational benefits, MALL largely preserved the instructional logic of CALL. Content remained standardized, and adaptation was limited, with pedagogical intelligence residing primarily in the teacher rather than the system itself.

1.9.3 The New Paradigm: AI-Mediated Language Education

The transition to AI-mediated language education represents a qualitative shift from technology as a passive repository or delivery mechanism to technology as an active "*thinking partner*." Artificial intelligence introduces systems capable of processing natural language, learning from learner data, and generating adaptive responses in real time.

This claim of novelty sits alongside a competing scholarly emphasis on continuity. One recent account frames ChatGPT as simply the latest in a lineage of 'teaching machines' reaching back to the behaviorist devices of the 1960s described earlier in this section, arguing that what has changed across six decades is not the aspiration to automate instruction but the sophistication with which a machine can now respond to language itself (Kohnke & Green-Eneix, 2025). Read this way, the paradigm shift is real, but it is a shift in degree of linguistic competence rather than in kind of pedagogical ambition.

Unlike CALL and MALL, which primarily supported instruction, AI-mediated environments participate in instructional decision-making. Platforms such as Duolingo employ machine learning algorithms to dynamically adjust task difficulty based on learner performance, exemplifying intricate customization driven by data analytics. This shift aligns closely with Cognitive Load Theory, as AI systems regulate task complexity to match learner capacity and prevent overload.

1.9.3.1 Generative AI as Tutor and Collaborator

The emergence of generative AI tools marks a further expansion of AI's pedagogical role. Tools such as ChatGPT function not only as information providers but also as collaborative tutors capable of simulating dialogue,

scaffolding writing, and supporting idea development. For example, advanced German learners have used ChatGPT to brainstorm and co-construct children's stories, positioning the AI as a writing partner rather than a mere correction tool (Lys, 2024).

Similarly, AI-powered chatbots such as Replika and TalkPal offer low-stakes environments for speaking practice, reducing anxiety and encouraging experimentation. These applications resonate strongly with constructivist learning theory, as learners actively construct knowledge through interaction with AI agents rather than passively consuming content (Czerkawski, 2024).

These illustrative cases reflect a broader pattern documented across the wider literature. A scoping review of 16 empirical studies found that ChatGPT consistently serves a dual role in EFL writing instruction—functioning simultaneously as a writing assistant and as an informal assessment tool—while noting that ethical questions and the need for task adaptation remain unresolved across the studies reviewed (Xiao et al., 2025). The tutor-and-collaborator role, that is, is not confined to isolated case studies; it recurs across the field, along with the concerns that accompany it.

1.9.3.2 Pedagogical and Theoretical Implications

AI-mediated language education is theoretically grounded in constructivism, sociocognitive SLA, and cognitive load theory. By adapting feedback, pacing, and task complexity, AI systems support individualized learning trajectories and cultivate learner autonomy. At the same time, research cautions against uncritical adoption, highlighting risks such as overreliance, surface learning, and diminished critical engagement (Liu, 2024).

A broader narrative review of AI technologies across second- and foreign-language education—spanning natural language processing, data-driven learning, automated writing evaluation, dynamic assessment, intelligent tutoring systems, automatic speech recognition, and chatbots—reaches a similar two-part conclusion: AI's integration into language education will only deepen, and the research documenting its effects has not yet caught up with the pace of its adoption (Son et al., 2025). That gap between adoption and evidence is a theme this book treats as a condition to be managed, not a problem to be wished away.

As a result, the teacher's role evolves from content provider to learning designer, mediator, and ethical gatekeeper. Rather than replacing educators, AI

amplifies the need for pedagogical expertise, critical digital literacy, and principled instructional design (Bengsch, 2024; Pan, 2024).

1.9.4 From Enhancement to Mediation

The progression from CALL and MALL to AI-mediated language education illustrates a shift from technological enhancement to pedagogical mediation. CALL introduced automation and individualized pacing; MALL expanded access and contextual learning; AI reconfigures the learning ecology itself by embedding intelligence, adaptability, and generativity within educational systems.

Rather than signaling the end of previous paradigms, AI builds upon their foundations, offering unprecedented opportunities for personalization, assessment, and learner engagement. The central challenge moving forward lies in ensuring that AI-mediated language education remains pedagogically grounded, ethically responsible, and fundamentally human-centered.

The experimental literature across this period bears out the trajectory just described. A systematic review of 30 studies using multiple-group experimental designs between 2014 and 2023 examined how chatbots and automated writing evaluation tools in language learning were designed, theoretically grounded, and evaluated (Qiao et al., 2025)—a decade of experimental work unfolding across exactly the MALL-to-generative-AI transition this section has traced. Whether that decade of experimentation has kept pace with the decade of adoption is a question this book returns to under each specific technology it examines.

Table 1. Three eras of technology-mediated language learning

Dimension	CALL (1960s–2000s)	MALL (2000s–2010s)	AI-mediated (2020s–)
Core device	Desktop computer, lab	Mobile phone, tablet	LLM-based systems across devices
Dominant logic	Program-controlled practice	Anytime-anywhere access	Generative, adaptive interaction
Content	Predefined, static	Standardized, portable	Generated, personalized on demand
Interaction	Learner–program	Learner–content–context	Learner–AI dialogue; multimodal

Dimension	CALL (1960s–2000s)	MALL (2000s–2010s)	AI-mediated (2020s–)
Feedback	Fixed, rule-based	Limited, often deferred	Immediate, adaptive, discourse-level
Teacher role	Selector of software	Facilitator of access	Designer, orchestrator, auditor
Key limit	Rigidity, isolation	Preserved CALL logic	Reliability, bias, overreliance

1.10 AI as a Paradigm Shift (LLMs, Multimodality)

The integration of Artificial Intelligence (AI) into educational contexts—particularly within English Language Teaching (ELT)—has emerged as one of the most profound transformations in the history of language education. Unlike earlier technological innovations that primarily functioned as supplementary tools, contemporary AI, and especially Generative Artificial Intelligence (GenAI), signals a *paradigmatic shift* in how languages are taught, learned, and assessed. This shift moves education away from linear, one-size-fits-all instructional models toward learning ecosystems that are *dynamic, adaptive, personalized, and data-driven*.

What distinguishes the current wave of AI from previous educational technologies is more than merely increased efficiency, but a fundamental reconfiguration of pedagogical roles, epistemological assumptions, and interactional patterns. This transformation is driven primarily by two interconnected developments: the emergence of *Large Language Models (LLMs)* and the expansion of *multimodal AI systems*, both of which redefine the relationships between learners, teachers, and machines within the language learning ecosystem.

1.10.1 The Engine of the Shift: Large Language Models (LLMs)

The evolution from early Computer-Assisted Language Learning (CALL) to Intelligent CALL (ICALL) represents a logical yet incomplete progression toward more adaptive systems. Early CALL applications and Intelligent Tutoring Systems (ITS) of the 1970s and 1980s were largely grounded in behaviorist learning theories, relying on repetitive drills, rule-based feedback, and predetermined learning paths. While innovative for their time, these systems lacked genuine linguistic understanding, contextual sensitivity, and the capacity for meaningful interaction.

Large Language Models represent a decisive break from these limitations. Trained on billions of human-generated texts and powered by deep learning architectures, LLMs are capable of interpreting, predicting, and generating language in ways that approximate human discourse. Models such as GPT-4 process linguistic input through hundreds of artificial neural layers, transforming tokens into high-dimensional vector representations that encode syntactic, semantic, and pragmatic relationships. This enables not only fluent text generation but also contextual reasoning, discourse coherence, and adaptive responsiveness.

In educational terms, this marks a shift from *reactive* to *generative* AI. Rather than merely responding to predefined inputs, LLMs can simulate dialogic interaction, scaffold learner output, and generate personalized learning trajectories. In language education, they function as *patient conversational partners*, formative feedback providers, and co-constructors of meaning. Research consistently shows that such systems expand opportunities for practice, reduce learner anxiety, and support fluency development beyond the temporal and spatial constraints of the classroom (Bonner et al., 2023; Wang et al., 2025; Wang, Zhang, Yao, & Seedhouse, 2025).

At a deeper level, LLMs challenge traditional distinctions between input and output, learning and assessment, and even learner and tutor. Automated feedback, idea generation, error diagnosis, and metalinguistic explanation—once the exclusive domain of teachers—are increasingly mediated by AI systems. This does not eliminate the teacher's role, but rather transforms it, demanding new forms of pedagogical design, orchestration, and ethical judgment.

1.10.2 Expanding the Horizon: Multimodality

The current paradigm shift extends beyond text-based interaction through the rise of *multimodal AI*. While early LLMs were primarily text-oriented, contemporary GenAI systems integrate multiple modalities, including speech, images, video, gesture, and even three-dimensional environments. For this reason, many current systems are more accurately described as *multimodal LLMs*.

The convergence of multimodal learning and Natural Language Processing (NLP) opens fundamentally new directions for ELT:

- *Immersive Environments*: Virtual Reality (VR) and Augmented Reality (AR) enable learners to participate in authentic, situated language use,

simulating real-world communicative contexts. Such environments have been shown to reduce affective barriers, enhance motivation, and promote experiential learning (Wu, 2024).

- *Digital Humans and Conversational Agents*: Platforms such as *D-ID Creative Reality Studio* combine face animation, speech synthesis, and LLM-driven dialogue to create embodied conversational agents that function as 24/7 language partners. These agents extend opportunities for interaction while supporting pronunciation, pragmatics, and discourse competence (Ayedoun et al., 2019; Wang & Li, 2025).
- *Multisensory Feedback*: By engaging visual, auditory, and kinesthetic channels simultaneously, multimodal AI systems offer feedback that more closely mirrors naturalistic language acquisition. This holistic engagement supports deeper processing and retention, particularly in speaking and listening development.

Multimodality thus aligns language education with contemporary understandings of communicative competence as inherently multimodal, contextual, and embodied. Language learning is no longer confined to text manipulation but becomes an integrated performance across semiotic resources.

1.10.3 Assessment, Autonomy, and Ethical Considerations

One of the most consequential implications of AI as a paradigm shift lies in assessment. AI enables continuous, formative, and diagnostic assessment embedded within learning activities rather than isolated testing events. However, this also raises critical concerns regarding validity, transparency, data privacy, and over-reliance on automated judgment (Zhang et al., 2025).

At the same time, when pedagogically scaffolded, AI has demonstrated strong potential to promote *learner autonomy and self-regulation.* Systematic reviews indicate that AI-supported environments promote independent practice, metacognitive awareness, and personalized goal setting—provided that learners receive explicit guidance and ethical framing (Sangkala & Mardonovna, 2024; Sayici & Aydin, 2025).

1.10.4 Toward a Hybrid and Sustainable AI Paradigm

The literature converges on the conclusion that the future of language education lies not in fully automated systems, but in *hybrid pedagogical models* that integrate AI affordances with human judgment, cultural sensitivity, and pedagogical intentionality. Ethical concerns—algorithmic bias, unequal access, misuse, and cognitive offloading—must be addressed proactively if AI is to support inclusive and sustainable language education.

In sum, AI constitutes a paradigm shift in foreign language education by transforming roles, expanding modalities of communication, and reshaping learning and assessment practices. Large language models and multimodal AI systems move ELT toward more personalized, interactive, and learner-centered paradigms, while simultaneously demanding new forms of professional competence, institutional policy, and ethical responsibility.

1.11 Scope of the Book: FL/L2 Teaching, Learning, and Assessment

The integration of Artificial Intelligence (AI) into foreign language (FL) and second language (L2) education represents a multifaceted transformation that affects every stage of the pedagogical process. Rather than functioning as a discrete technological add-on, AI increasingly operates as a *mediating infrastructure* that reshapes instructional practices, learner engagement, and assessment paradigms. This book, *Artificial Intelligence in Foreign Language Teaching, Learning, and Assessment*, explores how AI technologies bridge traditional methodologies with innovative, data-driven approaches, mapping a new educational landscape in which *human and machine intelligences function as collaborative partners.*

The scope of the book is deliberately holistic. Teaching, learning, and assessment are not treated as isolated domains but as *interdependent dimensions* of an AI-mediated language education ecosystem. Drawing on recent empirical studies, systematic reviews, and bibliometric analyzes, the book examines how AI redefines pedagogical roles, learner agency, and evaluative practices across diverse educational contexts (Wang, Zhang, Yao, & Seedhouse, 2025; Zhang, Hussin, & Abd Majid, 2025).

1.11.1 Scope in Foreign Language Teaching

In the domain of FL/L2 teaching, AI markedly reconfigures the instructor's professional role. Teachers are no longer positioned primarily as knowledge transmitters, but increasingly as *facilitators, cognitive engineers, cultural mediators, and techno-pedagogical designers*. The book examines how AI supports this transition while also introducing new responsibilities and challenges.

1. Instructional Design and Preparation

AI-powered systems enable the rapid generation of customized instructional materials, including lesson plans, discussion prompts, assessments, and interactive tasks aligned with specific curricular goals or learner profiles. Research shows that such tools can enhance instructional efficiency and creativity when used as design supports rather than replacements for pedagogical judgment (Bonner, Lege, & Frazier, 2023; Kovalenko & Baranivska, 2024).

2. The Teacher–AI–Student (TAS) Triad

A central conceptual focus of the book is the *Teacher–AI–Student (TAS) triad*, in which AI acts as an active mediating agent rather than a passive tool. Within this ecosystem, teachers leverage AI-generated learning analytics to design collaborative and task-based activities, while AI systems manage routine linguistic scaffolding and administrative tasks (Zhu & Li, 2025). The book critically examines how this triadic relationship reshapes authority, interaction, and responsibility in the language classroom.

3. Teacher Readiness and Professional Development

The book also addresses the growing need for *AI literacy and sustained professional development* among language educators. Studies consistently highlight gaps in teacher preparedness, ethical confidence, and institutional support, highlighting the importance of developing competencies related to critical evaluation of AI output, pedagogical integration, and responsible use (Hubbard & Schulze, 2025; Sayici & Aydin, 2025).

1.11.2 Scope in Foreign Language Learning

From a learning perspective, AI enables the creation of *personalized, adaptive, and learner-centered environments* that respond dynamically to individual needs, proficiency levels, and learning trajectories. The book examines AI-mediated learning through cognitive, interactionist, and sociocultural lenses, emphasizing both opportunities and risks.

Skill-Specific Language Development

The scope of the book covers AI applications across all language skills:

- *Reading and Listening*: Adaptive platforms adjust input complexity and provide multimodal support through synchronized text, audio, and visual cues, enhancing comprehension and engagement (Fitria, 2021).
- *Writing*: Automated Writing Evaluation (AWE) systems and generative AI assistants support brainstorming, revision, grammatical accuracy, and

stylistic refinement, while raising questions about authorship and originality (Lai, 2025).

- *Speaking and Pronunciation*: Speech recognition technologies and conversational agents (e.g., Apilingua App, ELSA Speak, ChatGPT-based tutors) provide low-anxiety, real-time oral practice and feedback, contributing to fluency and pronunciation development (Karim et al., 2023).

Affective and Motivational Dimensions

Beyond skill development, the book explores AI's impact on affective factors such as anxiety, confidence, and motivation. Empirical studies suggest that AI-mediated environments can create psychologically safe spaces for rehearsal and experimentation, particularly for learners reluctant to participate in traditional classroom interaction (Sangkala & Mardonovna, 2024).

Learner Autonomy and Self-Regulation

A key theme across chapters is *learner autonomy*. AI enables continuous, self-directed learning through 24/7 access to resources, personalized feedback, and adaptive challenges. However, the book also critically examines the risk of overdependence and superficial engagement, arguing that autonomy must be pedagogically scaffolded and accompanied by explicit AI literacy development (Li & Ding, 2025).

1.11.3 Scope in Assessment and Evaluation

Assessment represents one of the most transformative—and contested—areas of AI integration in FL/L2 education. The book addresses how AI shifts assessment from static, episodic measurement toward *dynamic, longitudinal, and formative evaluation*.

Automated Feedback and Scoring

AI systems increasingly provide instantaneous, targeted, and ostensibly impartial feedback on linguistic performance, enabling learners to reflect on errors and revise output in real time. Such systems have the potential to enhance learning efficiency when embedded within pedagogically meaningful tasks (Weng & Chiu, 2023).

Holistic Diagnostics and Learning Analytics

AI's capacity to analyze large datasets allows for fine-grained diagnostic insights, such as identifying persistent grammatical transfer patterns or pronunciation difficulties linked to learners' first languages. The book explores how these analytics can inform instruction while also raising concerns about transparency and interpretability (Wang et al., 2025).

Academic Integrity and Ethical Oversight

The book critically examines challenges related to AI-assisted plagiarism, authorship ambiguity, and assessment fairness. It discusses emerging rubric designs and hybrid evaluation models aimed at distinguishing human originality from machine-generated output, emphasizing the need for ethical frameworks and institutional policy development (Zaman, Hussain, & Tabassam, 2024).

1.12 Defining the Conceptual Boundaries of the Book

While comprehensive in scope, this book does not aim to function as a technical manual for specific AI tools. Instead, it adopts a *conceptual, pedagogical, and analytical orientation*, equipping researchers, educators, and policymakers with frameworks to critically understand, evaluate, and responsibly implement AI in FL/L2 teaching, learning, and assessment.

The integration of AI into language education can be likened to replacing a traditional flashlight with a *high-definition, 360-degree radar system*: where the flashlight illuminates only a fixed, user-directed point, the radar continuously maps the entire learning terrain, identifies hidden obstacles, and suggests adaptive pathways for each learner in real time. Whether this radar guides learners wisely or overwhelms them depends not on the technology itself, but on the *human intentionality* that shapes its use.

Pedagogical Companion for Chapter 1

Key Concepts: AI-mediated language education; personalization and adaptive learning; Teacher–AI–Student (TAS) triad; educational infrastructure; CALL and MALL; paradigm shift; augmentation versus automation; the PAGE framework.

Discussion Questions

1. The chapter argues that AI has moved from educational *technology* to educational *infrastructure*. What changes pedagogically when a tool becomes infrastructure—and what becomes harder to opt out of?

2. Consider the TAS triad from your own context: which of the three relationships (teacher–AI, student–AI, teacher–student) is currently changing fastest, and with what consequences?
3. The historical sequence CALL → MALL → AI-mediated education is presented as continuity rather than rupture. Which pedagogical problems have remained constant across all three eras?
4. "One-size-fits-all" instruction is identified as the constraint AI most directly addresses. What are the limits of personalization as an educational ideal?
5. What would count as evidence, in your setting, that AI adoption is augmenting rather than automating teachers' work?

Further Reading

- Warschauer, M., & Healey, D. (1998). Computers and language learning: An overview. *Language Teaching, 31*(2).
- Zhu, M., & Li, X. (2025). The Teacher–AI–Student triad. *English Language Teaching, 18*(7).
- Kasneci, E., et al. (2023). ChatGPT for good? *Learning and Individual Differences, 103*.

Chapter 2. Understanding Artificial Intelligence: Concepts, Technologies, and System Archetypes

2.1 Artificial Intelligence and the New Educational Revolution

The integration of Artificial Intelligence (AI) into foreign language teaching, learning, and assessment signals the onset of a new educational revolution—one that is often compared to earlier paradigm shifts such as the rise of print literacy, mass schooling, and digital learning technologies. Although the term *artificial intelligence* was formally introduced during the Dartmouth Summer Research Project in 1956, its contemporary educational significance extends far beyond its original conceptualization as machine-based problem solving (Kannan & Munday, 2018).

What began as a speculative inquiry into whether machines could simulate aspects of human reasoning has evolved into a globally embedded educational infrastructure. AI now operates at multiple levels of educational systems, from institutional analytics and automated assessment to individualized learner feedback and generative content creation. In this sense, AI is no longer an optional add-on to educational practice; rather, it functions as a structural layer that reshapes how knowledge is produced, mediated, and evaluated within formal learning environments (Tan et al., 2025).

In the context of English Language Teaching (ELT) and foreign language education more broadly, AI marks a qualitative shift rather than a simple technological upgrade. Early forms of Computer-Assisted Language Learning (CALL) primarily digitized existing pedagogical practices—grammar drills, vocabulary lists, and rule-based feedback. Contemporary AI-driven systems, by contrast, simulate selected aspects of human intelligence to support adaptive learning, dynamic interaction, and personalized instructional trajectories (Ahimsa & Khawa, 2025). This transformation has redefined the boundaries between teacher, learner, and technology, introducing new forms of agency and responsibility within the learning process.

At the same time, the rapid diffusion of AI has generated conceptual ambiguity. In both scholarly and practitioner discourse, terms such as *AI*, *machine learning*, *natural language processing*, *large language models*, and *generative AI* are frequently used interchangeably, obscuring important technological and pedagogical distinctions (Agyare et al., 2025; Masry-

Herzallah & Makaldy, 2025). Clarifying these distinctions is essential, not only for theoretical precision but also for the ethical and pedagogically sound integration of AI into foreign language education.

2.2 Defining Artificial Intelligence in Foreign Language Education

Artificial Intelligence in foreign language education can be defined as a constellation of computational systems designed to process linguistic input, model learner behavior, generate language output, and adapt instructional pathways in ways that approximate selected aspects of human cognition (Dong et al., 2025). Crucially, AI does not possess understanding, intentionality, or consciousness. Its apparent intelligence emerges from statistical pattern recognition and probabilistic inference rather than semantic comprehension (Bonner et al., 2023).

This distinction has important pedagogical implications. Treating AI as a thinking subject risks attributing epistemic authority to systems that fundamentally lack contextual awareness and moral judgment. As a result, contemporary educational research increasingly frames AI not as an autonomous intelligence but as a *pedagogical infrastructure*—a mediating system embedded within curricular aims, institutional policies, assessment cultures, and sociocultural contexts (Hubbard & Schulze, 2025).

From this infrastructural perspective, the educational value of AI does not reside solely in its technical capabilities. Instead, it depends on how effectively AI systems are aligned with learning theories, instructional design principles, and ethical commitments. In ELT, this alignment is particularly critical, as language learning is inherently social, contextual, and identity-sensitive. AI systems may facilitate practice, feedback, and personalization, but they cannot replace the interpretive, affective, and relational dimensions of human teaching.

AI-driven applications in language education now encompass automated writing feedback, speech recognition-based pronunciation training, adaptive vocabulary systems, formative assessment tools, and conversational agents. Collectively, these tools reflect a shift from static digital resources toward more dynamic adaptive, data-informed learning environments that continuously respond to learner input (Kohnke et al., 2025). This shift raises fundamental questions about authorship, assessment validity, learner autonomy, and the evolving role of teachers in AI-mediated classrooms.

2.3 Core Technologies Underpinning AI in Language Education

2.3.1 Artificial Intelligence and Machine Learning

Artificial Intelligence represents the overarching scientific and technological field concerned with designing systems capable of performing tasks traditionally associated with human intelligence, such as learning, reasoning, and decision-making (Agyare et al., 2025). Within this broader field, *Machine Learning (ML)* functions as the primary operational engine that enables AI systems to improve performance over time.

Unlike rule-based systems that depend on explicitly programmed instructions, ML algorithms learn from data. By identifying patterns across large datasets, ML systems generate predictive models that can be applied to new, unseen data points (Manorat et al., 2025). In educational contexts, this capacity enables systems to analyze learner behavior at scale, uncovering patterns that would be difficult or impossible for human instructors to detect in real time.

In language education, ML underpins learning analytics, adaptive task sequencing, automated scoring, and personalized feedback mechanisms. For example, ML-driven systems can track learner error patterns in writing or speech, adjust task difficulty dynamically, and recommend targeted instructional resources aligned with individual proficiency levels (Kovalenko & Baranivska, 2024). This data-driven adaptability represents a decisive departure from earlier CALL tools, which relied on static content and predetermined feedback pathways.

2.3.2 Natural Language Processing

Natural Language Processing (NLP) is the subfield of AI concerned with enabling machines to process, interpret, and generate human language (Abrusci et al., 2025). Historically, NLP evolved from symbolic, rule-based approaches that relied on manually encoded grammars and lexicons. While these systems offered transparency, they struggled to cope with linguistic ambiguity, variability, and authentic usage (Kannan & Munday, 2018).

The transition to statistical and neural NLP marked a major turning point. Contemporary NLP systems are trained on massive corpora of authentic language data, allowing them to model language probabilistically rather than deterministically. In language education, NLP underpins automated

essay scoring, grammar and coherence analysis, speech recognition, and real-time translation tools (Chen et al., 2025).

These technologies enable scalable interaction with learner language, providing immediate feedback that would otherwise be impractical in large classes or resource-constrained educational settings. However, their reliance on probabilistic modeling also introduces limitations, particularly in handling pragmatic nuance, intercultural appropriateness, and context-dependent meaning—dimensions that remain central to communicative competence.

2.3.3 Large Language Models

Large Language Models (LLMs) constitute one of the most consequential developments within contemporary Natural Language Processing (NLP), marking a qualitative shift in how machines generate and interact with human language. Architecturally, LLMs are deep neural networks—most commonly based on transformer frameworks—that are trained on massive corpora comprising billions of textual units drawn from diverse genres, registers, and communicative contexts. Through this large-scale training, LLMs learn statistical regularities in language use and operate as *autoregressive systems*, predicting each subsequent word based on probabilistic relationships within preceding context rather than semantic understanding (Bonner et al., 2023; Li et al., 2025).

A defining feature of LLMs is their general-purpose nature. Unlike earlier NLP systems that required task-specific training or rule-based programming, LLMs can perform a wide range of linguistic tasks—including summarization, translation, paraphrasing, argument construction, and extended discourse generation—without explicit retraining for each function. Models such as GPT-4 exemplify this flexibility, demonstrating performance across academic, professional, and everyday communicative tasks with minimal prompt-based guidance (de Winter, 2024). This capability has contributed substantially to the rapid diffusion of LLMs into educational settings, where versatility and scalability are highly valued.

In language education, the emergence of LLMs challenges long-standing assumptions about authorship, originality, and the nature of linguistic competence. Traditionally, language production in educational contexts has been viewed as a direct reflection of a learner's internalized linguistic

knowledge. LLMs complicate this relationship by enabling learners to co-construct texts with AI systems that can generate fluent, stylistically appropriate, and structurally coherent language. As a result, LLMs increasingly function as *cognitive scaffolds* that support drafting, revising, modeling genre conventions, and experimenting with alternative formulations. When used pedagogically, they can lower affective barriers, provide immediate exemplars, and extend learners' exposure to varied linguistic input.

At the same time, the pedagogical affordances of LLMs are inseparable from their limitations. Despite their surface-level fluency, LLM outputs are fundamentally probabilistic rather than epistemic. They do not possess awareness of truth, intention, or communicative purpose; instead, they generate language by estimating the most statistically likely continuation of a given input. This characteristic introduces well-documented risks in educational contexts, including hallucinated information, fabricated references, overgeneralized feedback, and discourse that appears coherent but lacks conceptual depth (Ibrahim, 2025; Nazaretsky et al., 2025).

The scale of these risks shows up in studies that test LLM output against learner or expert benchmarks rather than assessing it impressionistically. When 40 Korean EFL pre-service teachers used ChatGPT and EditGPT as revision aids, the tools corrected 86% of errors and improved lexical sophistication, yet word-choice and structural errors persisted even after AI-assisted revision (Chon & Shin, 2026)—fluency gains and error-free output are not the same thing. Reliability is uneven at the level of judgment as well as production: when 5 expert linguists rated ChatGPT-4.0's responses to 48 second-language-acquisition questions, most responses were judged good to very good, yet inter-rater agreement among the 5 experts was close to zero (Fleiss's kappa = 0.006), meaning trained evaluators could not consistently agree on what counted as an adequate answer (Alhazmi & Muftah, 2025).

From a pedagogical perspective, these limitations have profound implications. If LLM-generated output is treated as authoritative or inherently reliable, learners may defer critical judgment, leading to epistemic complacency and reduced metalinguistic awareness. Overreliance on LLMs may also obscure learners' actual proficiency, complicate assessment validity, and weaken opportunities for productive struggle—an essential component of language development. In turn, the educational value of LLMs depends

not on their autonomous use, but on how they are framed, mediated, and critically integrated by educators.

In this sense, LLMs are best understood not as replacements for human expertise, but as *probabilistic language simulators* whose outputs require interpretation, verification, and contextualization. When embedded within human-in-the-loop pedagogical models, LLMs can enhance language learning by supporting exploration, reflection, and iterative improvement. When used uncritically, however, they risk distorting learners' understanding of language, knowledge, and authorship. Recognizing this duality is essential for the responsible and pedagogically sound integration of LLMs into language education.

2.3.4 Generative Artificial Intelligence

Generative Artificial Intelligence (GenAI) represents the most recent and disruptive phase in the evolution of AI technologies applied to education. Unlike earlier AI systems that primarily focused on classification, prediction, or analysis of existing data, GenAI systems are designed to *produce new content*. This includes not only text, but also images, audio, video, code, and multimodal artifacts generated in response to complex prompts (Belkina et al., 2025). Technologically, GenAI systems rely on deep learning architectures—often large-scale neural networks—that model patterns within vast datasets and recombine these patterns to create outputs that appear novel, coherent, and contextually relevant. In addition, Linden et al. (2025), posit that AI can support human creativity without diminishing human originality, thus they theorize what one account calls "*Homo Promptus*"—the proposition that generative AI enables new forms of human-computer co-remembering, through which externalizing memory and ideation onto an AI system does not simply replace human cognitive work but potentially reshapes how human memory and creativity function over time.

In the context of English Language Teaching (ELT), GenAI has introduced capabilities that fundamentally alter instructional design and pedagogical practice. GenAI tools can generate essays, lesson plans, dialogues, role-play scenarios, assessment items, rubrics, and personalized learning prompts within seconds. This capacity enables unprecedented levels of customization and scalability, allowing instructional materials to be adapted rapidly to learners' proficiency levels, interests, and learning goals (Cheah et al., 2025; Choi et al., 2025). As a result, GenAI shifts the teacher's role

from primary content producer to pedagogical designer, curator, and evaluator of AI-generated materials.

Concrete implementations bear this out. In a course for 46 Japanese higher-education students that combined machine translation with generative AI tools for English learning, most participants reported high satisfaction, usefulness for English learning, and sustained motivation across the course (Oda & Saito, 2025)—evidence that GenAI's instructional value shows up most clearly when it is built into a designed course sequence rather than handed to learners as a standalone tool.

From a pedagogical perspective, GenAI expands the creative affordances of language classrooms. Learners can use GenAI to experiment with genre conventions, explore alternative formulations, simulate communicative scenarios, and engage in iterative drafting processes. When strategically integrated, GenAI can support exploratory learning, reduce affective barriers, and provide immediate linguistic input that extends beyond what is feasible through human instruction alone. These affordances align with learner-centered and constructivist approaches that emphasize agency, experimentation, and personalized pathways.

However, the generative nature of these systems also introduces significant pedagogical and epistemological challenges. Because GenAI produces content through probabilistic pattern generation rather than conceptual understanding, its outputs may appear fluent and persuasive while containing inaccuracies, oversimplifications, or fabricated information. Beyond this, when learners rely heavily on AI-generated text, there is a risk that cognitive effort shifts from language construction to surface-level editing, potentially diminishing opportunities for metacognitive engagement and productive struggle—processes that are central to durable language learning (Varsamidou, 2024).

This tension is visible across the wider evidence base, not only in isolated cases. A systematic review of 38 studies on generative AI and learning found that most reported positive effects on self-regulated learning (71.4% of studies) and critical thinking (62.5%), while the same review cautioned against the risk of overreliance (Sardi et al., 2025). The pattern this suggests is not a simple trade-off between benefit and risk but a conditional one: whether GenAI displaces or develops metacognitive engagement plausibly depends on how a task is designed around it, not on the technology by itself.

Uncritical adoption of GenAI may also obscure the boundaries between learner-generated and AI-generated language, complicating issues of authorship, originality, and assessment validity. If GenAI outputs are treated as neutral or authoritative, learners may develop an inflated sense of accuracy and competence, while teachers may struggle to interpret learner performance reliably. These risks highlight the importance of reframing GenAI not as an autonomous instructional agent, but as a mediated resource whose educational value depends on pedagogical intent and human oversight.

For this reason, effective use of GenAI in ELT requires *intentional instructional design*. Research emphasizes the need for structured tasks that clearly define the role of AI within the learning process, transparent framing that makes AI involvement explicit, and explicit instruction in critical AI literacy (Varsamidou, 2024). Learners must be guided to evaluate AI-generated output critically, interrogate its accuracy and appropriateness, and reflect on how it supports—or constrains—their own language development. In this way, GenAI can function as a catalyst for higher-order thinking rather than a shortcut that bypasses learning.

Consequently, GenAI's pedagogical significance lies not in its capacity to generate language, but in how it reshapes relationships between creativity, cognition, and instruction. When embedded within human-in-the-loop frameworks that prioritize reflection, verification, and learner agency, GenAI can enrich language education. When deployed without critical mediation, however, it risks reducing language learning to consumption rather than construction. Recognizing and managing this tension is essential for the responsible integration of Generative Artificial Intelligence into ELT.

2.3.5 How These Systems Fail: Hallucination, Brittleness, and Uneven Performance Across Languages

The pedagogical risks already noted in this chapter—hallucinated information, overgeneralized feedback, inconsistent judgments—are not incidental defects that better prompting will remove; they follow from how these systems are built, as this chapter has already argued. What follows from that architecture is brittleness: sharp, unpredictable drops in output quality once inputs move even slightly outside the statistical patterns a model was trained on. In language-education tools specifically, this brittleness has a documented shape. A systematic review of automated essay

scoring research found that most existing systems fail to adequately evaluate content relevance and discourse coherence (Ramesh & Sanampudi, 2021), which are precisely the dimensions of writing quality that require tracking meaning across a whole text rather than scoring local, surface features.

This unevenness is not only a matter of topic or genre; it tracks language itself, a fact with direct consequence for a readership teaching languages other than English. In Faruna (2026) benchmark testing of Gemini 2.5 Flash found refusal rates near 90% for English-language prompts but only 35% to 55% for the same category of prompts in Yoruba, Hausa, Igbo, and Igala, a pattern the study describes as severe cross-lingual safety degradation, for prompts in these languages, the same model behaves as though its safety training barely applies. The gap is not confined to safety. Generative AI tools evaluated on Bengali grammatical error correction competently flagged errors but largely failed to produce natural-language explanations a learner could actually use (Maity et al., 2024). Catching an error and explaining it are different operations, and the second may be the more fragile one when the language itself is lower-resourced.

The scale of the underlying gap is visible in infrastructure work. The most inclusive current language-identification models reliably cover only a fraction of the world's languages, with the best-performing system extending coverage to 1,665 languages (Kargaran et al., 2023), and a synthesis of natural language processing research on Armenian—a language with a long history—still describes it as persistently low-resource, with clear gaps in current large-language-model applicability (Avetisyan & Broneske, 2023). Nor is the gap fixed in place: cross-lingual retrieval-augmented prompting has raised zero-shot performance across 10 low-resource languages from 6 language families by 5.1% to 16.3%, depending on whether labeled examples were available (Nie et al., 2023), and a multilingual sentiment lexicon spanning French, English, Afrikaans, Sepedi, and Zulu reached 99% precision using a transformer model with explainable AI (Nkongolo, 2025). Targeted engineering, in other words, can close much of the gap for a given language—but closing it is an additional, language-specific undertaking, not something general-purpose model scale delivers on its own.

For a readership teaching across this spectrum of resourcing, the implication is direct: a tool's demonstrated competence in English-medium

contexts licenses no assumption about its competence in the language actually being taught. The unevenness documented here is one reason the human-in-the-loop principle argued for later in this chapter is not a generic caution but a calibrated one—the less resourced the language, the less warranted is any default trust in unsupervised system output, a point Chapter 12 returns to in discussing equity and multilingual futures.

2.4 AI System Archetypes in Language Education

The practical implementation of Artificial Intelligence in foreign language education is not monolithic; rather, it manifests through a set of distinct *system archetypes*, each designed to serve specific pedagogical purposes and to intervene at different stages of the learning process. These archetypes represent recurring design logics in AI-enhanced educational technologies and reflect how AI capabilities are operationalized to support interaction, instruction, feedback, and personalization. Understanding these system types is essential for educators and researchers seeking to move beyond tool-oriented adoption toward principled pedagogical integration.

One of the most visible and widely adopted archetypes in language education is the *AI chatbot*. Chatbots function as conversational interfaces that allow learners to engage in dialogue using natural language. In ELT contexts, they are frequently employed as virtual interlocutors that simulate communicative interaction, offering learners opportunities to practice speaking or writing in low-anxiety environments (Agyare et al., 2025). Unlike human interlocutors, chatbots are available on demand, non-judgmental, and infinitely patient, qualities that can reduce affective barriers such as speaking anxiety and fear of error. However, while chatbots are effective in promoting fluency and engagement, their limitations in pragmatic awareness, cultural nuance, and sustained discourse coherence necessitate pedagogical mediation to ensure communicative appropriateness and depth.

A second major archetype is the *Intelligent Tutoring System (ITS)*. ITS are structured AI-driven instructional environments that provide personalized guidance based on formal learner modeling and domain expertise. Typically, an ITS comprises three interrelated components: a learner model that tracks the learner's knowledge state and progress, a domain model that represents the target linguistic content, and a pedagogical model that

determines instructional strategies and feedback (Jung et al., 2025). In language education, ITS are particularly effective for grammar instruction, vocabulary development, and exam preparation, where clearly defined learning objectives and incremental progression are essential. Their strength lies in adaptive scaffolding—adjusting task difficulty, feedback timing, and instructional pacing in response to learner performance. Nevertheless, their structured nature may limit opportunities for open-ended language use unless complemented by communicative activities.

Beyond task-oriented instruction, *pedagogical agents* introduce a more social and affective dimension to AI-mediated learning. These agents—sometimes embodied as avatars or animated characters—are designed to guide learners through instructional content while nurturing engagement, motivation, and metacognitive awareness. Pedagogical agents often employ conversational strategies such as prompting reflection, encouraging self-explanation, and summarizing progress, thereby supporting learners' regulation of their own learning processes. In language education, such agents can help bridge the gap between cognitive instruction and affective support, although their effectiveness depends heavily on thoughtful design and alignment with pedagogical goals.

A fourth archetype is the *recommender system*, which leverages machine learning and learning analytics to personalize instructional pathways. Recommender systems analyze large volumes of learner data—such as engagement patterns, assessment results, and interaction histories—to suggest learning resources, tasks, or sequences tailored to individual needs (Wang et al., 2025). In ELT contexts, these systems can mitigate information overload by guiding learners toward appropriate materials and supporting differentiated instruction at scale. However, their reliance on data-driven prediction raises important questions about transparency, learner autonomy, and the risk of reinforcing narrow learning trajectories if recommendations are not pedagogically supervised.

Each archetype above is already instantiated in systems documented in the literature, not only in design proposals. AutoTutor exemplifies the ITS archetype: an intelligent tutoring system built around mixed-initiative dialogue rather than static question-and-answer sequencing (Graesser et al., 2005). At the level of pedagogical agents, Ayedoun et al. (2019) added communicative and affective strategies to an embodied conversational

agent designed to sustain second-language learners' willingness to communicate, illustrating how such agents combine instructional guidance with affective scaffolding rather than treating the two as separate functions. The chatbot archetype has likewise moved beyond single-agent dialogue: Aleedy et al. (2025) proposed a multi-agent chatbot architecture in which coordinated AI agents jointly support language-learning tasks. And the underlying logic of recommender functionality has been demonstrated, if not yet for a language classroom specifically, by Junaidi and Wahyono (2025), whose system generates competency recommendations from learners' attendance patterns and academic performance records—the same data-driven personalization principle an ELT-specific recommender would need to operationalize.

Taken together, these archetypes illustrate the functional diversity of AI applications in language education. More importantly, they highlight that the educational impact of AI does not derive from the technology itself, but from how specific system types are aligned with pedagogical intentions, learning theories, and contextual constraints. A chatbot used without instructional framing may result in superficial interaction, just as an ITS deployed without communicative integration may reduce language learning to rule acquisition. It follows that effective AI integration requires educators to select and combine system archetypes strategically, ensuring that technological affordances serve pedagogical aims rather than dictating them.

In this sense, the classification of AI system archetypes serves as a conceptual tool for resisting technological determinism. By foregrounding pedagogical function over technological novelty, educators and researchers can evaluate AI tools not in terms of what they *can* do, but in terms of what they *should* do within coherent, human-centered language learning environments.

2.5 Weak AI and Strong AI: Educational Relevance

The philosophical distinction between *Weak (Narrow) Artificial Intelligence* and *Strong (General) Artificial Intelligence* remains foundational for understanding the capabilities, limitations, and appropriate pedagogical roles of AI in language education. Originating in early debates within

cognitive science and philosophy of mind, this distinction provides a critical conceptual framework for evaluating what contemporary AI systems can—and crucially cannot—do within educational settings.

Weak AI refers to systems designed to perform specific, bounded tasks by optimizing performance within narrowly defined domains. All AI technologies currently deployed in educational contexts—including automated grading systems, intelligent tutoring platforms, speech recognition tools, and large language models—fall within this category. These systems excel at pattern recognition, probabilistic prediction, and data-driven optimization, enabling them to simulate aspects of intelligent behavior without possessing consciousness, intentionality, or genuine understanding (Herzallah & Makaldy, 2025). In language education, Weak AI has proven particularly effective in tasks such as providing surface-level feedback on grammar and coherence, identifying pronunciation deviations, generating practice materials, and supporting adaptive learning pathways.

By contrast, *Strong AI*—also referred to as Artificial General Intelligence—denotes a hypothetical form of machine intelligence capable of human-like cognition across domains, including abstract reasoning, self-awareness, contextual understanding, and moral judgment. Such systems would, in principle, be able to transfer knowledge flexibly across contexts, reflect on their own outputs, and engage in genuinely meaningful communication. Despite frequent popular discourse suggesting imminent breakthroughs, no existing AI system approaches this level of cognitive generality. The distinction is therefore not merely theoretical but serves as a necessary corrective to inflated claims regarding AI's "intelligence" in educational discourse.

The emergence of Generative AI and Large Language Models has, however, complicated this binary distinction. Models such as GPT-4 exhibit a degree of cross-domain fluency that can give the impression of general intelligence. They can generate extended discourse, respond coherently to diverse prompts, and adapt stylistically to different communicative contexts. Yet these capabilities arise from probabilistic language modeling rather than conceptual understanding. GenAI systems predict likely sequences of words based on statistical regularities in training data; they do not comprehend meaning, evaluate truth conditions, or possess epistemic awareness (Ng et al., 2025). This gap between apparent fluency and genuine

understanding is particularly consequential in language education, where meaning-making, pragmatics, and sociocultural context are central to communicative competence.

Empirical work has begun testing this gap directly rather than assuming it. Benedetto et al. (2025) examined how accurately large language models encode and apply the Common European Framework of Reference for Languages (CEFR), the proficiency framework against which most communicative language teaching is calibrated. That such testing is necessary at all is informative in itself: if LLM fluency reliably tracked CEFR-defined proficiency constructs, the question would need no empirical investigation. The Weak AI framing anticipates why the two can diverge—a system optimized to predict likely word sequences has no built-in mechanism ensuring its output also respects the graded functional criteria, such as register, discourse control, and sociolinguistic appropriateness, that CEFR descriptors define.

From an educational perspective, failing to recognize the Weak AI nature of contemporary systems risks attributing epistemic authority to tools that are fundamentally limited. When AI-generated feedback or content is treated as inherently correct, objective, or pedagogically optimal, learners and educators may defer judgment inappropriately, undermining critical thinking and reflective learning practices. Research has documented concerns related to AI "*hallucinations*," overgeneralized feedback, and culturally inappropriate language use—issues that stem directly from the absence of genuine understanding in Weak AI systems (Ng et al., 2025).

At the same time, acknowledging these limitations does not diminish the pedagogical value of AI. On the contrary, it clarifies its most productive role: that of a cognitive and procedural scaffold operating within learners' Zones of Proximal Development. Weak AI can efficiently handle repetitive, data-intensive, and surface-level tasks, thereby freeing human educators to focus on higher-order instructional goals such as discourse-level meaning, pragmatic competence, ethical reasoning, and learner identity formation. In this sense, the Weak/Strong AI distinction supports a pedagogical model grounded in *complementarity* rather than substitution.

Henceforth, the educational relevance of this distinction lies in expectation management. Recognizing that current AI systems are powerful but limited tools enables educators to integrate them strategically, avoiding both

technophobic rejection and uncritical adoption. A theoretically informed understanding of Weak AI helps ensure that AI-enhanced language education remains human-centered, epistemically responsible, and pedagogically sound.

2.6 Human–AI Collaboration versus Automation

The integration of Artificial Intelligence into language education has reignited a long-standing tension between *automation* and *pedagogical mediation*. On one end of this continuum, AI-driven automation promises efficiency, scalability, and consistency. Tasks such as grading, scheduling, test generation, plagiarism detection, and preliminary feedback can now be performed rapidly and at scale, significantly reducing teachers' administrative workload (Rodrigues, 2025; Tan et al., 2025). From an institutional perspective, such automation is often framed as a solution to resource constraints, large class sizes, and assessment standardization.

However, contemporary educational research increasingly cautions against equating efficiency with effectiveness. While automation may optimize procedural aspects of teaching, it does not address the inherently relational, interpretive, and ethical dimensions of language education. Language learning is not simply the accumulation of correct forms or lexical items; it is a socially situated process involving identity, affect, culture, and meaning-making. Fully automated instructional models risk flattening these dimensions by prioritizing measurable outputs over developmental processes.

In response to these concerns, current scholarship strongly favors *human–AI collaboration* models, often conceptualized through *human-in-the-loop* or *teacher–AI–student* frameworks. In these models, AI functions as a pedagogical co-pilot rather than a substitute for human expertise. AI systems handle data-intensive, repetitive, and surface-level processes—such as initial feedback generation or performance analytics—while human educators retain responsibility for pedagogical decision-making, ethical judgment, and affective support (Guan et al., 2025; Niloy & Islam, 2025).

This collaborative orientation redefines the role of teachers rather than diminishing it. Teachers become designers of learning environments, curators of AI-generated content, and critical mediators who contextualize

feedback, interpret learner needs, and respond to sociocultural nuances that AI systems cannot perceive. Within language education, this role is particularly crucial, as communicative competence extends beyond grammatical accuracy to include pragmatics, intercultural awareness, and discourse-level meaning.

Empirical studies consistently demonstrate that AI-enhanced instruction yields stronger learning outcomes when it is embedded within teacher-guided pedagogical frameworks. Learners benefit most when AI-mediated activities are contextualized, scaffolded, and critically discussed by instructors who can adapt instruction dynamically and respond to learners' emotional and motivational states (Chang et al., 2025; Herzallah & Makaldy, 2025). Conversely, when AI tools are deployed without pedagogical mediation, learners may experience confusion, overreliance, or misplaced trust in automated feedback.

A quasi-experiment with 124 L2 German learners gives this claim empirical texture rather than leaving it at the level of principle. Comparing a domain-specific large-language-model feedback system (Dr. Write) with generic-LLM feedback and with teacher feedback, Zheng et al. (2026) found that the domain-tuned system produced greater gains in writing self-efficacy and clause-level accuracy than alternative but did not produce a matching gain in affective engagement. This dimension is most tied to a learner's relationship with a human instructor. The finding cuts against a simple story in either direction: AI feedback engineered for a specific pedagogical purpose can outperform both generic AI and unaided teacher feedback on measurable linguistic accuracy, while still leaving a motivational and relational function that, in this study, only human mediation supplied. A systematic review of 30 multiple-group experimental studies conducted between 2014 and 2023 situates this finding within a wider, still-maturing evidence base on how such AI technologies—chatbots and automated writing evaluation among them—have actually been designed and evaluated in language learning, rather than assumed to work uniformly (Qiao et al., 2025). Collaboration, in other words, is not achieved merely by inserting an AI tool alongside a teacher; it depends on which functions the tool is asked to perform and which are deliberately left to the human in the exchange.

Importantly, human–AI collaboration also addresses ethical considerations that automation alone cannot resolve. Issues such as data privacy, algorithmic bias, epistemic trust, and assessment fairness require human oversight and institutional accountability. By maintaining humans at the center of instructional decision-making, collaborative models help ensure that AI serves educational values rather than redefining them.

Thus, the pedagogical question is not whether AI should automate certain tasks, but *how automation can be subordinated to human judgment*. A collaborative approach recognizes AI as a powerful amplifier of pedagogical capacity—one that enhances, rather than replaces, the human dimensions of teaching and learning.

2.7 Conclusion

This chapter has provided a comprehensive conceptual foundation for understanding Artificial Intelligence in language education by synthesizing technological, pedagogical, and philosophical perspectives. Through clarifying key distinctions among AI-related technologies, examining core system archetypes, and situating AI within contemporary instructional paradigms, the chapter has argued against simplistic narratives of technological replacement or inevitability.

Rather than positioning AI as an autonomous authority or neutral solution, current research in English Language Teaching frames it as a *powerful but limited mediator*—one whose educational value depends fundamentally on human interpretation, ethical governance, and pedagogical intentionality. AI systems can process language, generate content, and personalize instruction at unprecedented scale, but they cannot understand meaning, evaluate truth, or engage in the relational and affective dimensions that underpin language learning.

By foregrounding human–AI collaboration over automation, this chapter has emphasized a human-centered vision of AI integration that aligns with established theories of language acquisition, learner autonomy, and sociocultural mediation. In this vision, AI extends instructional capacity without displacing human responsibility, and technological affordances are evaluated in terms of pedagogical purpose rather than novelty.

The chapters that follow build on this conceptual groundwork by examining how AI reshapes language teaching practices, learner experiences, and assessment frameworks in concrete educational contexts. Together, they explore not only what AI *can* do in language education, but what it *should* do when guided by theory, ethics, and informed human judgment.

Pedagogical Companion for Chapter 2

Key Concepts: machine learning; natural language processing; large language models; generative AI; AI system archetypes; weak versus strong AI; human–AI collaboration; automation.

Discussion Questions

1. Which of the AI system archetypes described in this chapter are already present in your teaching or learning environment—and which are present without being recognized as AI?
2. Why does the distinction between weak and strong AI matter for the claims teachers and vendors make about educational technology?
3. LLMs generate fluent language without understanding it in the human sense. What follows for their use as language *models* for learners?
4. Where should the line between human–AI collaboration and automation be drawn in your own institution's practices?
5. Choose one core technology from Section 2.3 and explain it to a colleague with no technical background. What was hardest to convey accurately?

Further Reading

- Kasneci, E., et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*.
- Xi, Z., et al. (2023). The rise and potential of large language model based agents: A survey. arXiv.
- Molenaar, I. (2022). Towards hybrid human-AI learning technologies. *European Journal of Education, 57*(4).

Chapter 3. Theoretical Frameworks for AI-Mediated Language Education

The rapid integration of Artificial Intelligence (AI) into language education necessitates a re-evaluation of established theoretical frameworks. To understand the impact of AI on learning outcomes and pedagogical practices, scholars have increasingly adapted existing theories from Second Language Acquisition (SLA), psychology, and educational technology to accommodate the unique affordances of generative AI (GenAI) and Large Language Models (LLMs). Recent bibliometric and systematic reviews indicate that, despite exponential growth in AI-related language education research, theoretical grounding often lags behind technological adoption (Yang et al., 2025; Zhang et al., 2025).

This chapter anchors the book theoretically by synthesizing six interrelated frameworks: (1) *SLA perspectives, (2) sociocultural theory and mediation, (3) self-regulated learning and learner autonomy, (4) technology-enhanced learning models, (5) posthuman and ecological perspectives, and (6) assessment theory*. Rather than positioning AI as a neutral instructional tool, the chapter conceptualizes AI as a *pedagogically active mediator* that reshapes cognition, interaction, agency, and evaluation in language education.

3.1 SLA Perspectives: Input, Interaction, Output, and Noticing in AI-Mediated Language Education

Traditional second language acquisition (SLA) perspectives—especially those foregrounding *input, interaction, output, and noticing*—remain highly productive for theorizing AI-mediated language learning. What has changed is not the relevance of these constructs, but their conditions of enactment: in many learning episodes the interlocutor is no longer a teacher or peer but a *generative AI system* (e.g., chatbots, voice agents, LLM-based tutors), which alters the ecology of feedback, negotiation of meaning, and opportunities for uptake (Yang et al., 2025; Zhang et al., 2025). Recent reviews in AI-enhanced speaking and language learning also show that research increasingly frames AI effects through interactionist and CALL/SLA lenses, but with recurring caveats about context sensitivity, pragmatic development, and overreliance (Wang et al., 2025; Xu & Ismail, 2025).

3.1.1 Input: From "comprehensible input" to adaptive, on-demand input

Second language acquisition theory has long placed comprehensible input at the center of the acquisition process (Krashen, 1982, 1985). In AI-mediated environments, "input" is often *adaptive, responsive, and multimodal* rather than fixed. Learners can request explanations, paraphrases, simplified versions, examples, or vocabulary support, and systems can generate near-instant textual or spoken input. This aligns well with the SLA expectation that exposure quality and accessibility matter, while also introducing new variables: *what learners ask for*, *how the system frames the input*, and *how reliably the input aligns with curricular targets*.

Systematic reviews of AI for speaking instruction in tertiary EFL contexts report that AI chatbots and platforms provide frequent exposure to models of language use and offer learners expanded practice opportunities beyond classroom time—an input-related advantage in contexts where speaking turns are scarce (Wang et al., 2025). Similarly, literature syntheses on AI for spoken English suggest that AI tools support individualized practice by providing *immediate input–response cycles*, enabling learners to access input matched to perceived needs (Xu & Ismail, 2025). At the same time, scholars warn that input generated by AI may sometimes be *misaligned*, overly fluent, or insufficiently grounded in local curriculum—one reason teacher mediation and careful task design remain essential (Mateos-Blanco et al., 2024; Zhang et al., 2025).

Direct evidence on this point is sobering. Alhazmi and Muftah (2025) asked five expert linguists to rate ChatGPT-4.0's responses to forty-eight second-language-acquisition questions. The linguists judged most responses good to very good, yet their agreement with one another was negligible (Fleiss's kappa = .006). The finding does not show that AI-generated input is inaccurate; it shows that expert judgment of its accuracy is itself unreliable, which is arguably the more troubling result for a theory that treats input quality as measurable and controllable. Input controllability, in other words, presupposes an evaluative consensus that the evidence does not yet support.

A related point is input controllability. Unlike traditional materials, AI input is co-constructed through prompts. Bibliometric mappings show a growing cluster of research focused on "adoption," "acceptance," and "AI

frameworks" in education, this indicates that effective input access depends not only on tool quality but also on learners' and teachers' capacity to use it strategically (Yang et al., 2025; Zhang et al., 2025).

3.1.2 Interaction and negotiation of meaning: Human–AI dialogue as "interaction work"

The Interaction Hypothesis (Long, 1996) supports that acquisition is enhanced when learners engage in meaning-focused interaction that includes negotiation, feedback, and modified input. In AI-mediated contexts, negotiation increasingly occurs through *conversational agents*. Crucially, the *interactional labor* shifts: whereas teachers often initiate checks for comprehension, reformulate learner utterances, and manage turn-taking, GenAI systems frequently require learners to initiate and steer inquiry (Cheng et al., 2025). Cheng et al.'s (2025) comparison of human-tutor vs. GenAI-generated questions is especially important here: it suggests that learning gains are partly contingent on whether learners can generate or be prompted to generate the *right kinds of questions*, implying that interaction quality is not automatic; it is learned and designed.

This shift has two theoretical consequences for SLA:

- *Negotiation becomes prompt mediated***.** Learners negotiate meaning through clarification requests not only about content ("What does X mean?") but about system behavior ("Give me simpler examples," "Correct me but explain why," "Role-play as…"). This expands the repertoire of interactional moves while also raising the bar for metacognitive control and pragmatic precision (Cheng et al., 2025).
- *Interaction becomes scalable but potentially thinner***.** Systematic reviews in speaking instruction consistently show increased participation, practice frequency, and reduced anxiety in AI-supported environments, these benefits are compatible with interactionist predictions (Wang et al., 2025; Xu & Ismail, 2025). Yet they also note persistent challenges: AI may not fully support sociopragmatic nuance, may not always sustain authentic contingency, and may deliver feedback that is not equally meaningful across accents or discourse contexts (Wang et al., 2025; Xu & Ismail, 2025).

From an SLA standpoint, the key question is not whether AI "counts" as interaction, but what interactional properties it affords: (a) opportunities for modified input, (b) feedback that is contingent on learner meaning, and (c) repeated cycles of hypothesis testing through conversational turns. Current evidence suggests AI can provide (a) and (c) robustly, while (b) depends

heavily on task design and learners' ability to request the kind of interaction they need (Cheng et al., 2025; Wang et al., 2025).

A controlled test of this claim in the domain of pragmatics complicates it further. Saeedi and Soltani (2025) compared, in a pre-post study of 125 EFL undergraduates (87 analyzed), an AI chatbot delivering dynamic assessment of speech acts against classroom-based dynamic assessment and against direct feedback alone. The AI-mediated condition did not outperform either alternative; classroom dynamic assessment outperformed direct feedback at both immediate and delayed posttests ($p = .016$ and $p = .041$), while the chatbot condition produced no comparable advantage. If negotiation of meaning depends on an interlocutor's sensitivity to pragmatic context, as interactionist accounts assume, an AI system's failure to outperform even direct feedback for speech-act instruction suggests that conversational scalability and interactional quality are not the same thing.

3.1.3 Output: Low-stakes production, pushed output, and iterative rehearsal

The Output Hypothesis (Swain, 1995) emphasizes that producing language can "push" learners to process form more deeply, notice gaps, and test hypotheses. AI environments may amplify output in at least three ways:

1. *Increased volume and frequency of production:* Reviews of AI in EFL speaking instruction highlight that chatbots and speaking platforms can increase learners' speaking turns and willingness to communicate, partly because the interaction is low-pressure and repeatable (Wang et al., 2025). Similarly, spoken English literature reviews describe AI as enabling flexible, individualized rehearsal routines that are difficult to sustain in teacher-led contexts alone (Xu & Ismail, 2025).
2. *Private rehearsal that reduces affective barriers*: AI-mediated speaking practice is often experienced as psychologically safer than public classroom performance, supporting risk-taking and experimentation. This can yield more output attempts and thus more occasions for noticing and repair (Wang et al., 2025).
3. *Output becomes "iterable" through rapid feedback loops:* AI systems encourage repeated re-production: learners produce an utterance, receive feedback, revise, and try again—supporting a cyclical process consistent with output-driven development (Xu & Ismail, 2025). However, these benefits also introduce a risk repeatedly flagged in reviews: *overreliance,* where learners outsource linguistic problem-solving rather than developing internal monitoring strategies (Mohebbi, 2025; Xu & Ismail, 2025).

A direct comparison qualifies how this iterability plays out over time. Tang et al. (2026) followed sixty-one Chinese university students producing 244 texts over eight weeks, contrasting teacher feedback with AI-assisted peer feedback on L2 writing. Teacher feedback produced stronger short-term gains, but AI-assisted peer feedback showed a steadier, more sustained learning-curve effect across the eight weeks. Read against the Output Hypothesis, this suggests that AI-mediated iteration may trade immediate precision for durability: teachers close gaps faster in a single cycle, while AI-supported repetition accumulates smaller gains that persist longer.

3.1.4 Noticing: Feedback, comparison, and attention to form in LLM-mediated learning

Noticing—attention to linguistic form (Schmidt, 1990)—is central to linking interaction and output to acquisition. LLM-based tools can intensify noticing by making form-focused feedback immediate, salient, and on demand—yet the *quality* and *pedagogical alignment* of that feedback is decisive.

A key development is the shift from rule-based feedback to *adaptive, response-contingent feedback* generated by LLMs. In teacher education settings, Kinder et al. (2025) show that LLM-generated adaptive feedback can be effective, illustrating how feedback can be tailored to learner needs and context. From an SLA lens, this means feedback can be calibrated to maximize noticing while reducing extraneous processing—consistent with claims that adaptive feedback can lower cognitive burden by helping learners focus attention on relevant features (Kinder et al., 2025).

However, the literature also suggests two constraints on noticing in AI contexts:

- *Prompt quality as a gatekeeper of noticing:* If learners ask vague questions ("Help me"), feedback may be generic and fail to target the linguistic feature that would produce noticing. When prompts are precise ("Correct my tense errors and explain the rule with two examples at B1"), feedback becomes more usable for attention to form. This connects directly to findings that learners' ability to ask pragmatic, information-oriented questions mediates learning outcomes (Cheng et al., 2025).
- *Context sensitivity and trust:* Tools may offer fluent, confident feedback that appears authoritative, but can be misleading or insufficiently contextualized. Language-teaching tool surveys stress the need for AI literacy and critical evaluation to prevent uncritical uptake (Mateos-Blanco et al.,

2024). This is particularly important because noticing is only beneficial if the noticed feature is *accurately represented* and *instructionally relevant*.

Even well-targeted feedback only supports noticing if learners actually engage with it, and engagement is not guaranteed. Ziqi et al.'s (2024) study of forty-five L2 students revising argumentative essays with ChatGPT-generated feedback found that learners rejected or simply ignored roughly 38% of it -- 22% argued against the feedback, 16% disregarded it outright -- with content-level feedback rejected significantly more often than form-level feedback, largely because learners found it inconvenient or unhelpful to act on. For a theory that locates acquisition in noticed, attended-to form, a rejection rate approaching two-fifths is not a minor caveat: it means that a substantial share of the noticing opportunities AI feedback is assumed to create may never be taken up at all.

3.1.5 A synthesis: What SLA helps us see—and what it misses

SLA perspectives provide a powerful explanatory backbone for AI-mediated language learning, because AI tools often function as *input providers, interaction partners, output elicitors, and noticing triggers*. Empirical syntheses show that AI tools can enhance practice frequency, engagement, and certain performance indicators, particularly in speaking, while also foregrounding limitations related to pragmatic competence, contextual feedback, and the risk of overdependence (Wang et al., 2025; Xu & Ismail, 2025).

At the same time, bibliometric mappings indicate that the field is increasingly concerned with issues that classic SLA does not fully address on its own: adoption and acceptance, equity, teacher agency, and ethical governance (Yang et al., 2025; Zhang et al., 2025). This is why the subsequent sections of this chapter extend beyond SLA to sociocultural mediation, SRL and autonomy, ecological/posthuman perspectives, and assessment theory—building an integrated theoretical foundation rather than a single-theory explanation.

Quantitative synthesis lends the picture more precision, and some friction. Torres and Kahveci's (2025) multilevel meta-analysis of 117 effect sizes drawn from 46 EFL studies published between 2022 and 2025 found a medium-to-large overall effect of AI on language learning – but the effect was strongest for vocabulary and weakest for speaking, the very skill on which the interactionist and output-oriented arguments above rest most heavily.

A PRISMA-guided review of twenty-two empirical studies in school-level EFL settings reaches a compatible but more granular conclusion: AI tools produced significant gains across reading, writing, listening, speaking, and vocabulary alike, though digital-divide and teacher-readiness challenges persisted (Kundu & Bej, 2025). Taken together, the meta-analytic evidence supports SLA-grounded optimism about AI-mediated input, interaction, and output, while also showing that the size of the benefit is uneven across exactly the domains this section has treated as roughly interchangeable.

3.2 Sociocultural Theory, Constructivism and Mediation in AI-Mediated Language Education

While SLA perspectives primarily foreground cognitive mechanisms of language development, *sociocultural theory* conceptualizes learning as a *socially mediated, historically situated activity*, shaped by interaction with others and with cultural tools. From this standpoint, language learning is not just an internal process but a *distributed phenomenon*, emerging through participation in socially organized practices. The integration of artificial intelligence into language education therefore requires a reconceptualization of mediation, agency, and interactional roles.

Recent bibliometric analyzes indicate that sociocultural and constructivist perspectives are increasingly invoked to make sense of AI-mediated learning, particularly as researchers recognize that AI systems function not only as instructional tools but as *active mediators of learning processes* (Yang et al., 2025; Zhang et al., 2025). In AI-enhanced classrooms, learning is no longer organized solely around teacher–student interaction but around a *tripartite configuration* involving teachers, learners, and AI systems (Guan et al., 2025).

Constructivist learning theory provides a strong pedagogical foundation for the integration of artificial intelligence (AI) into foreign language education because it conceptualizes learners as active constructors of knowledge rather than passive recipients of linguistic input. Drawing on Piagetian and Vygotskian perspectives, constructivism emphasizes learner agency, prior knowledge, social interaction, contextualized experience, and reflection in the learning process (Alcazar-Gonzales, 2026). Within this framework, AI can function as a form of adaptive scaffolding by providing personalized learning pathways, immediate feedback, and opportunities for autonomous and interactive language practice. Alcazar-Gonzales (2026) argues that AI-

supported personalization and continuous feedback are closely aligned with constructivist principles because they allow learners to construct knowledge autonomously and refine their understanding through ongoing interaction and reflection. This connection is particularly relevant to foreign language education, where AI-mediated feedback and adaptive learning pathways can operate as digital scaffolds that support learners in progressing toward higher levels of language proficiency (Talehati, 2025). Moreover, AI-supported language environments can facilitate contextualized interaction, collaborative learning, and communicative practice, extending constructivist principles into technology-mediated learning spaces (Talehati, 2025). Thus, AI should not be conceptualized merely as a tool for delivering linguistic information, but as a pedagogical mediator that, when appropriately guided by teachers, can support the active, reflective, social, and personalized construction of foreign language knowledge.

3.2.1 Mediation and the Zone of Proximal Development (ZPD)

At the core of sociocultural theory is the notion of *mediation*, whereby learning is facilitated through tools, signs, and interaction with more capable others. Closely linked to this is the concept of the *Zone of Proximal Development (ZPD),* defined as the space between what learners can accomplish independently and what they can achieve with guidance.

In AI-mediated environments, digital systems increasingly act as *mediational artifacts* that provide forms of guidance traditionally associated with teachers or peers. NLP-driven tools and LLM-based systems can offer *graduated scaffolding*, prompting reflection, modeling strategies, and breaking complex tasks into manageable steps (Molenaar, 2022). From a sociocultural perspective, such functions resemble scaffolding within the ZPD, as learners are supported in performing tasks that exceed their unaided capacities.

Systematic reviews of AI in language education further confirm that AI tools often support learning through *immediate feedback, adaptive prompts, and iterative interaction*, all of which can extend learners' functional performance (Wang et al., 2025; Xu & Ismail, 2025). However, these reviews also emphasize that AI-mediated scaffolding is *qualitatively different* from human mediation. AI systems respond algorithmically and lack sensitivity to learners' emotional states, cultural histories, and identity

positioning—dimensions that are central to sociocultural accounts of learning (Mateos-Blanco et al., 2024; Zhang et al., 2025).

Accordingly, Guan et al. (2025) argue that AI should not be conceptualized as an autonomous tutor but rather as a *mediating partner* that augments, rather than replaces, human instructional roles.

3.2.2 Human-in-the-Loop Mediation and Teacher Agency

The growing prominence of AI has foregrounded the importance of *human-in-the-loop* models of instruction, which emphasize sustained teacher involvement in AI-mediated learning environments. From a sociocultural perspective, this involvement is not optional but foundational, as mediation is inherently relational and value-laden.

Research on AI acceptance and implementation consistently shows that teachers play a critical role in *interpreting, contextualizing, and regulating AI feedback*, ensuring that system outputs align with pedagogical goals and learners' developmental needs (Masry-Herzallah & Makaldy, 2025; Zhang et al., 2025). Teachers also function as *affective and ethical mediators*, providing emotional support, empathy, and motivational guidance—dimensions of learning that current AI systems cannot replicate (Mohebbi, 2025).

Systematic reviews on learner autonomy and self-regulation further suggest that when AI systems operate without sufficient pedagogical oversight, learners may become overly dependent on automated feedback, weakening their capacity for independent problem-solving and reflection (Mohebbi, 2025; Xu & Ismail, 2025). Sociocultural theory therefore reinforces the need to preserve *teacher agency*, positioning AI as a tool whose mediational value emerges only through deliberate pedagogical orchestration.

3.2.3 Collaborative Learning and AI as a Sociocultural Mediator

Sociocultural theory also emphasizes collaborative learning and the co-construction of knowledge through interaction. In this respect, AI technologies are increasingly embedded within *Computer-Supported Collaborative Learning (CSCL)* environments, where they function as mediators of group processes rather than direct instructors.

Empirical studies show that AI systems can analyze interaction patterns, monitor participation equity, and provide feedback on group dynamics,

thereby shaping how learners engage with one another (Aslan et al., 2025). Such systems act as *meta-level mediators*, guiding learners toward more productive collaborative practices without directly replacing peer interaction.

This aligns with findings from broader reviews of AI-enhanced language learning, which note that AI can support social interaction by structuring participation and highlighting communicative behaviors that facilitate collaboration (Wang et al., 2025; Yang et al., 2025). However, scholars also caution that excessive algorithmic regulation may constrain spontaneous interaction and reduce opportunities for learners to negotiate roles organically (Zhang et al., 2025).

From a sociocultural standpoint, effective AI-mediated collaboration requires a balance between *structure and openness*, ensuring that AI supports, rather than scripts, social interaction.

3.2.4 Sociocultural Implications for AI-Mediated Language Education

Taken together, sociocultural theory provides a powerful framework for understanding AI-mediated language education as a *distributed, relational, and tool-mediated process*. AI systems can extend learners' ZPDs, support reflective thinking, and enhance collaborative learning. Yet their effectiveness is contingent upon *human mediation*, ethical oversight, and pedagogical intentionality.

Bibliometric mappings of the field increasingly point to this convergence: AI in language education is moving away from purely instrumental accounts toward frameworks that emphasize *mediation, agency, and social context* (Yang et al., 2025; Zhang et al., 2025). Sociocultural theory thus acts as a critical bridge between cognitive SLA perspectives and broader ecological and posthuman accounts developed later in this chapter.

3.3 Self-Regulated Learning and Learner Autonomy in AI-Mediated Language Education

Support for *Self-Regulated Learning (SRL)* is one of the most consistently reported affordances of artificial intelligence in contemporary education, and particularly in AI-mediated language learning. Across systematic

reviews, bibliometric analyzes, and empirical studies, SRL emerges as a central explanatory construct for understanding how learners interact with AI tools, manage their learning processes, and develop autonomy over time (Mohebbi, 2025; Yang et al., 2025; Zhang et al., 2025).

SRL is typically conceptualized as a *cyclical process* involving forethought (goal setting and planning), performance (strategy use and monitoring), and self-reflection (evaluation and adaptation) (Heung & Chiu, 2025). AI systems, especially generative AI and learning analytics–driven platforms, are uniquely positioned to scaffold each of these phases by providing continuous feedback, adaptive prompts, and data-informed insights into learner behavior. However, the literature also highlights important tensions between *automation and agency*, raising questions about whether AI genuinely cultivates autonomy or merely simulates it.

3.3.1 AI as a Catalyst for Self-Regulated Learning

Recent systematic reviews demonstrate that AI tools can enhance learners' self-regulatory capacities by offering *immediate, individualized, and persistent support* that is difficult to sustain in traditional classroom settings (Mohebbi, 2025; Wang et al., 2025). In language learning contexts, AI chatbots, automated feedback systems, and adaptive platforms enable learners to practice skills independently, set personalized goals, and monitor progress without constant teacher intervention.

Mohebbi's (2025) systematic review on AI and learner independence shows that AI-mediated environments often encourage learners to take greater responsibility for pacing, task selection, and revision. Similarly, reviews of AI-enhanced speaking instruction indicate that learners engage more frequently in self-initiated practice when AI tools are available, suggesting a shift toward *self-directed engagement* (Wang et al., 2025; Xu & Ismail, 2025). From an SRL perspective, these affordances align with the forethought and performance phases of self-regulation.

A controlled comparison gives this claim empirical weight beyond review-level synthesis. Wei (2023) compared sixty EFL university students in experimental and control groups and found that the AI-mediated instruction group achieved significantly higher English achievement, L2 motivation, and self-regulated learning strategy use than the traditionally taught control group. The result is one of the few in this literature to measure SRL strategy

use as a direct outcome variable rather than inferring it from engagement or persistence data, and it lends the forethought-and-performance argument above a firmer empirical footing than the bibliometric mappings alone provide.

However, bibliometric mappings of the field caution that increased independence does not automatically translate into *effective self-regulation*. Yang et al. (2025) and Zhang et al. (2025) note that many studies report gains in engagement and efficiency but provide limited evidence of sustained metacognitive development, indicating the need for deeper theoretical integration of SRL frameworks.

3.3.2 COPES, Learning Analytics, and Metacognitive Scaffolding

Winne and Hadwin's *COPES model (Conditions, Operations, Products, Evaluations, and Standards)* has become a key framework for analyzing SRL in AI-mediated learning environments. Within this model, AI systems are conceptualized as part of the *external conditions* that shape learners' cognitive operations and strategic choices.

Learning analytics derived from AI systems can detect behavioral patterns such as hesitation, repeated errors, or disengagement, allowing systems to trigger *adaptive scaffolds* that prompt planning, monitoring, or strategy adjustment (Li et al., 2025). Empirical findings suggest that such analytics-driven interventions are positively associated with learning performance, particularly in self-paced and asynchronous environments where learners must regulate their own engagement (Li et al., 2025).

From a language education perspective, this form of metacognitive scaffolding can support learners in noticing gaps in their knowledge, selecting appropriate strategies, and evaluating outcomes—processes closely aligned with both SRL and SLA-informed noticing mechanisms. However, Mateos-Blanco et al. (2024) emphasize that the effectiveness of such scaffolding depends on *AI literacy*: learners must understand how to interpret feedback and analytics rather than accepting them uncritically.

3.3.3 Learner Autonomy and the Paradox of Automation

Learner autonomy, classically defined as the capacity to take responsibility of one's own learning (Holec, 1981; Little, 2002), has long been a central goal of language education, and AI tools appear, at least superficially, to support this goal by enabling self-paced, on-demand learning. Reviews of

AI in spoken English learning consistently report increased learner confidence and willingness to practice independently when AI tools are available (Xu & Ismail, 2025; Wang et al., 2025).

Structural modeling helps specify the mechanism behind this apparent support. Alqurni (2026), surveying 280 Saudi higher-education students, found that learners' perceived agency in AI systems predicted perceived usefulness, ease of use, and autonomy support, which in turn drove self-efficacy and self-directed learning motivation. The path is instructive: autonomy support is not a direct effect of AI availability but is mediated by how much agency learners perceive the system to have -- a perception that, as the paradox discussed next suggests, can just as easily transfer control away from the learner as toward them.

Yet the literature also identifies a paradox: the same automation that enables autonomy can also undermine it. When AI systems provide instant answers, corrections, and suggestions, learners may bypass productive struggle and reflective decision-making. Mohebbi (2025) notes that over-reliance on AI feedback can weaken learners' capacity for independent monitoring and evaluation, particularly if AI is positioned as an authoritative source rather than a learning aid.

This tension is reflected in bibliometric studies showing growing concern with issues such as dependency, ethics, and learner agency in AI-mediated education (Yang et al., 2025; Zhang et al., 2025). From an SRL standpoint, autonomy is more than merely about independence from teachers but about control over cognitive and metacognitive processes, this distinction becomes especially salient in AI-rich environments.

3.3.4 Self-Determination Theory (SDT) and Motivational Dynamics

Self-Determination Theory (SDT) provides an additional lens for examining learner motivation in AI-mediated contexts by focusing on the basic psychological needs of autonomy, competence, and relatedness. Empirical studies suggest that AI tools can support autonomy by offering flexible pacing and choice, and competence by providing immediate, individualized feedback that reinforces progress (Ode et al., 2025).

However, longitudinal analyzes reveal that these motivational benefits may not be stable over time. Bognár and Khine (2025) report that initial

increases in engagement and self-efficacy in AI-enhanced classrooms may decline as novelty effects diminish, particularly if learners are not guided toward deeper self-regulatory practices. Beyond this, SDT highlights the importance of relatedness, which can be compromised in AI-heavy learning environments if human interaction is reduced or marginalized (Ode et al., 2025).

These findings resonate with broader concerns in AI-in-education research about the risk of isolating learners or replacing meaningful social interaction with algorithmic feedback (Mateos-Blanco et al., 2024; Zhang et al., 2025).

A growth-modelling study offers a more textured picture of these dynamics than a single pre-post contrast can. Shi and Derakhshan (2026) surveyed 595 Chinese EFL students at three time points during an AI-powered course and modeled change using latent growth curve modeling; autonomy, relatedness, and competence -- the three SDT needs invoked above -- all showed growth across the three waves rather than a single post-novelty decline. This does not contradict Bognár and Khine's (2025) finding of engagement loss after the initial semester so much as suggest that different SDT needs may follow different trajectories, and that a two-point pre-post design and a three-wave growth model can license different conclusions about the same construct.

A meta-analysis of the broader agency literature adds a caution about generalizing from any single study in either direction. Zai and Zhou's (2026) meta-analysis of AI-assisted learning's effect on foreign language learner agency found positive associations across all dimensions examined, largest for engagement ($r = .648$) and smallest for enjoyment ($r = .392$), but flagged extreme heterogeneity across the included studies. The direction of the evidence is consistent; its magnitude is not, which is itself an argument for the theory-driven integration this section has called for rather than continued accumulation of loosely comparable effect sizes.

3.3.5 Implications for SRL and Autonomy in AI-Mediated Language Learning

Taken together, the literature suggests that AI has significant potential to support *self-regulated learning and learner autonomy,* but only under certain conditions. AI systems can scaffold planning, monitoring, and

reflection; provide analytics-driven insights; and enable flexible, individualized learning trajectories. At the same time, without careful pedagogical design and teacher mediation, AI may support superficial autonomy characterized by convenience rather than strategic self-regulation.

Bibliometric and systematic reviews increasingly call for theory-driven integration of SRL and autonomy frameworks into AI-mediated language education research (Yang et al., 2025; Mohebbi, 2025). From this perspective, SRL functions as a bridge between cognitive SLA theories and sociocultural and ecological accounts, emphasizing learner agency while acknowledging the powerful mediational role of AI.

3.4 Technology-Enhanced Learning Models in AI-Mediated Language Education

To operationalize the integration of artificial intelligence in language education, researchers have increasingly drawn on—and adapted—established *technology-enhanced learning (TEL) models*. While early TEL frameworks conceptualized technology as a supplementary instructional aid, AI systems introduce qualitatively different affordances, including adaptivity, generativity, and autonomous decision-making. As a result, traditional models such as the Technology Acceptance Model (TAM), Social Cognitive Theory, TPACK, and SAMR are being reinterpreted to account for *teacher agency, ethical responsibility, and pedagogical redesign* in AI-rich environments (Belkina et al., 2025; Tan et al., 2025; Yang et al., 2025).

Bibliometric analyzes indicate a clear shift in AI-in-education research from questions of *whether* technology should be adopted to *how* it can be pedagogically and ethically integrated (Yang et al., 2025; Zhang et al., 2025). Technology-enhanced learning models thus serve as critical bridges between theoretical principles and classroom implementation.

3.4.1 Technology Acceptance, Social Cognitive Theory, and Teacher Readiness

The *Technology Acceptance Model (TAM)* remains one of the most widely applied frameworks for understanding educators' adoption of AI tools. However, recent studies suggest that TAM alone is insufficient to explain teachers' complex emotional, cognitive, and motivational responses to generative AI. Accordingly, researchers increasingly extend TAM with

constructs drawn from *Social Cognitive Theory*, including *Technological Self-Efficacy (TSE)* and *Sense of Coherence (SOC).*

Empirical evidence demonstrates that teachers' perceived self-efficacy in using AI tools strongly predicts their willingness to integrate AI into pedagogical practice (Masry-Herzallah & Makaldy, 2025). Educators who believe they can competently manage AI systems are more likely to view AI as a pedagogical resource rather than a threat to professional identity. Similarly, SOC—defined as the extent to which individuals perceive new technologies as comprehensible, manageable, and meaningful—has been shown to moderate anxiety and resistance toward AI adoption (Masry-Herzallah & Makaldy, 2025).

Several recent applications of extended TAM and UTAUT2 models put numbers to this pattern. Gao and Yang (2026), surveying 326 faculty and interviewing 22 more at Shanghai University, found that performance expectancy, perceived usefulness, institutional support, and AI self-efficacy all predicted adoption attitude, while perceived risk reduced it. Among Moroccan preservice teachers evaluating AI for inclusive education, perceived usefulness emerged as the strongest single predictor of acceptability in a TAM2 model (Amouri et al., 2025), and a UTAUT-based survey of 498 preservice teachers found AI self-efficacy and digital literacy jointly predicting 51.4% of the variance in generative-AI acceptance (Bulut, 2026). Not every study locates the predictive weight in the same place, however: a UTAUT2 study of 400 Spanish university students found experience, performance expectancy, hedonic motivation, and habit driving behavioral intention to use ChatGPT (Romero-Rodríguez et al., 2023), while a parallel UTAUT2 study of 425 Ecuadorian university teachers found high acceptance of educational AI tracking most closely with constructivist pedagogical beliefs and teaching modality (Cabero-Almenara et al., 2024). Self-efficacy is the one construct that recurs as significant across nearly every model reported here; which other construct joins it seems to depend on context.

Comparative research on pre-service teachers' AI readiness further highlights contextual and cultural variation in acceptance patterns. Lucas and Dorotea (2025) report that differences in institutional support, professional preparation, and national policy frameworks considerably shape teachers' readiness to engage with AI-enhanced instruction. These findings underline

that technology acceptance is not merely an individual psychological disposition but a situated phenomenon, embedded within broader institutional and socio-cultural contexts.

Other studies complicate even the primacy of usefulness itself. A PLS-SEM survey of 306 higher-education students found that satisfaction was the strongest predictor of AI-tool use in curricula -- ahead of perceived utility, which had only a moderate effect -- while content quality and credibility showed weak, non-significant effects (Mat Yusoff et al., 2025). Self-efficacy fares more consistently: among 416 Thai K-12 teachers, self-efficacy strongly predicted generative-AI adoption ($r = .68$), though the same teachers voiced ethical and critical-thinking concerns and cited infrastructure gaps even as they adopted the tools (Tongchai & Malakul, 2025). Other moderators complicate the picture further -- gender differences confined specifically to perceived ease of use among 239 Kazakhstani university students, not to overall attitude (Yilmaz et al., 2023), and, among 356 Turkish university students, AI learning anxiety moderating the acceptance pathway even as AI literacy and self-efficacy mediated it (Türk et al., 2025). Technology acceptance in AI-mediated language education, in short, is directionally consistent but structurally inconsistent: adoption is reliably positive, but the psychological architecture producing it differs by population, instrument, and national context.

From a TEL perspective, teacher readiness is thus conceptualized not only as technical competence but as a form of pedagogical agency, enabling educators to critically evaluate AI tools, align them with curricular goals, and mediate their classroom use. Supporting this perspective, Can and Mangır (2024) found that hands-on engagement with ChatGPT enhanced language teachers' willingness to integrate AI, while concerns regarding time, infrastructure, and implementation persisted. Their findings underscore the importance of sustained professional development and institutional support for meaningful and effective AI integration.

3.4.2 TPACK and SAMR in the Age of Generative AI

The *Technological Pedagogical Content Knowledge (TPACK)* framework (Mishra & Koehler, 2006) has long served as a foundational model for integrating technology into subject-specific teaching. In the context of generative AI, however, scholars argue that traditional TPACK must be reconceptualized to address new forms of knowledge and responsibility. This has

led to the emergence of *GenAI-TPACK,* which foregrounds ethical awareness, prompt literacy, and pedagogical judgment as core competencies for teachers (Belkina et al., 2025; Tan et al., 2025).

Systematic reviews of GenAI implementation in higher education reveal that successful integration depends not on technical sophistication alone but on teachers' ability to design tasks that leverage AI's generative capacities while maintaining instructional coherence and academic integrity (Belkina et al., 2025). Prompt design, feedback interpretation, and decision-making about when and how AI should be used become integral components of pedagogical expertise.

The *SAMR model* (Substitution, Augmentation, Modification, Redefinition) (Puentedura, 2010) is frequently employed to categorize levels of AI integration. Current research suggests that many AI applications in language education remain at the substitution or augmentation levels, such as grammar checking or automated feedback (Belkina et al., 2025). While these uses may enhance efficiency, they do not fundamentally transform learning processes.

The pedagogical aspiration, emphasized across recent studies, is redefinition—where AI enables learning tasks previously inconceivable without it. Examples include dynamic role-play simulations with adaptive interlocutors, real-time personalized discourse modeling, and AI-mediated collaborative problem-solving environments (Belkina et al., 2025; Tan et al., 2025). TEL models thus shift attention from technological novelty to pedagogical transformation.

3.4.3 TEL Models as Ethical and Pedagogical Frameworks

Beyond instructional design, TEL models increasingly function as ethical frameworks for AI integration. Bibliometric analyzes reveal growing scholarly concern with transparency, accountability, and the distribution of agency between humans and machines (Yang et al., 2025; Zhang et al., 2025). In this sense, TEL models are no longer neutral descriptors of integration stages but normative guides that help educators navigate ethical dilemmas related to authorship, assessment, and learner autonomy.

3.5 Posthuman and Ecological Perspectives on Language Learning

Moving beyond tool-centered approaches, *posthuman and ecological perspectives* conceptualize AI as part of a complex socio-technical ecosystem in which agency is distributed across human and non-human actors. These perspectives challenge anthropocentric assumptions underlying many TEL and SLA models, offering alternative ways of understanding learning, identity, and knowledge production in AI-mediated environments.

3.5.1 Socio-Ecological Models of AI Integration

Socio-ecological models emphasize that successful AI integration requires alignment across multiple levels of the educational system:

- *Macro level:* national policy, legal frameworks, and ethical guidelines
- *Meso level:* institutional leadership, infrastructure, and professional development
- *Micro level:* classroom practices, teacher beliefs, and learner engagement

Empirical research demonstrates that misalignment across these levels often leads to superficial or unsustainable AI adoption (Ng et al., 2025). For example, schools may introduce AI tools without adequate teacher training or ethical governance, resulting in resistance or misuse.

From a socio-ecological perspective, AI is not an external add-on but an intrinsic component of the learning environment, shaping and being shaped by institutional norms, pedagogical values, and social relationships (Ng et al., 2025; Zhang et al., 2025). This view reinforces the need for holistic, system-level approaches to AI integration in language education.

3.5.2 Posthuman Inquiry and Distributed Agency

Posthumanist perspectives further destabilize traditional distinctions between human subjects and technological tools by framing AI as a co-participant in educational practice. Rather than viewing AI as a passive instrument, posthuman inquiry examines how interactions with AI systems reshape teacher identity, learner agency, and epistemic authority.

Qualitative studies indicate that learners often attribute expertise, authority, or even personality to AI systems, influencing their sense of authorship and trust in knowledge production (Nazaretsky et al., 2025). Similarly, research on tripartite interaction in EFL education shows that AI systems mediate

not only cognitive processes but also social relationships and power dynamics in the classroom (Guan et al., 2025).

From this perspective, learning is understood as emerging from assemblages of humans, technologies, data, and discourses, rather than residing solely within individual learners. This ontological shift has profound implications for language education, particularly in areas such as writing, assessment, and collaborative meaning-making.

3.5.3 Implications for Language Education Theory

Posthuman and ecological perspectives compel language educators and researchers to reconsider foundational assumptions about agency, responsibility, and learning outcomes. When AI becomes a co-participant in instructional processes, questions arise regarding authorship, accountability, and the ethics of delegation.

Bibliometric mappings suggest that these concerns are gaining prominence in AI-in-language-education research, signaling a move toward critical, theory-driven inquiry rather than purely instrumental adoption (Yang et al., 2025). In this sense, posthuman and ecological perspectives complement sociocultural and SRL frameworks by extending analysis beyond individual learners to encompass broader socio-technical configurations.

3.6 Assessment Theory in AI-Mediated Language Education: Validity, Fairness, and Washback

Assessment occupies a pivotal position in language education, shaping not only how learning outcomes are measured but also how teaching and learning are enacted. The integration of artificial intelligence into language assessment fundamentally challenges established assumptions about *validity, fairness, reliability, transparency, and washback*. While AI-enabled assessment systems promise efficiency, scalability, and immediacy, recent research cautions that these benefits are accompanied by significant theoretical and ethical tensions (Mateos-Blanco et al., 2024; Yang et al., 2025).

Bibliometric analyzes indicate that assessment has emerged as one of the most contested domains within AI-mediated language education, with growing scholarly attention to issues of bias, explainability, and construct representation (Yang et al., 2025; Zhang et al., 2025). Assessment theory

therefore provides a crucial lens for evaluating not only *how* AI is used to assess language but *what* kinds of language competence are being privileged and legitimized.

3.6.1 Reliability, Consistency, and Scalability

One of the most frequently cited advantages of AI-based assessment is its potential to enhance *reliability and consistency*, particularly in large-scale and high-stakes contexts. Automated scoring systems can apply uniform criteria across large populations, reducing variability associated with human fatigue, subjectivity, or inconsistency (Xi, 2010).

Empirical evidence from international large-scale assessments supports this claim. Jung, Tyack, and von Davier (2025), examining the implementation of automated scoring in international large-scale assessments with TIMSS 2023 data, demonstrate that carefully engineered scoring pipelines—combining machine translation with artificial neural networks—can approach the reliability of trained human raters across multilingual constructed responses, provided that scalability and quality-control safeguards are built into the process. Although their study concerns science assessment rather than language proficiency, its implications for language assessment are substantial, particularly for contexts involving multilingual populations and constructed responses.

From an assessment theory perspective, increased reliability strengthens score comparability and procedural consistency. However, reliability alone is insufficient; consistent scoring does not guarantee that what is being measured aligns with intended constructs of communicative competence.

3.6.2 Construct Validity and Alignment with Language Proficiency Frameworks

Validity, particularly *construct validity*, remains the most critical and contested issue in AI-mediated language assessment. The central question is whether AI systems genuinely assess language ability or merely measure alignment with algorithmic patterns.

Recent empirical studies raise serious concerns in this regard. Benedetto et al. (2025) examine how accurately large language models encode and apply the Common European Framework of Reference for Languages (CEFR). Their findings indicate that LLMs frequently misclassify proficiency levels, struggle to simulate learner-typical errors, and inconsistently distinguish

between adjacent levels (e.g., B1 vs. B2). Such limitations challenge the validity of AI-based tools for both formative and summative assessment when used without rigorous calibration.

Similarly, Willert and Würz (2025) show that while Bayesian and AI-assisted approaches can support formative assessment design, automated systems require careful task modeling and human oversight to ensure alignment with intended learning outcomes. These findings underscore a key assessment-theoretical concern: AI systems often perform well at surface-level linguistic patterning but struggle with contextualized, pragmatic, and discourse-level competence.

From a theoretical standpoint, this misalignment threatens the interpretive argument underlying assessment validity, calling into question the inferences drawn from AI-generated scores.

3.6.3 Fairness, Bias, and Algorithmic Justice

Beyond technical validity, *fairness* has become a central concern in AI-mediated assessment. Traditional assessment theory recognizes fairness as both a statistical and a social construct, encompassing bias, equity, and stakeholder perceptions.

Research on Automatic Short Answer Grading (ASAG) using models such as GPT-4 suggests that AI systems may demonstrate a degree of *algorithmic consistency* across demographic groups. Rodrigues et al. (2025) report that LLM-based graders tend to be more lenient than human raters but show relatively stable scoring patterns across gender and native-language groups. This finding challenges assumptions that AI necessarily amplifies bias and suggests potential for reducing certain forms of human rater variability.

However, fairness is not solely an empirical property of scores; it is also a matter of perceived legitimacy. Jones-Jang et al. (2025) show that students' trust in AI grading systems is strongly influenced by transparency and explainability. When assessment decisions are perceived as opaque or unchallengeable, learners may experience dissatisfaction and reduced acceptance—even if statistical fairness is achieved.

This tension highlights a fundamental issue in AI-mediated assessment: procedural justice. Without clear explanations of how scores are generated, AI systems risk undermining learners' sense of agency and fairness, regardless of technical performance.

The pattern these two studies illustrate is not idiosyncratic. Plakans and Lee's (2025) state-of-the-art review of 869 peer-reviewed articles on second language writing assessment published over twenty-five years identifies five major themes running through the field and explicitly calls for a fairness- and justice-oriented, multilingual turn in how writing assessment -- automated or otherwise -- is theorized and conducted. Read against a quarter-century of accumulated research rather than against any single study, the tension this chapter has been tracing between statistical consistency and perceived legitimacy is not a transitional problem that better models will resolve; it is a structural feature of the assessment enterprise that AI has inherited rather than created, and one the multilingual turn the review calls for would require assessment theory -- not just assessment technology -- to address.

3.6.4 Explainability, Transparency, and the "Black Box" Problem

Assessment theory has long emphasized transparency as a prerequisite for ethical and pedagogically sound evaluation. In AI-mediated assessment, transparency is complicated by the "black box" nature of many machine-learning models.

Studies consistently emphasize the need for *explainable AI (XAI)* in educational assessment. Jones-Jang et al. (2025) argue that explainability is essential not only for ethical compliance but also for learner acceptance and instructional usefulness. Without interpretable feedback, AI-generated scores risk becoming detached from learning processes, reducing assessment to a purely classificatory function.

Mateos-Blanco et al. (2024) similarly stress that AI tools used in language education must be accompanied by clear pedagogical rationales and user literacy, enabling teachers and learners to understand the scope and limitations of automated evaluation.

3.6.5 Washback Effects and Pedagogical Consequences

Washback—the influence of assessment on teaching and learning—is a critical dimension of assessment theory that takes on new significance in AI-mediated contexts. AI-driven assessment systems provide continuous, data-rich feedback that can shape instructional priorities and learner strategies in powerful ways.

On the positive side, AI-enabled formative assessment can support self-regulated learning by offering immediate feedback, progress tracking, and opportunities for revision (Mohebbi, 2025; Xu & Ismail, 2025). Such practices align assessment more closely with learning, reinforcing formative rather than purely summative functions.

A controlled study outside the language-education literature narrowly construed offers concrete support for this formative pathway. Al-Aghbari et al. (2025) implemented AI-based formative assessment with seventy eighth-grade female students in Oman's Muscat Governorate and found significant improvements in both academic achievement and students' perceptions of the classroom assessment environment relative to conventional assessment. Although the study concerns science rather than language learning, its outcome variable -- perceptions of the assessment environment, not merely test scores -- is precisely the washback-relevant measure that language-education research on AI feedback rarely reports, and its inclusion here extends the geographic and disciplinary base of the claim beyond the East Asian and higher-education contexts that dominate this chapter's other citations.

However, negative washback risks also emerge. Bibliometric and review studies warn that learners may begin to optimize performance for AI systems rather than developing communicative competence, particularly if assessment criteria privilege surface-level accuracy or fluency (Yang et al., 2025; Zhang et al., 2025). In such cases, assessment practices may narrow curricular focus and reinforce test-oriented learning behaviors.

From an assessment-theoretical standpoint, this highlights the importance of construct-aligned design and teacher mediation in AI-based assessment systems.

3.6.6 Toward Responsible AI-Mediated Language Assessment

Taken together, the literature suggests that AI has the potential to transform language assessment by enhancing reliability, scalability, and formative feedback. At the same time, it raises profound challenges for validity, fairness, transparency, and washback.

Assessment theory thus plays a gatekeeping role in AI-mediated language education, ensuring that technological innovation does not outpace pedagogical and ethical responsibility. Recent scholarship increasingly calls for *human-in-*

the-loop assessment models, where AI supports—but does not replace—professional judgment (Mateos-Blanco et al., 2024; Zhang et al., 2025).

In this sense, AI-mediated assessment is best understood not as an autonomous evaluative mechanism but as a socio-technical practice, embedded within institutional norms, pedagogical values, and theoretical commitments. This perspective prepares the ground for the integrated framework proposed in the concluding section of this chapter.

Pedagogical Companion for Chapter 3

Key Concepts: comprehensible input; interaction and negotiation of meaning; pushed output; noticing; sociocultural mediation; zone of proximal development; self-regulated learning; learner autonomy; TPACK and SAMR; posthumanism; validity, fairness, and washback.

Discussion Questions

1. Krashen's input hypothesis predates AI by four decades. In what specific ways does adaptive, on-demand input change—or fail to change—the theoretical picture?
2. Can interaction with an AI system constitute "negotiation of meaning" in the interactionist sense? Defend both answers before choosing one.
3. Apply the ZPD to an AI tutoring scenario: what would it mean for the system to work *within* a learner's ZPD rather than simply below or above it?
4. The chapter closes with assessment theory. Why must validity arguments change when the scorer is an algorithm rather than a rater?
5. Which theoretical lens in this chapter best explains a failure of AI-supported learning you have observed or can imagine?

Further Reading

- Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. Longman.
- Chapelle, C. A. (2001). *Computer applications in second language acquisition*. Cambridge University Press.
- Xi, X. (2010). Automated scoring and feedback systems: Where are we and where are we heading? *Language Testing, 27*(3).

PART II

AI in Language Teaching

Chapter 4. AI as a Teaching Assistant: Planning, Materials, and Tasks

4.1 Introduction

As we have already stated the role of Artificial Intelligence (AI) in foreign language education has undergone a significant conceptual and practical transformation over the past decade. Initially positioned as a form of peripheral technological support—often limited to automated drills, grammar checkers, or isolated computer-assisted language learning (CALL) tools—AI has increasingly evolved into an integrated teaching assistant that actively supports instructional planning, materials development, task design, and learner differentiation. This shift moves technology from a passive repository of information to an active "*co-pilot*" or "*co-pedagogical agent*" that augments instructional decision-making.

On the research agenda, recent bibliometric and systematic reviews confirm that AI-supported instruction has also emerged as one of the most dominant and rapidly expanding research clusters in applied linguistics and language education, particularly since 2018 (Kartal & Yeşilyurt, 2024; Jaleniauskienė et al., 2023; Yang et al., 2025). Rather than framing AI as a substitute for teachers, contemporary scholarship emphasizes its role as a "*more capable peer*" that facilitates cognitive growth through scaffolded support and social interaction (Zhang et al., 2025). Within this human-centered perspective, AI is understood not as an autonomous authority but as a mediating resource whose educational value depends on how it is integrated into instructional design, classroom interaction, and assessment practices (Edmett et al., 2024; Mohebbi, 2025).

4.2 Lesson Planning with AI

AI-supported lesson planning has become one of the most visible and practically influential applications of AI in foreign language teaching. Generative AI (GenAI) markedly enhances the efficiency and creativity of lesson planning by providing educators with "*idea scaffolding*" (Li et al., 2025). By automating and scaffolding aspects of instructional preparation, AI enables teachers to streamline "back-office" tasks—such as curriculum mapping, warm-up generation, and worksheet drafting—thereby allowing

educators to dedicate more time to student interaction and higher-order pedagogical concerns (Tan et al., 2025).

In practical terms, educators utilize Large Language Models (LLMs) to generate lesson plan templates, identify curriculum topics, and suggest activities. A prominent advancement in this area is the development of theory-grounded systems like the Culturally Responsive Lesson Planner (CRLP). This tool utilizes an Interactive Semi-Automated (ISA) prompt architecture that engages teachers in a structured dialogue to collect cultural and contextual parameters before generating a lesson plan, ensuring the pedagogy aligns with Culturally Responsive Pedagogy (CRP) and honors students' lived experiences (Nyaaba & Zhai, 2025). Further, AI supports student-centered models, such as the 5E instructional model (Engage, Explore, Explain, Elaborate, Evaluate), by acting as a creative partner that generates hooks, explanatory drafts, and complex group projects (Li et al., 2025).

However, the literature consistently emphasizes that effective lesson planning with AI requires active pedagogical mediation. While AI can produce coherent structural outlines, it often struggles to strategically integrate prior course content with future learning outcomes without explicit, high-quality prompting from the teacher (Tan et al., 2025). Consequently, effective collaboration depends on the teacher's "GenAI-TPACK" (Technological Pedagogical Content Knowledge), a framework necessary for educators to integrate these tools ethically and effectively (Belkina et al., 2025).

Empirical work on what teachers actually do when planning with AI, rather than on what the tools afford in principle, gives the mediation claim a more concrete footing. In a four-week quasi-experiment with 60 in-service teachers, GenAI-agent-supported reflection—whether self-directed or peer-directed—significantly increased both teachers' self-reflection efficacy and the depth of their instructional-design reflection relative to a conventional-technology control group, with no advantage for peer- over self-directed use (Cao et al., 2026). Time savings prove similarly real but conditional: across three K–12 case studies of generative-AI item-development tools, adoption reduced development time and cost while improving the accuracy and consistency of content alignment (Bolender et al., 2024). A multi-institution mixed-methods study of AI-driven personalized English instruction reports a parallel pattern at the level of outcomes—improved

motivation, participation, and language skills—while flagging unequal technology access and unmet teacher-training needs as the barriers that determine who benefits (Bernal Parraga et al., 2025).

4.2.1 From Idea Scaffolding to a Worked Example

The practical value of AI-assisted planning becomes clearest when examined at the level of a concrete lesson. Consider a 90-minute speaking lesson for a B1-level university EFL class on the theme of sustainable consumption. Working from the curriculum objective—learners can state and justify an opinion in a short discussion—the teacher first drafts the communicative outcome and the evidence of learning, and only then turns to the model for support. The division of labor matters: decisions about outcomes, sequencing, and assessment remain with the teacher, while the model contributes candidate materials and variation (Tan et al., 2025; Edmett et al., 2024).

In the preparation phase, the teacher asks the model for three opinion-gap scenarios calibrated to B1 vocabulary, each with a role card containing a position, two supporting arguments, and one counterargument to anticipate. The output is treated as a draft: the teacher removes an idiom cluster that exceeds the level, replaces a culturally distant example with one drawn from students' local context, and adds a pre-taught lexical set. During the lesson, the model plays no live role; the interaction is human. After the lesson, the teacher submits anonymized samples of learner discourse and asks for a frequency summary of hesitation phenomena and recurring lexical gaps, which informs the next lesson's focus. A planning cycle of this kind operationalizes the augmentation thesis of this book within a single teaching unit: preparation is accelerated, pedagogical judgment is exercised at every checkpoint, and the post-lesson analysis strengthens rather than replaces the teacher's diagnostic skill (Oliveira et al., 2025; Mohebbi, 2025).

Table 2 summarizes the cycle as a reusable planning pattern.

Phase	Teacher decision (retained)	AI contribution (delegated)	Quality checkpoint
Outcome setting	Communicative objective, assessment evidence	—	Alignment with curriculum
Material drafting	Selection among variants; cultural fit	Scenario drafts, role cards, lexical sets	Level check, bias check
Task calibration	Grouping, timing, interaction pattern	Differentiated versions of the same task	Cognitive load; step size
Enactment	All live instructional moves	—	—
Post-lesson analysis	Interpretation, next-step planning	Descriptive summary of learner output	Data privacy; anonymization

Table 2. A teacher-directed AI planning cycle for a language lesson.

4.2.2 Prompt Design as Pedagogical Reasoning Made Explicit

The quality of AI-generated planning support is largely a function of the instructional information encoded in the prompt (Tan et al., 2025). Effective prompts externalize the same variables that structure sound lesson design: learner profile, proficiency level, objective, constraints, and evaluation criteria. In this sense, prompt formulation is not a technical trick but pedagogical reasoning made explicit—which is why prompt quality correlates with teachers' broader instructional expertise and their GenAI-TPACK (Belkina et al., 2025; Nyaaba & Zhai, 2025).

The following template family, adaptable across levels and skills, illustrates the principle. Each template deliberately specifies role, learner context, task, constraints, and output format, and each ends with an instruction that keeps the teacher in an evaluative position.

- *Template 1—Leveled text generation.* "You are assisting an English teacher. Write a [180–220]-word text on [topic] for [CEFR B1] learners aged [17–19] in [country/context]. Use only vocabulary from [list/level]; recycle these target items at least twice each: [items]. Avoid idioms and culturally specific references that require background knowledge. Then list the ten least frequent words you used, so I can verify level fit."
- *Template 2—Text simplification with rationale.* "Simplify the following text to [A2]. Preserve the information structure and discourse markers. Output

two columns: the simplified text, and a note for each change explaining what was simplified and why. Do not change quoted speech. [text]"

- *Template 3—Interactive dialogue for role-play.* "Create a two-role dialogue ([customer]/[shop assistant]) of 12–16 turns practicing [making polite requests]. Embed these three target structures naturally: [structures]. Insert two communication breakdowns that require clarification requests. Provide a follow-up card with three 'what would you say if…' variations."
- *Template 4—Differentiated task set.* "From this reading passage, produce three task versions for the same comprehension objective: (a) scaffolded with sentence starters and a glossary, (b) standard, (c) extension with an inference and evaluation question. Keep all three answerable from the passage alone. [passage]"
- *Template 5—Feedback pre-screening.* "Here are anonymized learner sentences with a recurring error pattern. Classify the errors by type, order them by frequency, and propose one inductive noticing activity targeting the most frequent type. Do not correct the sentences themselves; I will decide what to address in class. [samples]"

Two design features of these templates deserve comment. First, each includes a verification affordance—a word list, a rationale column, an answerability constraint—that makes the output checkable rather than merely usable, addressing the documented tendency of teachers to over-accept fluent AI text (Anders & Speltz, 2025; Hooshyar et al., 2025). Second, none of the templates delegates an evaluative decision: what to teach, what to correct, and what counts as adequate remain outside the model's remit, consistent with the human-in-the-loop position developed throughout this book (Ho & Lee, 2025; Oliveira et al., 2025).

The specific case of AI-generated assessment items shows both sides of this ledger with unusual clarity. A case study of one teacher using AI to generate foreign-language exam questions found genuine gains in practicality, customization, and efficiency, but also documented challenges in readability, validity, and ownership (Kic-Drgas & Kılıçkaya, 2024). The ownership problem is easy to underestimate: an item a teacher has not personally drafted is an item whose fitness for a specific class the teacher must verify rather than simply trust, and the same case study's readability and validity concerns are exactly what such verification is meant to catch. The lesson generalizes beyond exams—wherever a template promises to save planning time by generating scoreable content, the saved time is only realized net of the verification labor the teacher then owes the artifact, labor

that Section 4.4.1's screening rubric is designed to make routine rather than improvised.

4.2.3 Co-Planning Dialogues: The Assistant as Thinking Partner

The templates of the previous section treat the model as a producer of artifacts. A second, complementary use treats it as an interlocutor in the planning process itself—a use closer to what experienced teachers do with a trusted colleague than to what they do with a resource bank. In a co-planning dialogue, the teacher does not request a finished plan; they think aloud about a design problem and instruct the system to respond with questions, alternatives, and objections rather than solutions. A representative opening move: "I am planning a first lesson on narrative tenses for a group that has repeatedly failed to transfer this grammar into free speech. Before proposing anything, ask me five questions a mentor teacher would ask about this group and my previous attempts." The model's questions—about proficiency profile, prior instruction, error patterns, task types already tried—externalize a diagnostic routine that novice teachers in particular have not yet internalized (Belkina et al., 2025).

Used this way, the assistant functions as a rehearsal space for pedagogical reasoning, and the interaction itself has professional development value independent of any artifact it produces. Theory-grounded planning tools point in the same direction: systems that engage teachers in structured dialogue about cultural and contextual parameters before generating anything produce plans that teachers modify more actively and adopt more critically than single-shot generation does (Nyaaba & Zhai, 2025). The design insight is that the *sequence* of the interaction matters: elicitation before generation preserves the teacher's authorship of the pedagogical problem, whereas generation before elicitation anchors the teacher on the machine's framing—a small architectural difference with large consequences for whose thinking leads.

Independent evidence supports treating this kind of engagement as a genuine site of professional learning rather than only a metaphor. A qualitative study of nine Chinese EFL teachers in a university-based professional learning community examined how sustained, collegial engagement with large language models shaped their developing digital literacy (Xue, 2024). The community setting looks like it matters as much as the tool: literacy built through joint reflection among colleagues sits closer to the

disagreement-engineering and peer scrutiny this section recommends than to any teacher's solitary use of a chat interface.

Two ground rules keep co-planning from degrading into outsourced planning. The first is that the dialogue ends before the plan is written: the teacher drafts the actual lesson from the clarified problem, optionally returning to the model for the materials-production tasks of *Section 4.3*. The second is that disagreement is engineered in: instructing the system to argue against one's chosen approach ("give the three strongest objections a skeptical colleague would raise") converts the model's fluency from a validation hazard into a stress test. Both rules operationalize, at the scale of a planning session, the intentional friction that Chapter 14 identifies as a design requirement for agentic systems: the productive difficulty is preserved on the human side of the exchange, where the learning it produces belongs to the teacher.

4.3 AI-Generated Texts, Dialogues, and Multimodal Input

The capacity of LLMs to produce context-aware, coherent, and multilingual content has revolutionized the creation of instructional materials, effectively democratizing content creation. Teachers can now generate bespoke texts, dialogues, and multimodal resources tailored to the immediate needs of their learners (Aldreabi et al., 2025).

- *Text and Dialogue Generation*: AI tools are used to create reading passages, simplify complex texts for lower proficiency levels, and generate example sentences containing specific vocabulary (Tan et al., 2025). Beyond static text, tools like ChatGPT and other conversational agents allow for the creation of dynamic, interactive dialogues. These systems can simulate role-plays or act as Socratic tutors, answering open-ended questions and providing students with a safe, low-anxiety environment to practice communication (Caccavale et al., 2025; Kohnke et al., 2025).
- *Algorithmic Narrativity*: AI-generated storytelling introduces "algorithmic narrativity," a practice where the learner and AI collaboratively shape a story. These narratives embed target vocabulary within emotionally resonant contexts to improve retention and reduce learner anxiety (Namaziandost & Çakmak, 2025).
- *Multimodal Input*: Recent advancements have expanded AI's utility beyond text. Generative AI can now produce audio and visual materials to support diverse learning styles, such as "audio-learning modules" that summarize textbook content (Jafarian & Kramer, 2025). Fine-tuned

multimodal LLMs also enable systems to evaluate "picture-cued writing tasks," where students describe real-life scenarios depicted in images. These systems provide immediate, adaptive feedback by analyzing the connection between visual stimuli and the student's textual output (Zhuang et al., 2025).

- *Conversational Agents (CAs):* Tools like Storio, an e-book integrated with a bilingual agent named "Mia," emulate high-quality adult-child conversations. These agents provide differentiated, context-aware feedback and use strategies like "recasts" and "translanguaging" to support language acquisition (Xiao et al., 2025).

Evidence on how learners actually experience these AI-generated dialogues qualifies the promise somewhat. A study of 100 international students found that ChatGPT showed significant potential for personalized multilingual learning support and cultural comprehension, though interface friendliness and interaction naturalness still needed improvement (Yu et al., 2024). The gap is instructive for materials design: a dialogue can be linguistically well-formed and pedagogically well-targeted and still fail as an interaction, because naturalness is a property of the exchange's timing, repair, and register rather than of any single generated turn, and is correspondingly harder for a teacher to screen for than a factual or grammatical error.

Scholars caution that AI-generated content must be subjected to critical pedagogical evaluation regarding factual inaccuracies, cultural bias, and pragmatic oversimplification (Edmett et al., 2024; Faiz & Hazilan, 2024).

A concrete instance of what such caution catches: a qualitative study of sixth-semester Indonesian EFL students using AI translation tools found that literal idiom transfer and register mismatches persisted in their output even after AI assistance, and were reduced only once guided post-editing was introduced (Budiyanto et al., 2026). The finding matters for materials generation specifically because idiom and register errors are difficult for a non-expert reader—including many learners—to detect; a text can read as fluent and still misrepresent how the target language is actually used in the register it depicts. This is precisely the failure mode a screening pass must be trained to catch, since fluency and accuracy are independent properties of generated text and only one of them is visible on a casual read.

4.3.1 Working Multimodally: From Generated Text to Generated Voice, Image, and Interaction

The materials workflows described above extend naturally beyond text, and it is at the multimodal level that AI assistance begins to alter what a single teacher can feasibly produce. A listening lesson illustrates the point. Traditionally, a teacher wanting level-appropriate listening material on a specific topic faced a choice between published materials that rarely fit the local syllabus and authentic recordings that rarely fit the level. A current workflow inverts this: the teacher drafts or generates a level-controlled script using the constraints of Template 1, has it voiced by a speech-synthesis system in more than one accent, and—critically—generates comprehension and noticing tasks from the same script, so that task and text remain aligned by construction (Jafarian & Kramer, 2025). Robust speech recognition models such as Whisper close the loop in the other direction, producing usable transcripts of authentic recordings that the teacher can then simplify, gloss, or convert into tasks (Radford et al., 2022). The pedagogical gain is not merely convenience; it is the ability to hold topic, level, accent, and task type constant while varying exactly one of them—a degree of design control over listening input that classroom teachers have never previously had.

Comparable workflows now exist for visually anchored tasks. Fine-tuned multimodal models can evaluate picture-cued writing, analyzing the relationship between an image and the learner's description of it (Zhuang et al., 2025); the same capability, run in reverse, lets teachers generate image sets calibrated to target vocabulary and then attach automated formative feedback to the resulting tasks. Conversational agents embedded in reading materials—such as e-book agents that conduct dialogic reading, recasting learner utterances and offering translanguaging support—show how interaction itself becomes a designed material rather than an unpredictable by-product of classroom life (Xiao et al., 2025).

Three cautions keep this multimodal expansion pedagogically honest. First, synthesis quality varies across languages and varieties, and a voice that sounds fluent may still model unnatural prosody; audition before use is non-negotiable. Second, generated images inherit the representational biases discussed in *Section 4.4.1*, and image sets should be audited with the same systematic variation of names, settings, and depicted roles. Third, multimodal abundance tempts teachers toward input maximalism—more

channels, more stimulus, more novelty—when the learning-design question remains what it has always been: which mode, in which combination, serves this objective for these learners. The multimodal workflow, like the textual one, ends with a teacher's judgment, not a generator's output.

A field-defining review of second-language-acquisition scholarship synthesizes more than a decade of work on multimodal composing—learners drawing on audio, visual, gestural, and spatial resources alongside text—establishing that the pedagogical case for combining modes predates, and does not depend on, generative AI (Lim & Kessler, 2023). What AI principally changes is the tooling available for this kind of composing, not the underlying rationale for it. A more recent review of digital multimodal composition in second-language writing identifies a persistent gap between what the research documents as beneficial and what actually happens in classrooms—a reminder that AI tools will not by themselves close a gap that has more to do with curriculum, assessment norms, and teacher preparation than with tool availability (Yu et al., 2024). For listening and speaking specifically, generative AI's capacity to sustain spoken dialogue and interpret prosodic and interactional cues remains an emerging capability whose classroom-scale limitations, alongside its possibilities, are only beginning to be mapped systematically (Goh & Aryadoust, 2025).

4.3.2 Choosing the Toolset: Criteria Before Catalogues

Discussions of AI-assisted materials development often begin with inventories of tools; this chapter has deliberately begun with workflows, because tools change quarterly while workflows persist. Nevertheless, teachers and institutions must make selection decisions, and four criteria organize them more durably than any catalogue.

1) *Data handling*: where do learner texts, voices, and images go, under what retention terms, and does the tool offer an education-specific deployment with contractual data protection—the governance baseline of Chapter 12 applied at the point of purchase.
2) *Language coverage*: performance in and about the target language, including varieties; a tool excellent for English may be poor for the languages a program teaches, and vendor benchmarks rarely say so.
3) *Pedagogical configurability*: whether the teacher can constrain level, length, register, and error tolerance, or whether the tool's defaults are the only pedagogy on offer.

4) *Exportability*: whether outputs leave the platform in formats the institution's existing workflow can use, or whether the tool's convenience is purchased with lock-in. Notably absent from this list is output brilliance, which varies less across mainstream systems than marketing suggests, and which matters less than the surrounding workflow in any case (Weng & Chiu, 2023). A modest model inside a disciplined editorial workflow outperforms a frontier model inside an undisciplined one.

4.4 Task Design and Differentiation

AI technology facilitates a shift from "one-size-fits-all" instruction to precision education through dynamic task adaptation and automated generation (Hooshyar et al., 2025). By leveraging student data, AI systems can classify learners into different achievement clusters and automatically generate tailored learning scenarios for each group.

- *Differentiation Strategies*: For high-achieving students, AI can generate advanced topics and critical-thinking quizzes, while for students requiring support, it can produce foundational summaries and drill-and-practice items (Pesovski et al., 2025). Adaptive Learning Platforms (ALPs) utilize algorithms to dynamically adjust the sequence and difficulty of materials based on real-time performance. By identifying knowledge gaps, these platforms can redirect students to review foundational concepts before progressing to complex tasks like argumentative writing (Tan et al., 2025b; Li, 2025).
- *Automated Question Generation (AQG)*: AI tools can transform plain texts into formative practice questions, such as multiple-choice questions (MCQs), fill-in-the-blanks, and matching exercises. These tasks align with the "learning by doing" philosophy and provide immediate feedback essential for retrieval-based learning (Erdős et al., 2025; Chan et al., 2025).
- *Interactive and Scaffolding Tasks*: AI facilitates "response-contingent scaffolding," where students engage in writing tasks and receive immediate, adaptive feedback on grammar and structure (Kinder et al., 2025). In collaborative settings, AI agents can mediate interactions, providing preliminary feedback that allows the human teacher to focus on higher-order emotional and motivational support (Guan et al., 2025).

The differentiation claims above-described capability; evidence on realized outcomes is more mixed and worth stating before the mechanism is taken as settled practice. A PRISMA-guided review of 22 empirical studies drawn from 189 screened records found that AI tools produced significant gains across reading, writing, listening, speaking, and vocabulary in school EFL settings, but that digital-divide and teacher-readiness challenges

recurred across the studies examined (Kundu & Bej, 2025). At the level of a single curriculum, Saudi university EFL/ESP writing students using ChatGPT showed significant gains in grammar accuracy, vocabulary, creativity, and critical thinking, despite technical and ethical implementation challenges (Elmotri et al., 2025). Differentiation and curriculum-level AI use, on this evidence, deliver real gains that arrive bundled with real preconditions and real friction.

4.4.1 Quality Control for AI-Generated Materials

Because generative systems produce fluent output regardless of whether it is accurate, level-appropriate, or culturally suitable, materials evaluation becomes a distinct professional skill rather than an optional afterthought (Edmett et al., 2024; Faiz & Hazilan, 2024). The evaluation rubric in *Table 3* consolidates criteria recurring across the literature into a screening instrument that teachers can apply in under five minutes per artifact. It is deliberately criterion-referenced rather than holistic: a single failing criterion (for example, factual inaccuracy in a reading text) is sufficient grounds for revision, whatever the artifact's other merits.

Criterion	Guiding question	Typical failure mode	Action if failed
Factual accuracy	Are all checkable claims correct?	Plausible fabrication, outdated facts	Verify against sources; regenerate with corrections
Level fit	Does lexis/syntax match the target level?	Hidden low-frequency items; long noun phrases	Run frequency check; simplify
Objective alignment	Does the task elicit the target behavior?	Attractive but off-objective activity	Re-prompt with explicit objective
Cultural appropriateness	Are contexts inclusive and locally intelligible?	Western-centric defaults, stereotyped roles	Localize; replace examples
Bias and representation	Who is visible, in what roles?	Skewed gender/ethnicity roles	Rebalance; audit systematically

Criterion	Guiding question	Typical failure mode	Action if failed
Pragmatic authenticity	Is the language socially plausible?	Over-polite, uniformly formal dialogue	Inject register variation
Answerability/validity	Can tasks be completed from given input?	Questions requiring outside knowledge	Test-solve before use

Table 3. A screening rubric for AI-generated language teaching materials (synthesized from Edmett et al., 2024; Nyaaba & Zhai, 2025; Hooshyar et al., 2025; Faiz & Hazilan, 2024).

Bias auditing deserves particular emphasis in language education, where materials model social identity as much as linguistic form. Research on dialect prejudice in language models shows that systems can reproduce covert biases against non-standard varieties even when overt outputs appear neutral (Hofmann et al., 2024)—a finding directly relevant to teachers generating listening scripts, character dialogues, or "example learner" texts. A practical audit habit is to regenerate the same artifact with varied character names, settings, and language varieties, and to inspect whether quality, register, or role assignment shifts systematically.

Automated question generation supplies a further, quantifiable instance of the same pattern. A design-and-development study built a ChatGPT-based automatic question-generation system around a prompt-engineering protocol for English-language education and found statistically significant question validity overall, though some question types were not optimal for ChatGPT to generate (Lee et al., 2024). In the light of *Table 4.2*, the result cuts against a single blanket verdict on AI-generated assessment items: validity is not a fixed property of the generator but varies by item type, which means a screening rubric applied uniformly across a bank of AI-generated questions will under-detect failures concentrated in the types the model handles poorly.

4.4.2 A Differentiation Decision Matrix

Differentiation is the area where AI assistance yields the largest measurable time savings, but also where indiscriminate use most quickly produces pedagogically incoherent classrooms (Pesovski et al., 2025; Hooshyar et al., 2025). The decision matrix in Table 4 links common learner-profile signals

to differentiation moves that AI can draft and the teacher validates. The matrix embodies a constraint emphasized across the differentiation literature: vary the pathway and the scaffolding, not the objective. All learners work toward the same communicative outcome; what differs is support density, input complexity, and product format.

Learner signal	Differentiation move	AI draft task	Teacher validation focus
Reads below level	Graded input	Simplified parallel text + glossary	Meaning preserved; no infantilization
Completes early, accurate	Extension	Inference/evaluation questions; register transformation	Genuine depth, not more volume
High anxiety in speaking	Rehearsal scaffold	Private dialogue practice script with hints	Low-stakes framing; exit ramp to live interaction
Uneven vocabulary retention	Recycling set	Spaced retrieval items from personal error log	Item quality; interval logic
Strong oral, weak written	Modality bridge	Transcribe-and-upgrade task from learner's own speech	Ownership; visible progress

Table 4. Differentiation decision matrix: signals, moves, and validation focus.

Case Box 4.1—Differentiated instruction with mobile technology in Turkish classrooms

The professional skills that AI-assisted differentiation demands—diagnosing learner profiles, designing tiered tasks, and validating technology-drafted materials—predate generative AI. In a line of research on technology-supported differentiation in Türkiye, Şimşek and Can (2020) documented teachers' use of tablets to differentiate content, process, and product according to learners' readiness, interests, and needs. Teachers judged the technology a useful implementation tool for differentiation, yet consistently pointed beyond the device itself: realizing a differentiated curriculum required deliberate organization, supportive attitudes, and integration into existing practice. A subsequent multi-country study of 26 teachers designing and implementing mobile-supported differentiation activities within an

Erasmus+ project (Şimşek, Can, & Küçük, 2025) reinforced this picture: differentiated activities increased motivation, participation, interaction, and cooperation, but their success depended on infrastructure, sustained training, and—notably—peer learning among the teachers themselves. In the context of the present chapter, these findings transfer directly to the GenAI era: AI planning assistants will differentiate instruction only as well as the diagnostic routines, institutional support, and collaborative professional culture into which teachers embed them.

Constraints of this kind are not specific to mobile technology or to the Turkish context. An exploratory study of 15 ESP teachers in Moroccan higher education found that integrating a 6Cs framework with inclusive AI-supported learning design was hindered by low student motivation, limited materials, and time constraints (Belkbir, 2024)—the same triad of institutional and motivational preconditions identified above, recurring in a different educational system. Differentiation frameworks built around AI, in short, travel across contexts more easily than the conditions for realizing them do.

4.4.3 From Lesson to Syllabus: Curriculum-Level Planning with AI

The planning support examined so far operates at the scale of the single lesson or task. A less visible but arguably more consequential application operates at the scale of the course. Curriculum work—mapping a syllabus against a proficiency framework, sequencing objectives, ensuring that vocabulary and structures are recycled rather than encountered once and abandoned—is precisely the kind of large, structured, consistency-demanding task at which generative systems assist well and human working memory does not (Tan et al., 2025; Li et al., 2025).

Concretely, teachers and program coordinators now use LLMs for four curriculum-level operations. The first is *framework alignment*: submitting an existing syllabus and asking for a mapping of each unit's objectives onto CEFR descriptors, with discrepancies flagged rather than silently repaired—the output is an audit, and the teacher adjudicates it. The second is *recycling analysis*: extracting the target lexis of each unit and tracing where, if anywhere, each item recurs in later units; the resulting recycling map makes visible the one-exposure vocabulary that course books quietly accumulate, and spaced re-encounters can then be designed deliberately—the same logic that drives competency-recommendation systems at the institutional scale (Junaidi & Wahyono, 2025). The third is *coherence*

checking: asking whether the skills load of each week matches the course's stated weighting, and whether assessment tasks sample the syllabus representatively or cluster around what is easiest to test. The fourth is *variant generation*: producing a parallel syllabus strand for a different intensity (a twice-weekly evening cohort alongside a daily intensive one) while preserving objectives—a task whose manual version is so laborious that it is often simply not done.

A brief worked example makes the recycling operation concrete. A coordinator submits a twelve-unit B1 syllabus and receives a recycling map showing that 41 of 120 target items appear exactly once, concentrated in Units 3 and 7—the two units built around topical vocabulary (environment, health) rather than high-frequency verbs. The map itself decides nothing; what follows does. The team schedules three-minute retrieval openers in Units 5, 8, and 11 that deliberately re-encounter the orphaned items in new contexts, converts two of Unit 7's single-exposure sets into the lexical basis of a Unit 10 speaking task, and—where an item is judged not worth recycling—removes it from the assessed list rather than pretending one exposure taught it. The entire intervention costs one planning meeting. What the AI contributed was neither the pedagogy nor the decisions but the visibility: a complete, accurate account of a design property that no member of the team could have held in mind, produced at a cost low enough that the audit actually happened. This is the general shape of high-value curriculum assistance—machine visibility feeding human redesign—and it scales from a single course to a program review without changing character.

The systematic-review evidence on intelligent CALL gives this curriculum-level use a theoretical warrant: learning outcomes in AI-supported language instruction depend heavily on instructional design quality, not on tool sophistication (Weng & Chiu, 2023). A model that helps a teacher see the design—its gaps, its dead vocabulary, its misweighted assessment—intervenes at the point of highest leverage. But the same evidence base grounds a warning. Generative systems optimize local plausibility: each generated unit looks coherent, while the course-level progression may drift, repeat, or contradict itself across units (Hooshyar et al., 2025). Curriculum work with AI therefore inverts the usual quality-control direction: at lesson level, the teacher checks the details of a plausible whole; at course level, the teacher must supply the whole and use the system to check details at a scale no human reviews reliably. Holding that distinction is what separates

AI-assisted curriculum design from a folder of AI-generated lessons in search of a course.

A controlled comparison from vocational language-service training illustrates the same design principle operating at program scale. In a study spanning 120 students in an AI- and big-data-supported 'Belt and Road' language-service curriculum against 120 students in a conventional comparison group, together with 100 industry surveys and 10 practitioner interviews, the AI-supported curriculum produced significantly higher learning efficiency and translation quality than the conventional one (Cheng, 2025). The result is consistent with this chapter's claim that AI assists curriculum work best as a design-visibility instrument rather than a generator of finished courses: the comparison group did not lack technology so much as the systematic redesign that the technology's diagnostic output made possible.

4.5 Risks of Deskilling vs. Pedagogical Augmentation

The integration of AI as a teaching assistant introduces a critical tension: the potential for pedagogical augmentation versus the risk of professional deskilling.

- *Pedagogical Augmentation*: When implemented effectively, AI functions as an "exoskeleton" or "co-teacher" that augments human expertise rather than replacing it (Tan et al., 2025a). The *DRIVE* framework (Directive Reasoning Interaction and Visible Expertise) emphasizes that students and teachers should maintain agency by critically steering AI interactions. In this "collaborative intellectual partnership," the human user provides strategic direction while the AI handles the cognitive load of content processing (Oliveira et al., 2025). Experimental studies have shown that AI-integrated blended learning approaches allow teachers to cover substantially more content and dedicate increased time to personalized student interaction (Niloy et al., 2025).
- *Risks of Deskilling*: Conversely, there is a valid concern regarding "*deprofessionalization*" or "*cognitive offloading*," where educators and students may uncritically accept AI-generated content (Anders & Speltz, 2025; Hooshyar et al., 2025). Excessive reliance on AI for lesson planning and assessment may lead to skill atrophy, where teachers lose the ability to critically evaluate curriculum design or diagnose student needs independently (Rahimi, 2025). What is more, widespread use of AI for generating materials raises concerns about the homogenization of content and the loss of the teacher's unique instructional voice (Lyu et al., 2025).

- *The "Black-Box" Effect and Bias*: Many AI models remain "black boxes," providing outputs without transparent reasoning, and may reproduce cultural biases found in Western-centric training data (Hooshyar et al., 2025; Nyaaba & Zhai, 2025). To mitigate these risks, a *Human-in-the-Loop (HITL)* approach is essential. Teachers must maintain pedagogical authority, critically evaluating AI outputs for factual accuracy and ethical alignment (Ho & Lee, 2025). Cultivating "AI literacy" is vital to ensuring that technology serves as a catalyst for creative inquiry while maintaining human agency (Anders & Speltz, 2025; Jin et al., 2025).

Direct empirical evidence on this question—whether AI-assisted planning strengthens or erodes teacher expertise—remains scarce, but what exists points in more than one direction at once, which is itself informative. A phenomenographic study of 15 Iranian EFL teachers found that AI use reshaped their professional identity across six themes, including a transformation of role, closer pedagogical alignment, and increased self-reflective, technologically fluent practice (Ghiasvand & Seyri, 2025). Read against a simple deskilling narrative, that last theme cuts the other way: these are teachers reporting more reflective engagement with their own practice, not less. This does not settle the deskilling question so much as relocate it—the same teachers who report richer self-reflection remain, on other evidence already cited in this chapter, exposed to the over-acceptance risk that unverified fluency creates (Anders & Speltz, 2025). Whether augmentation or erosion predominates looks less like a fixed property of the technology than an outcome of whether reflection and verification are actually built into a teacher's routine of use.

4.5.1 Assessment-Aware Task Design

A teaching assistant that can draft any text can also complete any homework. Task design in AI-rich classrooms therefore has to anticipate learner access to the same tools the teacher uses. One productive response is to design tasks in which AI use is either structurally irrelevant (in-class, interaction-based performance) or explicitly staged: learners generate a draft with an LLM and are then assessed on their ability to critically critique, edit, and reference the AI-generated artifact (Perkins et al., 2024a). Frameworks such as the AI Assessment Scale make the permitted level of AI involvement explicit per task, replacing blanket prohibition with graduated transparency (Perkins et al., 2024a). Chapter 11 develops the assessment-integrity dimension in depth; the planning-stage implication here is that

every generated task should carry a deliberate decision about disclosed AI use—for the teacher and for the learner alike.

A qualitative case study of nine Ontario EAP instructors found that their engagement with generative AI was shaped by pedagogical viability and institutional policy, alongside concerns over diminished critical thinking and the marginalization of multilingual learners (Haidar & Tassis, 2025). Assessment-aware task design of the kind sketched above is one direct response to the first concern; the second is a matter of whose language variety AI-generated dialogue and text treat as the default norm—a question *Section 4.4.1*'s bias-auditing habit is built to surface but not, by itself, to resolve.

4.5.2 The Teacher as Editor-in-Chief: A Professional Workflow, Its Conditions, and Its Costs

The threads of this chapter—planning cycles, prompt design, materials screening, differentiation matrices, curriculum audits—converge on a recognizable professional identity. The teacher working productively with AI assistance operates as an editor-in-chief: commissioning drafts against a specification, rejecting or reworking most of what arrives, retaining unilateral authority over what is published to learners, and remaining accountable for every word of it. The metaphor is exacting rather than flattering. An editor who publishes unread copy is negligent; an editor who writes everything personally has no staff worth the name. The professional skill lies in the specification and the verification, and both are teachable (Edmett et al., 2024; Kohnke et al., 2025).

The available evidence suggests that, under this working model, the time economics are real but conditional. Experimental work on AI-integrated blended instruction reports that teachers cover more content while devoting more time to individual student interaction (Niloy et al., 2025)—the reallocation, not the raw saving, is the pedagogically meaningful outcome, and it echoes the augmentation logic of the *DRIVE* framework, in which the human supplies strategic direction while the system absorbs processing load (Oliveira et al., 2025). Reviews of AI and learner independence point the same way: the value of automated support materializes when it is embedded in a deliberately designed pedagogical structure, and dissipates when it substitutes for one (Mohebbi, 2025).

The conditions, however, are institutional as much as individual. The mobile-differentiation research summarized in *Case Box 4.1* found that technology-supported differentiation succeeded where infrastructure, training, and peer collaboration among teachers were present, and stalled where they were absent (Şimşek, Can, & Küçük, 2025)—and there is no reason to expect AI-assisted planning to behave differently. An editor-in-chief needs an editorial infrastructure: time allocated for verification rather than only for production; shared repositories where colleagues' validated prompts, screened materials, and audit findings accumulate instead of evaporating; and professional development that treats prompt specification and output evaluation as assessable competencies, in line with emerging teacher-competency frameworks for AI (Miao & Cukurova, 2024).

The institutional risk is not hypothetical, and evidence from adjacent higher-education contexts quantifies it. A cross-national analysis of industry-academia AI micro-credential programs in the United States, the European Union, and Singapore found a 40 percent decline in faculty autonomy under market-driven governance arrangements, with instructors repositioned from designers of instruction toward moderators of externally produced content (Soomro et al., 2026). Nothing in language teaching specifically guarantees immunity from the same dynamic: an editor-in-chief model of AI-assisted planning presupposes that the institution treats the teacher's edited judgment as final, and the finding above shows what happens to that model once procurement and governance decisions are made elsewhere and the teacher's role is redefined around them rather than by them.

Named honestly, the costs complete the picture. Verification labor is real labor, and institutions that celebrate AI's generative speed while budgeting no time for editorial control are commissioning unread copy at scale. Skill maintenance requires deliberate practice: an editor who never writes loses the craft judgment that editing depends on, which is why the deskilling risks of *Section 4.5* are managed, not escaped, by the editorial model—teachers need protected occasions to plan and write unassisted, precisely so that their assisted judgments remain expert. And the model's authority claim must be institutional fact rather than rhetoric: where learning platforms or administrators can push AI-generated content directly to learners over a teacher's judgment, the editor-in-chief has become a proofreader, and the augmentation thesis of this book has been quietly abandoned. The *PAGE* framework

of *Chapter 13* supplies the test: pedagogy-first commissioning, agency-preserving authority, governance-bound disclosure, and equity-oriented auditing, applied to the everyday workflow of getting Tuesday's lesson ready.

4.6 Conclusion

This chapter has framed AI as a teaching assistant that supports lesson planning, material generation, task design, and learner differentiation in foreign language education. The evidence demonstrates that AI's pedagogical value lies in *augmentation* rather than automation. While AI enhances efficiency, personalization, and learner engagement, its effectiveness ultimately depends on informed *pedagogical mediation*. The central challenge for language educators is not whether to use AI, but how to integrate it responsibly—ensuring that human expertise, ethical judgment, and instructional intentionality remain at the core of foreign language teaching practice (Edmett et al., 2024).

The chapter's argument can be stated as a single professional norm: nothing generated is taught unread. Around that norm, the preceding sections have assembled the working apparatus—planning cycles that fix where machine contribution enters and where teacher decision rules; prompt families whose verification affordances make outputs checkable; screening rubrics and bias audits that convert quality from an impression into a procedure; differentiation and curriculum routines that scale the teacher's design intelligence rather than replace it; and an editorial identity that names both the authority and the accountability involved. None of this apparatus is technologically fragile: it survives model upgrades, vendor churn, and the arrival of the agentic systems examined in *Chapter 14*, because it is anchored in instructional design rather than in any tool's current capabilities. What it does not survive is institutional neglect—the withdrawal of verification time, the erosion of teacher authority over published content, the substitution of procurement for pedagogy. The teaching assistant of this chapter's title, in short, is only as good as the teacher's working conditions permit it to be, which is why the professional and policy arguments of *Chapters 6* and *12* are not adjacent to the materials question but constitutive of it.

Pedagogical Companion for Chapter 4

Key Concepts: AI as teaching assistant; idea scaffolding; prompt design; GenAI-TPACK; materials evaluation; differentiation; automated question generation; deskilling; human-in-the-loop.

Discussion Questions

1. Using the planning cycle in Table 4.1, identify which decisions in your own lesson preparation you would delegate to AI—and which you would refuse to delegate, and why.
2. Rewrite one of the prompt templates in Section 4.2.2 for your own teaching context. What context-specific information did you need to add for the output to be usable?
3. Apply the screening rubric (Table 4.2) to a real AI-generated text. Which criterion failed first?
4. The chapter warns of a "homogenization of content" and loss of the teacher's instructional voice. Is this a real pedagogical loss or professional nostalgia? Argue with reference to learner outcomes.
5. Design one assessment task for your learners in which AI use is structurally irrelevant, and one in which it is explicitly staged. Which was harder to design?

Further Reading

- Edmett, A., Ichaporia, N., Crompton, H., & Crichton, R. (2024). *Artificial intelligence and English language teaching* (2nd ed.). British Council.
- Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). The Artificial Intelligence Assessment Scale (AIAS). *Journal of University Teaching and Learning Practice, 21*(6).
- Weng, X., & Chiu, T. K. F. (2023). Instructional design and learning outcomes of intelligent computer assisted language learning. *Computers and Education: Artificial Intelligence, 4*.

Chapter 5. AI-Supported Teaching of Language Skills

5.1 Introduction: From Skill Isolation to Intelligent Skill Ecology

The teaching of language skills has traditionally been organized around discrete domains—reading, listening, writing, and speaking —often treated as separable components within language curricula. Although communicative and integrated-skills approaches have long been advocated, institutional constraints such as class size, time limitations, and assessment pressures have frequently restricted opportunities for sustained, individualized skill development. The integration of Artificial Intelligence (AI) into language education introduces a structural shift away from isolated skill instruction toward what may be conceptualized as an *intelligent skill ecology*, in which learners engage dynamically with language input, output, feedback, and revision across multiple modalities and skills.

Recent bibliometric and systematic reviews consistently identify language skills instruction as the most dominant application cluster in AI-mediated language education research, particularly since the emergence of generative AI and advanced speech technologies (Kartal & Yeşilyurt, 2024; Yang et al., 2025; Zhang, Hussin, & Abd Majid 2025). Unlike earlier generations of Computer-Assisted Language Learning (CALL), contemporary AI systems integrate natural language processing (NLP), machine learning, automatic speech recognition (ASR), and multimodal analytics, enabling instructional support that operates simultaneously across linguistic, cognitive, and affective dimensions of learning.

Above all, current scholarship rejects deterministic narratives that position AI as a replacement for teachers. Instead, AI is framed as a *pedagogical amplifier* that expands instructional capacity while preserving the teacher's central role in scaffolding, ethical mediation, and meaning making (Edmett et al., 2024; Mohebbi, 2025). Against this backdrop, this chapter synthesizes current research on AI-supported teaching across four core language skills: writing, speaking and pronunciation, reading and vocabulary development, and listening and multimodal comprehension.

5.2 AI-Supported Reading and Vocabulary Development

5.2.1 Adaptive Reading Systems

AI-supported reading instruction increasingly relies on *adaptive reading systems* that dynamically align texts and tasks with learners' proficiency levels. These systems analyze linguistic features such as lexical density, syntactic complexity, cohesion markers, and discourse structure in order to calibrate reading materials to individual learners' needs. Unlike static graded readers or uniform classroom texts, adaptive reading environments continuously adjust input based on learner performance, comprehension patterns, and reading behavior.

Bibliometric analyzes identify reading comprehension and vocabulary development as *core thematic clusters* within AI-mediated language learning research, reflecting sustained scholarly attention to receptive skills as foundations for overall language proficiency (Jaleniauskienė et al., 2023; Yang et al., 2025). Empirical studies consistently demonstrate that adaptive reading systems enhance comprehension accuracy, reading speed, and inferencing ability by maintaining an *optimal balance between challenge and accessibility,* this condition is widely recognized as essential for effective input processing.

Direct outcome evidence for adaptive and AI-mediated reading support, though still thinner than the corresponding evidence for writing, is beginning to accumulate. A randomized controlled trial compared parent-led and LLM-powered bilingual conversational-agent dialogic reading using children's e-books, testing effects on EFL comprehension and conversational skills in young learners (Xiao et al., 2025)—extending adaptive-reading research, concentrated heavily on secondary and university learners, down into early childhood and into a dialogic rather than purely text-adaptive format. A PRISMA-guided review of 22 empirical studies, drawn from 189 records screened, found that AI tools produced significant gains across reading, writing, listening, speaking, and vocabulary specifically in school-level EFL settings, though digital-divide and teacher-readiness challenges persisted (Kundu & Bej, 2025).

From a cognitive perspective, adaptive reading environments support comprehension by regulating *cognitive load*. When texts are appropriately calibrated, learners are less likely to experience overload caused by excessive

unfamiliar vocabulary or syntactic complexity. At the same time, adaptive systems avoid oversimplification by gradually increasing difficulty as learners demonstrate readiness. This fine-tuned modulation of input facilitates sustained engagement and supports the development of fluent reading processes rather than fragmented decoding strategies.

A further pedagogical advantage of AI-supported reading lies in its capacity to provide *real-time comprehension support.* Many adaptive systems integrate features such as predictive glosses, summarisation prompts, comprehension checks, and visual scaffolds that respond to learner interaction. These tools enable learners to resolve comprehension breakdowns autonomously, reducing dependence on teacher intervention and promoting strategic reading behaviors. Over time, learners develop greater awareness of how to deploy strategies such as skimming, scanning, inferencing, and monitoring comprehension.

Significantly, adaptive reading systems also generate rich learner data, allowing for the analysis of reading trajectories across time. This longitudinal visibility supports diagnostic insight into learners' strengths and difficulties, enabling more targeted pedagogical support. However, the literature cautions that adaptivity alone does not guarantee deep comprehension. Without pedagogical framing, learners may engage superficially with texts, prioritising task completion over interpretive depth. As a result, scholars emphasize the integration of adaptive reading tools within *strategy-based and meaning-oriented reading instruction*, where teachers guide learners to reflect on content, purpose, and interpretation.

5.2.2 Vocabulary Acquisition and Retention

Vocabulary development represents one of the most consistently and robustly supported outcomes of AI integration in language education. AI-driven tools facilitate vocabulary learning through a combination of *contextualization, repetition, and personalization*, enabling learners to encounter lexical items across varied contexts and modalities. Dynamic glossing, usage-based examples, semantic mapping, and spaced repetition algorithms work together to promote both immediate comprehension and long-term retention.

Empirical studies consistently demonstrate that AI-supported vocabulary tools outperform traditional dictionary-based approaches in fostering

lexical breadth and depth (Ngo, 2024). Rather than presenting isolated word meanings, AI systems embed vocabulary within authentic or semi-authentic contexts, allowing learners to observe collocations, register variation, and pragmatic usage. This contextual richness supports deeper lexical processing and reduces the likelihood of inert vocabulary knowledge that cannot be readily activated in communication.

This advantage is consistent with the broader meta-analytic pattern noted in section 5.2: across a synthesis of 117 effect sizes from 46 EFL studies, vocabulary showed the strongest overall AI-associated gains of the four skills this chapter surveys, while speaking showed the weakest (Torres & Kahveci, 2025)—plausibly because vocabulary is more tractable to discrete-item measurement and spaced-repetition intervention than the holistic, multidimensional construct of oral proficiency. Quasi-experimental evidence bears this out: in a study of 200 language learners, chatbot and virtual-assistant use improved vocabulary retention, fluency, motivation, and autonomy while reducing language anxiety relative to conventional instruction (Jara Chiriboga et al., 2025), and a literature-based analysis of generative AI in university EFL learning likewise identifies vocabulary development as the domain where generative tools most consistently enhance personalization, autonomy, and motivation while reducing anxiety (Chetveryk, 2026).

From a cognitive standpoint, AI-mediated vocabulary learning aligns with principles of *distributed practice and retrieval-based learning*. Spaced repetition algorithms schedule encounters with lexical items at optimal intervals, reinforcing memory consolidation and reducing forgetting. At the same time, adaptive systems prioritise words based on learner performance, frequency of exposure, and communicative relevance, thereby maximising learning efficiency. Such personalized sequencing would be difficult to achieve through manual instruction alone, particularly in large or heterogeneous classrooms.

AI-supported vocabulary instruction also plays a significant role in cultivating *learner autonomy and motivation*. Studies indicate that learners who engage with AI-driven vocabulary tools report increased control over learning pace, content selection, and review strategies, contributing to heightened motivation and sustained engagement (Mohebbi, 2025).

This autonomy is particularly valuable in extensive reading contexts, where vocabulary growth is incremental and requires long-term commitment.

However, the literature also highlights potential limitations. Overreliance on automated glosses may discourage inferencing strategies, and excessive focus on individual word learning may fragment textual comprehension. In turn, researchers advocate pedagogical approaches that integrate AI-supported vocabulary learning with *meaning-focused reading tasks*, encouraging learners to move fluidly between word-level attention and discourse-level understanding.

Crucially, vocabulary development in AI-supported environments is most effective when positioned as part of a *reading–lexicon synergy*, rather than as a standalone objective. Adaptive reading systems that integrate vocabulary tracking allow learners to encounter newly learned words repeatedly across texts, reinforcing form–meaning connections and facilitating transfer to productive skills. Teachers play a key role in reinforcing this synergy by designing tasks that prompt learners to reuse vocabulary in writing and speaking, thereby consolidating learning across modalities.

In sum, AI-supported reading and vocabulary development exemplify the strengths of adaptive, data-informed pedagogy. When thoughtfully integrated, AI tools enhance comprehension, promote durable lexical acquisition, and support learner autonomy. Yet their effectiveness ultimately depends on pedagogical mediation that foregrounds strategic reading, contextualized vocabulary use, and the integration of receptive skills within broader communicative goals.

5.2.3 Reading, Vocabulary, and Learner Diversity

The adaptive reading and vocabulary research reviewed above concerns almost exclusively adolescent and adult learners in comparatively well-resourced settings. A smaller literature considers how the same mechanisms—immediate feedback, contextualized input, personalized pacing—operate for younger, disabled, and linguistically marginalized populations, complicating the optimistic picture above. A mixed-methods study of children's second-language learning found that AI tools such as speech recognition and interactive environments did enhance L2 skills through immediate feedback, but flagged ethical concerns and a risk of overdependence specific to child learners (Alkhutaba & Alnajjar, 2026)—a population with

even less-developed capacity than the adult learners discussed in section 5.2 to exercise the critical judgment this chapter treats as a precondition for AI's benefits.

Access to these affordances, where they exist, is also unevenly distributed. A systematic review conducted during the COVID-19 period found that AI tools supported written language, reading, listening, memory, and arithmetic for children with special educational needs, but that access barriers specific to developing-nation contexts constrained who could actually benefit (Bah & Artaria, 2020)—a reminder that for the learners this review describes, reading support does not arrive in isolation from support for other cognitive domains, and the barriers to accessing it are infrastructural before they are pedagogical.

Taken together, these studies expose an asymmetry in the evidence base itself: research on AI-supported reading and vocabulary for young, disabled, and low-resource-context learners is thinner, more review-based, and less experimentally controlled than the adult, higher-resource-context literature dominating this section. The gap matters because these are precisely the populations least equipped to compensate for a tool's limitations through independent critical judgment, and for whom the stakes of the affordances and limitations discussed throughout this chapter are highest. Section 5.6 returns to this equity dimension at the level of the chapter as a whole.

5.3 AI-Supported Listening and Multimodal Comprehension

5.3.1 Multimodal Input and Cognitive Processing

Listening instruction has historically been constrained by limited exposure to authentic input, restricted classroom time, and the transient nature of spoken language, which affords learners little opportunity to revisit or analyze input once it has passed. AI-enhanced learning environments address these constraints by enabling *multimodal input integration*, combining audio with video, transcripts, captions, visual annotations, and interactive comprehension support. This multimodal orchestration fundamentally reshapes how learners engage with spoken language, transforming listening from a linear, ephemeral activity into a flexible and revisitable process.

Meta-analytic findings indicate significant gains in listening comprehension when learners are granted control over *playback speed*, access to *adaptive captions*, and opportunities to receive *immediate feedback* on comprehension (Wu, 2024). Such affordances allow learners to regulate input according to their cognitive and linguistic readiness, thereby reducing processing pressure and facilitating more effective form–meaning mapping. From a cognitive standpoint, these features support segmentation of speech, lexical recognition, and syntactic parsing—processes that are often compromised in real-time listening.

This meta-analytic optimism should be read against more mixed evidence specific to listening as a discrete skill. Whereas AI-enhanced instruction has been shown to produce a large, significant effect on speaking proficiency (SMD = 1.033), the same synthesis of 19 studies found only a positive, non-significant trend for listening (SMD = 0.714) (Jantakoon et al., 2025)—one of the few instances in this chapter where a meta-analysis has failed to reach significance for its central outcome. Listening's comparative resistance to demonstrable AI-supported improvement may simply reflect that comprehension leaves no directly observable output for a system to evaluate and correct, unlike an utterance or a drafted paragraph.

Some of this asymmetry likely reflects how comparatively underexplored listening remains relative to the other three skills. Conceptual work mapping the affordances of generative AI—spoken dialogue systems, intelligent personal assistants, and related technologies—for foreign language listening and speaking concludes that the underlying capabilities are real but that both the pedagogical and technical literature on listening specifically lag behind the equivalent work on speaking (Goh & Aryadoust, 2025). Even reviews reporting significant listening gains tend to bundle them together with reading, writing, speaking, and vocabulary rather than isolating them as an independent finding (Kundu & Bej, 2025), making it difficult to know how much of any reported gain is listening specific.

Multimodal listening environments also enhance comprehension by supporting *dual-channel processing*, whereby auditory input is reinforced through visual and textual cues. When carefully aligned, multimodal input can scaffold comprehension by clarifying referential meaning, reinforcing key information, and supporting inferencing. For example, captions can assist learners in identifying word boundaries and unfamiliar vocabulary,

while visual context can aid pragmatic interpretation and situational understanding. Over time, such support contributes to improved listening fluency and reduced reliance on compensatory strategies.

However, the literature consistently cautions that multimodality is not inherently beneficial. Poorly designed multimodal environments—characterized by excessive visual stimuli, redundant information, or misaligned modalities—can increase *extraneous cognitive load,* diverting attention away from meaning construction (Edmett et al., 2024). Learners may become overwhelmed by competing inputs or rely excessively on captions at the expense of auditory processing. As a result, scholars emphasize the importance of *pedagogical alignment* between learner proficiency, task objectives, and modality selection. Effective multimodal listening instruction requires deliberate sequencing, such as gradually reducing visual support as listening competence develops, to ensure that scaffolding facilitates rather than replaces auditory comprehension.

AI-supported systems offer a further advantage by enabling *adaptive multimodal scaffolding.* Based on learner interaction and performance data, AI environments can dynamically adjust the level and type of support provided—such as toggling captions, highlighting key phrases, or prompting strategic pauses. This adaptivity supports personalized listening pathways and encourages learners to develop metacognitive awareness of how different modalities support comprehension. When learners are guided to reflect on their use of multimodal resources, listening instruction becomes not only more accessible but also more strategic.

5.3.2 Immersive and Contextualized Listening

Beyond multimodal enhancement, emerging research explores the potential of *immersive and contextualized listening environments,* including virtual reality (VR), augmented reality (AR), and simulated social settings. These environments situate listening tasks within meaningful, goal-oriented contexts, such as navigating public spaces, participating in professional interactions, or engaging in everyday social exchanges. By embedding listening within action-oriented scenarios, immersive technologies support a shift from decontextualized comprehension to *situated meaning-making*.

From a sociocognitive perspective, immersive listening environments enhance comprehension by providing rich contextual cues that support inferencing,

pragmatic interpretation, and situational awareness. Learners are not simply decoding linguistic input but interpreting meaning in relation to speaker intent, social roles, and environmental affordances. Preliminary empirical findings suggest that such contextualization improves learners' ability to comprehend implied meaning, manage ambiguity, and respond appropriately in real-world listening situations (Mohebbi, 2025).

Immersive listening environments also appear to exert a positive influence on *engagement and motivation*. The experiential nature of simulated contexts encourages active participation and sustained attention, addressing one of the longstanding challenges of listening instruction—learner passivity. When learners perceive listening tasks as purposeful and relevant, they are more likely to persist, repeat activities, and invest cognitive effort. Early studies report increased retention and transfer of listening skills when immersive environments are integrated with reflective follow-up tasks.

This description is beginning to receive larger-scale controlled support, albeit indirectly. A randomized controlled trial with 477 Chinese undergraduate EFL learners found that AI-enhanced immersive virtual reality—augmented through natural-language processing, machine learning, or sensor-based methods—produced measurable improvements in CEFR-aligned grammatical and lexical performance relative to a non-immersive control condition, along with within-VR comparative effects on communication and intercultural competence (Niu, 2026). Its outcome measures, however, target grammatical and lexical performance rather than listening comprehension specifically, illustrating a broader pattern in this section: much of the immersive-VR evidence base speaks to language performance in general rather than to listening as a discrete, separately measured construct.

Nevertheless, the empirical base for immersive listening remains emergent, and researchers urge caution against premature generalisation. Technical limitations, accessibility concerns, and variability in learner responses present ongoing challenges. Moreover, immersive environments may inadvertently prioritise novelty over pedagogical coherence if not carefully designed. As with other AI-supported modalities, the educational value of immersive listening depends on *instructional integration*, where immersive experiences are embedded within broader learning sequences that include pre-listening preparation, guided interaction, and post-task reflection.

The teacher's role remains central in mediating immersive listening experiences. Teachers guide learners in setting listening goals, interpreting contextual cues, and reflecting on comprehension strategies. When immersive listening is framed as *situated rehearsal* rather than as an end in itself, it complements rather than competes with traditional listening instruction. In this capacity, immersive AI technologies function as extensions of the listening ecology, offering learners access to rich, contextualized input that would otherwise be inaccessible within classroom constraints.

In sum, AI-supported listening and multimodal comprehension represent a significant expansion of instructional possibilities. Through adaptive multimodal scaffolding and immersive contextualization, AI technologies can enhance comprehension, engagement, and pragmatic awareness. However, these benefits are contingent on thoughtful pedagogical design that balances cognitive support with strategic challenge and situates listening within meaningful communicative contexts. When aligned with human-centered instructional goals, AI-enhanced listening instruction contributes not only to improved comprehension but also to learners' broader communicative competence.

5.4 AI-Supported Writing Instruction

5.4.1 Automated Feedback as Cognitive Scaffolding

Writing has emerged as one of the most extensively researched domains of Artificial Intelligence (AI) application in language education, largely due to its compatibility with data-driven analysis and iterative learning processes. Automated Writing Evaluation (AWE) systems and generative AI tools are now capable of providing immediate, fine-grained feedback on multiple dimensions of writing performance, as documented in a growing body of systematic review evidence (Shi & Aryadoust, 2024), including grammatical accuracy, syntactic complexity, lexical choice, cohesion, and discourse organization. This shift represents a significant departure from traditional feedback models that rely primarily on delayed, selective, and often summative teacher commentary.

Meta-analytic and systematic review evidence consistently demonstrates that AI-supported writing instruction leads to measurable improvements in linguistic accuracy, fluency, and textual complexity, particularly when learners

engage in *multiple cycles of drafting, feedback, and revision* (Wu, 2024). These gains are not merely mechanical; rather, they reflect increased opportunities for hypothesis testing, error noticing, and form–meaning alignment—processes long recognized as central to second language writing development. By enabling learners to receive immediate responses to their linguistic choices, AI systems effectively transform writing from a product-oriented activity into an ongoing process of experimentation and refinement.

The concrete evidence behind this claim now spans several instructional contexts and target languages. In an eight-week quasi-experiment with 53 intermediate-to-advanced undergraduate learners of Chinese as a second language, ChatGPT-4.0 feedback produced greater writing gains and higher self-efficacy than traditional teacher support (Huang & Chen, 2026). A quasi-experiment with 124 L2 German learners found that a domain-specific LLM feedback system, Dr. Write, produced greater gains in writing self-efficacy and clause-level accuracy than generic-LLM or teacher feedback, though without matching gains in affective engagement (Zheng et al., 2026). In India, 40 first-year ESL engineering students using an integrated combination of DeepL Write and Claude AI showed significant gains in argumentative-writing scores and metacognitive awareness of argument structure (Divya & Alamelu, 2026), while 60 Omani EFL university learners receiving Gemini-assisted feedback outperformed a traditional-feedback control group on both an immediate and a delayed posttest, with higher motivation (Ng et al., 2025). The underlying technology is not intrinsically English-specific: a neural automated writing evaluation system trained on a Korean learner corpus achieved high reliability in predicting Korean L2 writing proficiency scores (Lim et al., 2023), evidence that automated writing evaluation is becoming a cross-linguistic technology rather than an artifact of English-language NLP resources.

From a cognitive perspective, automated feedback functions as a form of *externalised scaffolding.* By offloading lower-level monitoring tasks—such as detecting grammatical inconsistencies or surface-level cohesion problems—AI reduces extraneous cognitive load during revision. This cognitive redistribution allows learners to allocate attentional resources to higher-order writing concerns, including argument development, rhetorical organization, and audience awareness (Mohebbi, 2025). In this sense, AI-supported feedback aligns closely with cognitive load theory and

sociocognitive models of writing, which emphasize the importance of freeing working memory capacity for meaning construction and strategic decision-making.

A further pedagogical advantage of AI-based feedback lies in its *temporal and adaptive nature*. Unlike teacher feedback, which is often constrained by institutional workload, class size, and assessment schedules, AI feedback is continuous, scalable, and responsive to learner input in real time. Learners are thus able to revise texts while the cognitive and communicative context of writing remains salient, increasing the likelihood that feedback will be processed meaningfully rather than superficially. Empirical studies suggest that this immediacy promotes deeper engagement with feedback and encourages learners to take greater responsibility for their revision decisions.

Despite these benefits, the literature also identifies critical limitations that complicate uncritical adoption of AI feedback systems. Empirical studies repeatedly note that AI tools tend to privilege *surface-level linguistic accuracy* over deeper rhetorical and pragmatic dimensions of writing. While AI systems can reliably identify grammatical errors or suggest lexical alternatives, they often struggle to evaluate communicative intent, authorial voice, genre-specific conventions, and culturally embedded pragmatic norms—particularly in argumentative, creative, or discipline-specific academic writing (Pham, 2025). As a result, there is a risk that learners may produce formally polished texts that lack rhetorical coherence, critical depth, or contextual appropriateness.

Above mentioned limitations are visible even in studies that report net gains. Forty Korean EFL pre-service teachers who used ChatGPT and EditGPT to revise their writing saw 86 percent of their errors corrected, but word-choice and structural errors persisted (Chon & Shin, 2026). Left to their own judgment, learners do not simply accept what remains: 45 L2 students revising argumentative essays with ChatGPT feedback rejected or ignored roughly 38 percent of it—22 percent by arguing against it, 16 percent by ignoring it—rejecting content-level feedback significantly more than form-level feedback as inconvenient or unhelpful (Ziqi et al., 2024). A study of 25 B1-level Turkish EFL students revising opinion essays found a similar asymmetry: most language-use feedback was accepted, but organization-level feedback met mixed engagement, leaving over a third

uncorrected (Cengiz et al., 2025). AI feedback, across these studies, is acted on most readily where it is most surface-level.

Benefits are also uneven across proficiency levels. Low-proficiency EFL learners interacting with ChatGPT feedback found the process cognitively demanding and the feedback itself of limited effectiveness for improving their writing (Li, 2023), suggesting the scaffolding claimed above presupposes a baseline competence not all learners bring. Comparative work on reformulation makes a related point: when seven EFL instructors and three AI chatbots—ChatGPT-3.5, Claude2, and Bing Chat—reformulated the same Japanese EFL students' writing, human reformulations most closely preserved the original text's length and vocabulary (Nakai & Shintani, 2025), raising the question of whose voice survives an AI-mediated revision. Most striking is a null result: across 270 EFL students' texts collected over a thirteen-week semester, neither teacher, automated, nor combined feedback produced significant gains in syntactic complexity, with effects varying by proficiency level rather than feedback source (Thi & Nikolov, 2023)—a direct complication of this subsection's opening claim of consistent gains.

These limitations have prompted a growing consensus in the literature in favor of *hybrid feedback models*, in which AI and human expertise are deliberately integrated rather than positioned in opposition. In such models, AI tools are employed to support form-focused revision, error detection, and micro-level language refinement, while teachers retain responsibility for guiding discourse-level development, genre awareness, and critical engagement with content (Edmett et al., 2024). This division of labour not only preserves the pedagogical authority of the teacher but also positions AI as a complementary cognitive tool rather than an evaluative authority.

It is worth noting that research emphasizes that the effectiveness of automated feedback depends less on the sophistication of the technology itself than on how learners are taught to *interpret, evaluate, and act upon AI-generated suggestions*. Without explicit instruction in feedback literacy and critical engagement, learners may adopt AI recommendations uncritically, leading to overreliance or diminished authorial agency. On this basis, contemporary scholarship increasingly frames AI-supported writing instruction as a *process-oriented, metacognitively mediated pedagogy*, in which

automated feedback serves as a catalyst for reflection, negotiation, and informed linguistic choice rather than a substitute for human judgment.

This literacy is increasingly the subject of dedicated measurement rather than general appeal. Using exploratory and confirmatory factor analysis with 385 and 420 Chinese EFL university students respectively, one study validated a 22-item, four-dimensional scale operationalizing AI literacy in L2 writing as understanding, application, ethics and risk awareness, and self-efficacy (Huang, Z., et al., 2025). Qualitative work converges on a similar picture: focus-group analysis of Spanish applied-linguistics students found their critical AI literacy blends technical understanding with ethical awareness and an explicit insistence that AI augment rather than replace human expertise (Pérez-Paredes et al., 2025), as is mentioned before this book adopts augmentation-over-automation principle (Chapter 13).

5.4.2 Revision Practices and Metacognitive Development

Beyond surface-level error correction, one of the most significant pedagogical contributions of AI-supported writing environments lies in their capacity to reshape learners' *revision practices* and promote deeper *metacognitive engagement* with the writing process. Revision, long recognized as a central component of effective writing development, is no longer confined to a single post-draft stage but has become an ongoing, recursive activity facilitated by continuous AI-generated feedback.

Traditional second language writing instruction has often struggled to cultivate meaningful revision behaviors. Learners frequently interpret teacher feedback as evaluative rather than developmental, leading to minimal or cosmetic revisions focused on correcting isolated errors. AI-supported writing environments, by contrast, encourage repeated cycles of drafting, feedback, and modification, thereby normalising revision as an integral and expected part of writing rather than as remediation for failure. Systematic reviews indicate that learners exposed to AI-mediated feedback environments engage in more frequent and sustained revision cycles than those receiving teacher feedback alone (Mohebbi, 2025).

From a metacognitive perspective, AI-supported revision practices encourage the development of *self-regulated learning (SRL)* skills, particularly in the domains of monitoring, evaluation, and strategic decision-making. When learners interact with AI feedback, they are prompted—implicitly or

explicitly—to assess the relevance of suggestions, compare alternative formulations, and decide whether and how to implement changes. This process requires learners to move beyond passive compliance and engage in *metacognitive judgment*, evaluating linguistic accuracy, communicative effectiveness, and alignment with writing goals.

Empirical studies demonstrate that learners who engage critically with AI feedback exhibit higher levels of metacognitive awareness and writing autonomy rather than accepting suggestions automatically. These learners are more likely to question feedback, seek clarification, and make selective revisions based on communicative intent, whereas learners who adopt AI feedback uncritically tend to prioritise formal correctness at the expense of meaning and rhetorical coherence (Mohebbi, 2025). This distinction makes clear the importance of *revision literacy*, defined as the ability to interpret feedback, evaluate its relevance, and apply it strategically within a broader writing plan.

This distinction has recently been operationalized rather than left at the level of description. Drawing on focus groups with 12 students and a subsequent survey of 542 university L2 writers, one study validated a 22-item, four-dimensional scale measuring cognitive, behavioral, emotional, and ethical engagement with ChatGPT feedback (Dong, 2025)—an instrument for measuring exactly the distinction between critical and uncritical engagement this chapter has so far only described qualitatively. The risk of the uncritical pole is not merely theoretical: twelve students using ERNIE Bot generative-AI feedback reported real technological, educational, and social benefits, but the same study identified AI-literacy gaps and a risk that sustained reliance could weaken independent writing ability over time (Mi et al., 2026).

Importantly, recent research highlights that AI-generated feedback can alter learners' *epistemic stance toward text ownership and authorship*. Studies investigating revision behavior in AI-supported writing contexts reveal a tendency for learners to invest less cognitive effort in revising texts perceived as AI-generated, a phenomenon attributed to overtrust in machine output or reduced sense of personal responsibility (Radtke & Rummel, 2025). This finding has significant implications for pedagogy, as it suggests that AI may inadvertently encourage superficial engagement unless learners

are explicitly trained to treat AI output as provisional and negotiable rather than authoritative.

To address this risk, contemporary scholarship emphasizes the need to embed *explicit metacognitive instruction* within AI-supported writing pedagogy. Learners must be guided to ask critical questions such as: *Why is this suggestion being made? What linguistic principle does it reflect? Does it align with my intended meaning, audience, and genre?* When revision is framed as an act of decision-making rather than correction, AI feedback becomes a catalyst for reflective learning rather than a shortcut to textual polish.

Furthermore, AI-supported revision environments enable learners to externalise and observe their own writing processes over time. Through interaction logs, version histories, and iterative drafts, learners can trace patterns in their errors, stylistic choices, and revision strategies. This longitudinal visibility supports *metacognitive calibration*, allowing learners to identify persistent weaknesses, recognize improvement, and adjust learning strategies accordingly. Such process awareness has been shown to strengthen learner confidence and motivation, particularly in academic writing contexts where affective barriers such as anxiety and self-doubt are prevalent.

Nevertheless, the literature cautions that metacognitive benefits are not inherent to AI use but are *pedagogically mediated*. Without instructional guidance, learners may focus narrowly on error elimination, treating revision as a mechanical task rather than as an opportunity for conceptual refinement. As a result, researchers increasingly argue that AI-supported writing instruction must be situated within *process-oriented pedagogical frameworks,* where revision is explicitly taught, modeled, and discussed as a meaning-making activity (Faiz & Hazilan, 2024).

A micro-genetic pilot study of Hong Kong primary school pupils illustrates what such deliberate design can look like at the younger end of the age range this chapter otherwise underrepresents. Using a generative AI agent built around what its designers call the E.F.F.E.C.T. framework, the study found that explicit scaffolding supported self-regulated learning and reduced writing anxiety among primary-age L2 writers (Wong, 2026)—evidence that the metacognitive scaffolding logic developed above for university writers can extend to much younger learners when made explicit rather than assumed.

In this respect, the role of the teacher remains central. Teachers function not as feedback providers in the traditional sense but as *metacognitive coaches*, helping learners interpret AI feedback, reflect on revision choices, and connect linguistic adjustments to communicative goals. When AI feedback is integrated into guided reflection activities—such as revision commentaries, peer discussion, or reflective journals—its potential to foster durable writing competence and learner autonomy is significantly enhanced.

In summary, AI-supported revision practices represent a critical intersection between technological affordance and learner cognition. When thoughtfully integrated, AI tools can transform revision from a corrective afterthought into a sustained, reflective, and self-regulated learning process. However, this transformation depends on deliberate pedagogical design that foregrounds metacognitive development, preserves authorial agency, and positions AI as a partner in learning rather than an unquestioned authority.

5.4.3 Genre Awareness and Academic Literacy

Beyond supporting linguistic accuracy and revision practices, AI-supported writing environments increasingly play a significant role in the development of *genre awareness and academic literacy*, particularly in higher education and research-oriented contexts. Genre knowledge—the ability to recognize, interpret, and produce texts that conform to socially situated conventions—is a cornerstone of academic writing competence. For many second and foreign language learners, difficulties in academic writing stem less from grammatical limitations than from limited access to implicit disciplinary norms governing argumentation, structure, stance, and epistemic positioning.

AI-driven writing tools offer new affordances for making genre conventions *visible, analysable, and negotiable*. Generative AI systems can model prototypical academic genres—such as research articles, literature reviews, reflective essays, or argumentative papers—by providing structural outlines, rhetorical move sequences, and example formulations. Empirical studies indicate that when learners use AI tools during the planning and drafting stages, they demonstrate improved organizational coherence, clearer rhetorical progression, and increased awareness of genre-specific expectations (Tran et al., 2025). In this sense, AI functions as a *genre*

mediator, bridging the gap between tacit disciplinary knowledge and learner production.

A discipline-specific study bears this out concretely. Saudi university EFL/ESP writing students who used ChatGPT showed significant gains in grammar accuracy, vocabulary, creativity, and critical thinking, notwithstanding technical and ethical implementation challenges (Elmotri et al., 2025)—gains reported not only in the surface-accuracy dimensions AI is best known for but in creativity and critical thinking as well, where genre modeling was embedded within an ESP curriculum rather than left to incidental tool use.

From an academic literacy perspective, AI-supported genre scaffolding aligns with sociocultural models of writing, which conceptualize literacy as participation in disciplinary discourse communities rather than the acquisition of decontextualized language skills. By exposing learners to recurring rhetorical patterns—such as hedging strategies, stance markers, and citation practices—AI tools can support learners' gradual enculturation into academic discourse. This is particularly valuable for novice writers and multilingual students, who often lack access to sustained mentoring or exemplars within their academic contexts.

However, the literature also highlights a critical tension: while AI can model genre conventions effectively, it may also *reify formulaic writing practices* if used uncritically. Studies caution that learners who rely heavily on AI-generated templates or phrasing may reproduce surface-level genre features without developing a deeper understanding of epistemic purpose or disciplinary variation (Kartal & Yeşilyurt, 2024). This risk is especially pronounced in genres that require evaluative judgment, theoretical positioning, or original synthesis, such as research proposals or argumentative essays.

Seen in this light, scholars argue that AI-supported genre instruction must be framed not as genre imitation but as *genre inquiry*. When learners are encouraged to compare AI-generated texts with authentic disciplinary samples, identify divergences, and reflect on rhetorical choices, AI becomes a resource for critical genre analysis rather than prescriptive conformity. In such pedagogical designs, learners interrogate not only *how* academic texts are structured but *why* particular rhetorical strategies are valued within specific discourse communities.

Another crucial dimension of academic literacy concerns citation practices, source integration, and ethical authorship. Generative AI tools can assist learners in paraphrasing, summarising, and synthesising sources, thereby supporting engagement with academic texts. At the same time, the propensity of AI systems to generate fabricated references or overly polished paraphrases has intensified concerns about plagiarism, authorship ambiguity, and epistemic integrity (Zhang, Hussin, & Abd Majid (2025)). As a result, recent research emphasizes the integration of ethical AI literacy into academic writing instruction, positioning genre competence as inseparable from responsible knowledge production.

A qualitative account from a Thai university's English-medium-instruction writing class gives this concern a concrete shape. In a poetic autoethnography combining the perspectives of one EFL writing teacher and two international students, ChatGPT was described as functioning as a pervasive "ghostwriter" in academic writing, present throughout the writing process in ways requiring ongoing ethical navigation rather than a one-time disclosure decision (Yang et al., 2024). A ghostwriter, unlike a plagiarized source, leaves no discrete trace to detect—exactly what makes the authorship questions above resistant to policy solutions built on detection alone.

In this context, genre awareness extends beyond textual form to include an understanding of academic values such as originality, transparency, and accountability. Learners must be guided to distinguish between AI as a *supportive drafting tool* and AI as an illegitimate source of content generation. Empirical studies show that explicit instruction in AI-assisted citation, acknowledgment of AI use, and reflective commentary on writing processes considerably reduces misuse and enhances learners' sense of authorial responsibility (Edmett et al., 2024).

Importantly, AI-supported genre learning also intersects with issues of *equity and epistemic access.* Academic genres often privilege implicit norms associated with dominant linguistic and cultural traditions. By making these norms more explicit, AI tools have the potential to democratise access to academic discourse, particularly for first-generation students and multilingual writers. However, this potential is contingent on careful pedagogical mediation; without critical framing, AI may simply reproduce existing power structures embedded in academic genres rather than challenge them.

The contingency is not hypothetical. An evaluation of generative AI applied to Bengali grammatical-error correction found that tools could competently flag errors in this lower-resourced language but largely failed to produce the learner-usable, natural-language explanations this chapter has repeatedly identified as central to feedback's pedagogical value (Maity et al., 2024). Where explanation fails, feedback reduces to correction, and the metacognitive engagement this chapter treats as the mechanism connecting AI feedback to learning has little to work with—the democratizing potential of AI-supported writing instruction remains bounded by the same linguistic-resource asymmetries that shape natural language processing more broadly.

The role of the teacher, therefore, remains indispensable. Teachers act as *genre brokers*, helping learners interpret AI-generated models, contextualise genre conventions within disciplinary practices, and negotiate tensions between conformity and creativity. When AI-supported writing is embedded within genre-based pedagogy—such as scaffolded tasks, reflective annotations, and comparative genre analysis—it supports not only textual proficiency but also learners' participation in academic knowledge-making.

To sum up, AI-supported writing instruction holds considerable promise for enhancing genre awareness and academic literacy, provided it is implemented within a critical, process-oriented framework. When used judiciously, AI tools can illuminate the often invisible conventions of academic discourse, support learners' rhetorical development, and cultivate ethical engagement with scholarly practices. However, these benefits depend on sustained pedagogical guidance that foregrounds agency, critical reflection, and disciplinary situatedness over mere textual conformity.

Collectively, these findings highlight a central principle: *AI-supported writing instruction is pedagogically effective only when it is human-centerd, process-oriented, and critically mediated.* Rather than replacing teacher expertise, AI reshapes instructional roles, positioning educators as designers of writing ecologies that integrate automated feedback, reflective revision, and genre awareness within coherent learning trajectories. In this integrated model, AI functions not as an evaluative authority but as a cognitive and rhetorical partner—one whose educational value depends fundamentally on informed pedagogical orchestration.

5.4.4 Comparing AI, Teacher, and Peer Feedback

Because AI feedback is so often discussed in relation to what teachers do, it is worth asking what happens when the two are compared directly. The answer depends on timescale and outcome. In an eight-week study of 61 Chinese university students' writing across 244 texts, teacher feedback produced stronger short-term gains, while AI-assisted peer feedback produced a steadier, more sustained learning-curve effect (Tang et al., 2026), this finding cautions against ranking AI and teacher feedback on a single scale when their advantages surface on different timescales. Elsewhere the comparison favors AI more directly: EFL college learners receiving ChatGPT-assisted feedback achieved higher holistic scores and fewer grammatical and lexical errors than learners receiving peer feedback, though some feared over-reliance (Mun, 2024).

Perceived effectiveness and stated preference are not the same thing. In a pilot study of 15 EFL student pairs, both ChatGPT and teacher corrective feedback were judged effective, yet students still preferred teacher feedback, or a combined teacher–ChatGPT approach, over ChatGPT alone (Zeevy-Solovey, 2024). Reliability comparisons complicate matters further: when three human raters and three AI tools—ChatGPT, CiCi, and Bing—scored the same ESL essays, inter-rater consistency between human and AI raters was low, even though overall score levels did not differ significantly (Guibangguibang & Hesula, 2026)—a disagreement in where scores come from, not in their average value, with direct bearing on the assessment-theoretic apparatus of Chapters 10 and 11.

Teachers themselves are not of one mind about any of this. A Q-methodological study in which 40 teachers sorted 42 statements about AI-generated feedback identified four distinct viewpoints, ranging from instructor-led trust, through strategic resistance, to transformative embrace (Jinowat et al., 2026)—"*the teacher's view*" of AI feedback is better treated as contested terrain than a single position. That contest has a practical edge: one formulation asks directly whether automated writing evaluation and generative AI feedback can relieve EFL teachers of feedback fatigue while genuinely improving student writers, rather than merely their texts (Hyland, 2025), anticipating Chapter 6's fuller treatment of teacher workload under AI-mediated conditions.

5.5 AI-Supported Speaking and Pronunciation

5.5.1 Automatic Speech Recognition and Phonological Development

Speaking and pronunciation instruction has benefited substantially from recent advances in automatic speech recognition (ASR) technologies, which now enable fine-grained analysis of learner speech in real time. Contemporary ASR-based systems are capable of processing both segmental features (such as vowel and consonant production) and suprasegmental features (including stress patterns, rhythm, intonation, and speech rate), offering learners immediate, data-driven feedback on oral performance. This represents a significant pedagogical shift from traditional pronunciation instruction, which has often relied on limited classroom time, delayed teacher feedback, and subjective auditory judgments.

Meta-analytic findings consistently demonstrate that speaking-related skills yield some of the largest effect sizes among AI-supported language learning outcomes, particularly in pronunciation accuracy and fluency development (Wu, 2024). These gains are frequently attributed to the frequency and consistency of practice enabled by ASR systems. Unlike classroom-based speaking activities—where individual learners may receive only minimal speaking time—ASR-based tools allow learners to engage in repeated, focused rehearsal without institutional or temporal constraints. This practice intensity supports the automatisation of phonological patterns, a key condition for fluent speech production.

This ranking is not settled across the meta-analytic literature, and the disagreement is worth making explicit. A more recent multilevel meta-analysis of 117 effect sizes from 46 EFL studies published between 2022 and 2025 found an overall medium-to-large effect of AI on language learning—but located this effect very differently across skills, with vocabulary showing the strongest gains and speaking the weakest (Torres & Kahveci, 2025). Part of the discrepancy is likely definitional: primary studies coded as measuring "speaking" vary widely in whether they operationalize it as pronunciation accuracy, fluency, or self-rated confidence, and such construct heterogeneity is a familiar threat to comparability across second-language meta-analyses.

A third synthesis narrows the comparison usefully. A meta-analysis of 19 studies published between 2017 and 2024 found that AI-enhanced English instruction produced a significant, large effect on *speaking proficiency* (SMD = 1.033), but only a positive, non-significant trend for listening (SMD = 0.714) (Jantakoon et al., 2025)—a pattern suggesting, alongside Torres and Kahveci's contrasting ranking, that the size and even the ordering of skill-specific AI effects depends heavily on which outcome measures and study pools a synthesis draws on. Conversational practice specifically tells a more moderate story: a meta-analysis of 70 effect sizes from 28 studies of chatbot-assisted conversation practice found an overall positive but moderated effect on language-learning performance rather than a uniformly large one (Wang et al., 2024). The safest generalization is a modest one: AI reliably supports speaking development, but a single dominant effect size for any one skill should be read with the instability of the underlying evidence base in mind.

From a cognitive and psycholinguistic perspective, ASR-based feedback facilitates *phonological noticing,* a mechanism central to second language speech development. By explicitly highlighting mismatches between learner production and target-like forms, ASR systems help learners become aware of phonetic contrasts that may otherwise remain perceptually inaccessible. Over time, repeated exposure to such feedback supports the restructuring of phonological representations and improves articulatory control. Importantly, ASR systems provide consistent feedback that is unaffected by teacher fatigue or classroom variability, thereby offering a stable reference point for learner self-monitoring.

A consistently reported advantage of ASR-based practice is its impact on the *affective dimension* of speaking. Oral production is widely recognized as one of the most anxiety-provoking aspects of language learning, particularly in teacher-fronted or peer-evaluative contexts. Studies consistently report that AI-mediated speaking environments reduce performance anxiety and increase learners' willingness to communicate, as learners can practise privately, repeat attempts without embarrassment, and experiment with language forms without fear of negative social judgment (Xu & Ismail, 2025). This reduction in affective barriers is particularly salient for lower-proficiency learners and those from educational cultures where oral error correction is perceived as face-threatening.

Despite these advantages, the literature also identifies notable limitations in ASR-based pronunciation instruction. Current systems may struggle with accent diversity, non-standard pronunciation patterns, and speech produced in noisy or low-quality recording conditions. In addition, while ASR feedback excels at identifying deviations from target norms, it may oversimplify the complex relationship between intelligibility, comprehensibility, and accentedness. As a result, researchers caution against equating ASR-generated scores with communicative effectiveness and argue for pedagogical models that prioritise intelligible speech over native-like pronunciation.

Set against these limitations, the net effect in large controlled comparisons remains positive. In a 12-week study, 120 ESL undergraduates split into an AI-chatbot group and an equal-sized control group, with the chatbot group significantly outperforming controls in fluency, pronunciation, grammar, vocabulary, and self-rated speaking confidence (Grab, 2025). A qualitative case study of ChatGPT's responses across EFL writing, grammar, vocabulary, and conversation tasks reaches a compatible conclusion from a different angle, finding the tool linguistically accurate and useful overall while its limitations cluster around pragmatics and cultural sensitivity rather than core linguistic accuracy (Islomova, 2026)—the same accuracy-versus-appropriateness distinction this section has already drawn for pronunciation.

As a result, scholars increasingly advocate hybrid pronunciation pedagogies, in which ASR tools are used to support repetitive practice, self-monitoring, and phonological awareness, while teachers guide learners in setting realistic pronunciation goals and interpreting feedback within communicative contexts. In such models, ASR functions as a powerful diagnostic and rehearsal tool, but pedagogical judgment remains firmly human-led.

5.5.2 Conversational Agents and Task-Based Interaction

Beyond pronunciation-focused practice, AI-supported speaking instruction has expanded through the use of conversational agents, including chatbots and voice-based virtual interlocutors. These systems enable learners to engage in simulated spoken interaction through role-plays, interviews, problem-solving tasks, and scenario-based dialogues. Unlike scripted drills, conversational AI systems are designed to generate *dynamic responses*, thereby approximating elements of real-time interaction and supporting extended oral production.

Systematic reviews focusing on Asian tertiary education contexts identify conversational agents as particularly effective when integrated into task-based language teaching (TBLT) frameworks (Wang et al., 2025). Within such frameworks, AI agents serve as interlocutors that prompt learners to *negotiate meaning, clarify misunderstandings, and reformulate output* across multiple interactional turns. Empirical studies indicate that repeated engagement with AI-mediated tasks leads to improvements in *fluency, lexical diversity, and discourse coherence,* especially when tasks are designed to increase cognitive and communicative complexity progressively.

From an interactionist perspective, conversational AI supports key conditions for second language development by creating opportunities for modified output and interactional feedback. Learners are encouraged to sustain interaction, respond to prompts, and adjust their language use in response to system-generated cues. Over time, these interactional cycles contribute to proceduralisation of speaking skills and increased confidence in spontaneous speech production.

The empirical base for these claims now extends well beyond the Asian tertiary contexts most heavily represented in this literature. A longitudinal study of Swedish lower-secondary students practising English through a spoken dialogue system built around embodied conversational agents found sustained engagement over time and a slightly positive trend in students' social and emotional learning experience (Ericsson & Johansson, 2023). Mixed-reality implementations extend the same logic to group interaction: a mixed-reality multi-agent system supporting group conversation, tested with 21 second-language learners, increased engagement and willingness to communicate while easing speaking-related fear (Fernandez-Espinosa et al., 2025). The paradigm has also moved beyond English-medium contexts: a multi-agent chatbot supporting English–Arabic translation and language learning, evaluated by 40 undergraduates and four faculty members, significantly enhanced learner engagement and pedagogical value (Aleedy et al., 2025).

Task and dialogue design, moreover, are not interchangeable. A triple-blind randomized controlled trial with 456 EFL learners at CEFR B1–B2 compared three AI dialogue modes—guided feedback, interactive question-and-answer, and inquiry-driven scaffolding—and found differential gains in computational thinking and language proficiency depending on which

mode learners were assigned to (Rui, 2025), substantiating with experimental evidence the pedagogical claim above: that the value of conversational AI lies in how the interaction is designed, not in the technology as such.

Nevertheless, the literature consistently cautions against overstating the communicative competence of AI interlocutors. Despite advances in natural language processing, AI agents lack full pragmatic and sociocultural awareness, particularly in contexts requiring sensitivity to politeness norms, cultural references, or emotionally nuanced interaction. As Huang et al. (2025) note, conversational AI systems often operate on probabilistic language models that prioritise coherence over contextual appropriateness, which may result in pragmatically odd or culturally neutralised responses.

For this reason, researchers emphasize that conversational AI should be conceptualized primarily as a rehearsal space rather than a substitute for authentic human communication. AI-mediated interaction is most pedagogically effective when used to prepare learners for real-world speaking tasks, providing a low-stakes environment for experimentation and practice. When learners transition from AI-based rehearsal to human interaction—such as classroom discussion or assessment—the benefits of increased fluency and reduced anxiety are more likely to transfer.

Pedagogically, the effectiveness of conversational agents depends heavily on *task design* and *instructional framing*. Tasks must be goal-oriented, meaningful, and aligned with communicative objectives rather than treated as novelty-driven exchanges. Teachers play a crucial role in contextualising AI-mediated interaction, debriefing task performance, and drawing learners' attention to pragmatic and discourse-level features that AI systems may not fully model.

In sum, conversational AI systems offer substantial potential to expand speaking opportunities and support task-based interaction, particularly in contexts where access to authentic interlocutors is limited. However, their educational value lies not in replacing human interaction but in augmenting the speaking ecology, enabling learners to rehearse, reflect, and refine oral performance within a human-centerd pedagogical framework.

5.5.3 Speaking Anxiety, Affect, and Learner Identity

Speaking anxiety has long been identified as one of the most persistent affective barriers in second and foreign language learning. Oral production requires learners to perform linguistic knowledge publicly and spontaneously, often under conditions of time pressure and evaluative scrutiny. For many learners, particularly those in exam-oriented or teacher-centerd educational contexts, speaking is associated with fear of negative evaluation, loss of face, and perceived inadequacy of linguistic competence. These affective constraints frequently limit participation, reduce willingness to communicate, and inhibit opportunities for meaningful oral practice.

AI-supported speaking environments introduce a qualitatively different *affective ecology* by decoupling oral practice from immediate social judgment. Empirical studies consistently report that learners experience lower levels of anxiety and increased confidence when practicing speaking with AI systems, as these environments allow for repeated attempts, private rehearsal, and error-tolerant interaction (Xu & Ismail, 2025). Unlike classroom interaction, where mistakes are publicly visible and socially consequential, AI-mediated speaking offers a *low-stakes rehearsal space* in which errors are framed as data for learning rather than as failures.

Quasi-experimental and correlational evidence from Chinese EFL contexts substantiates this pattern with named populations. In a six-week quasi-experiment with 204 Chinese EFL students, AI-based instruction increased positive achievement emotions and willingness to communicate more than conventional instruction (Chen et al., 2025). A structural equation model fitted to 517 Chinese university EFL students found that AI literacy raises willingness to communicate indirectly, by increasing AI-learning self-efficacy and reducing foreign-language classroom anxiety (Zhang et al., 2025)—positioning self-efficacy and anxiety not as side effects of AI-mediated speaking practice but as the mechanisms through which it operates.

From an affective–cognitive perspective, reduced anxiety facilitates more efficient language processing and greater attentional availability during speech production. When learners are not preoccupied with self-monitoring or fear of evaluation, they are more likely to focus on *message formulation, fluency, and communicative intent*. Over time, repeated exposure to successful speaking experiences in AI-mediated contexts can lead to affective recalibration, whereby learners gradually reframe speaking as manageable

and controllable rather than threatening. This shift has been shown to increase willingness to communicate in subsequent human interaction settings, suggesting that affective benefits can transfer beyond the AI environment.

Beyond anxiety reduction, AI-supported speaking practices also shape *learner identity*—how learners perceive themselves as language users. Identity-oriented research in applied linguistics emphasizes that learners' participation in speaking is influenced not only by linguistic competence but also by beliefs about legitimacy, ownership, and voice. AI-mediated speaking environments provide learners with opportunities to experiment with new linguistic identities without the immediate risk of social sanction. In doing so, AI tools can support the development of a more confident and agentive speaker identity, particularly for learners who have previously positioned themselves as "silent" or "weak" speakers.

However, the literature also identifies important tensions in the relationship between AI-mediated speaking and learner identity. While AI interaction may promote confidence, there is a risk that learners may feel more competent in machine-mediated contexts than in human interaction, leading to a form of context-bound confidence. Some studies report that learners who perform well with AI interlocutors experience renewed anxiety when transitioning back to classroom discussions or assessments, where social dynamics, turn-taking, and pragmatic expectations are more complex. This finding underlines the need for pedagogical designs that explicitly bridge AI rehearsal and human interaction rather than treating them as separate domains.

Equally, researchers caution that AI systems are not neutral interlocutors. Conversational agents are designed based on dominant linguistic norms and may implicitly *privilege standardized varieties of language*, potentially marginalising learners whose accents or speech patterns deviate from these norms. Such *biases* can influence learners' perceptions of linguistic legitimacy and reinforce deficit-oriented views of non-native speech. As a result, scholars argue for a *critical affective pedagogy* that frames AI feedback within intelligibility-based rather than native-speaker-based standards and explicitly validates linguistic diversity.

The role of the teacher is therefore crucial in mediating the affective and identity-related dimensions of AI-supported speaking. Teachers function as

affective guides, helping learners interpret AI feedback, reflect on emotional responses to speaking tasks, and contextualise AI performance within broader communicative goals. When AI-mediated speaking is integrated with reflective activities—such as self-assessment, peer discussion, or guided debriefing—learners are better positioned to transfer confidence, manage anxiety, and negotiate speaker identity across contexts.

As a result, AI-supported speaking environments offer significant potential to reduce speaking anxiety, enhance affective engagement, and support the emergence of more confident learner identities. However, these benefits are neither automatic nor uniform. They depend on pedagogical frameworks that recognize affect and identity as central components of oral proficiency, position AI as a supportive rehearsal partner rather than an evaluative authority, and maintain a clear trajectory from private practice to authentic social communication. When these conditions are met, AI-mediated speaking instruction can contribute not only to improved fluency and pronunciation but also to learners' long-term participation and agency as language users.

Case Box 5.1—Immersive language learning before the LLM era: 3D virtual worlds at Istanbul University

The pedagogical logic of today's immersive, AI-populated practice environments was worked out in language education well before large language models. In a program of research at Istanbul University's English Language Teaching department, Can and Şimşek (2015) designed and implemented a ten-week course component for pre-service foreign language teachers on the 3D virtual world platform Second Life, running a "Comparative Education" class in parallel face-to-face and in-world formats with mixed-methods triangulation. Two findings from this line of work anticipate themes now central to AI-mediated language learning. First, avatar-based participation individualized the learning experience and lowered affective barriers: attending as an avatar reduced learners' stress and anxiety and raised motivation and willingness to take communicative risks—the same affective mechanism now invoked for low-anxiety conversational practice with AI interlocutors (Chapter 7). Second, learner ownership of the environment mattered: participants could build, own, and modify the virtual space itself, making agency a design feature rather than an aspiration. A companion strand explored immersive game-based vocabulary teaching in

the same platform (Can, Elia, & Şimşek, 2017). Read from the present, this work suggests continuity rather than rupture: what generative and agentic AI add to immersive language learning is not the pedagogy—presence, embodiment, lowered anxiety, learner ownership—but the linguistic intelligence of the interlocutors that populate such spaces.

5.6 Pedagogical and Ethical Implications

Across reading, listening, writing, and speaking the literature converges on a set of pedagogical and ethical principles that fundamentally reshape how Artificial Intelligence (AI) should be conceptualized and implemented in language education. Foremost among these is the recognition that human–AI collaboration consistently yields stronger learning outcomes than fully automated instructional models. Large-scale bibliometric mappings and systematic reviews demonstrate that AI-mediated instruction is most effective when embedded within human-led pedagogical frameworks that provide contextual interpretation, affective support, and ethical oversight (Yang et al., 2025; Zhang, Hussin, & Abd Majid (2025)). Empirical studies further indicate that *teacher-guided AI* use enhances *learner engagement, autonomy, and transfer of skills*, whereas fully automated approaches risk reducing learning to optimisation and compliance, neglecting the social, interpretive, and developmental dimensions of language use (Mohebbi, 2025; Wang et al., 2025).

A second convergent principle concerns the *primacy of process-oriented integration over product-focused use*. Across skills, AI-supported learning is most impactful when learners engage in iterative cycles of practice, feedback, reflection, and revision. Evidence from writing, speaking, and vocabulary research consistently shows that AI tools contribute most meaningfully when they support *how* learning unfolds rather than merely evaluating *what* learners produce (Ngo, 2024; Xu & Ismail, 2025). Product-oriented uses—such as automated scoring, surface-level correction, or decontextualized performance metrics—may offer efficiency, but they rarely support durable competence, metacognitive development, or cross-skill transfer (Mohebbi, 2025; Mateos-Blanco et al., 2024).

Crucially, the expansion of AI-mediated instruction foregrounds the necessity of ethical and critical AI literacy as a core pedagogical objective rather

than an ancillary concern. Multiple reviews emphasize that learners must be equipped to understand the *affordances and limitations* of AI systems, interrogate *algorithmic bias*, and make informed decisions about when and how AI support is *pedagogically legitimate* (Mateos-Blanco et al., 2024; Zhang, Hussin, & Abd Majid (2025)). Without such literacy, learners may over-rely on automated feedback, accept AI output uncritically, or misunderstand the epistemic boundaries between assistance and authorship—particularly in academic writing and assessment contexts (Mohebbi, 2025). Ethical AI literacy thus becomes inseparable from language literacy itself, especially in higher education and professional domains where accountability, originality, and transparency are paramount.

These shifts have profound implications for the role of the teacher. Rather than positioning educators as dispensable intermediaries, AI reshapes instructional roles by *recentring teachers as designers of learning ecologies*, facilitators of reflection, and guardians of communicative authenticity. Bibliometric and conceptual analyzes consistently frame teachers as essential human agents who curate tasks, sequence AI affordances, mediate feedback interpretation, and ensure alignment between linguistic development and communicative purpose (Yang et al., 2025; Edmett et al., 2024). In this model, pedagogical authority does not diminish; instead, it becomes more strategic, interpretive, and ethically grounded.

Equity considerations further complicate the pedagogical landscape. While AI has the potential to democratise access to practice, feedback, and authentic input, uneven technological access, language-model bias, and cultural normativity risk reproducing existing educational inequalities (Zhang, Hussin, & Abd Majid (2025); Mohebbi, 2025). Research in diverse educational contexts highlights that AI systems often *privilege dominant language varieties* and learning assumptions, underlining the *need for inclusive design and pedagogical mediation* (Mateos-Blanco et al., 2024). Ethical AI integration therefore requires sustained attention to accessibility, linguistic diversity, and contextual sensitivity to ensure that AI-supported instruction does not marginalise already disadvantaged learners.

Attitudinal evidence from early-childhood contexts illustrates how this equity concern intersects with age-appropriateness. A survey of 543 early-childhood teachers in Mecca, Saudi Arabia found broadly positive attitudes toward using ChatGPT to teach English as a second language, but also

concrete concern about the risks such tools pose for young children with still-limited digital skills (Allehyani & Algamdi, 2023). This reminds us that the equity question here concerns not only who has access to AI tools but whether the youngest, least digitally literate learners can safely exercise the critical judgment this chapter treats as a precondition for AI's benefits.

Finally, the literature underscores the need for future research agendas that move beyond short-term performance gains. Bibliometric mappings reveal a predominance of short-duration interventions and cross-sectional designs, with limited evidence on long-term transfer, skill integration, and real-world communicative impact (Yang et al., 2025; Zhang, Hussin, & Abd Majid (2025)). Longitudinal and mixed-methods studies are therefore needed to examine whether AI-supported skill development persists over time, transfers across modalities, and supports learners' participation in authentic communicative practices. At the same time, integrative research designs are required to explore how AI-mediated instruction across writing, speaking, reading, and listening can be coordinated within coherent curricula rather than implemented as isolated interventions.

This call for methodological rigor is reinforced by reviews that interrogate not just what chatbot-mediated language learning finds but how solidly it is shown. A systematic review of recent chatbot research explicitly evaluates the strength of evidence behind claimed learning-outcome gains, rather than simply tallying positive results, cautioning that enthusiasm for chatbots' usefulness in language learning has in places outpaced the rigor of the designs used to demonstrate it (Şahin Kızıl et al., 2025).

Thus, the pedagogical and ethical implications of AI in language education demand a *human-centerd, critically mediated, and ecologically informed approach*. Consistent with current empirical and conceptual scholarship, AI's educational value lies not in automation but in *augmentation*—enhancing instructional capacity while preserving the relational, interpretive, and ethical foundations of language teaching (Mohebbi, 2025; Yang et al., 2025). When thoughtfully integrated, AI becomes not a substitute for pedagogy, but a catalyst for reimagining language education as adaptive, reflective, and socially responsible.

Pedagogical Companion for Chapter 5

Key Concepts: automated writing evaluation; cognitive scaffolding; revision and metacognition; genre awareness; automatic speech recognition; conversational task-based interaction; speaking anxiety; adaptive reading; vocabulary retention; multimodal listening; immersion.

Discussion Questions

1. The chapter presents automated feedback as *cognitive scaffolding*. Under what classroom conditions does the same feedback become cognitive *outsourcing*?
2. Speaking practice with AI reduces anxiety partly because no human is listening. What is gained and what is lost when the audience is removed from communication?
3. How should a teacher sequence AI-supported reading tools so that adaptive simplification does not become a ceiling on learners' text complexity?
4. Case Box 5.1 traces today's immersive practice environments to pre-LLM virtual worlds. Which affordances attributed to AI in this chapter are actually affordances of *immersion*, independent of AI?
5. Across the four skills, where is the evidence for AI support strongest and weakest? What explains the asymmetry?

Further Reading

- Shi, H., & Aryadoust, V. (2024). A systematic review of AI-based automated written feedback research. *ReCALL, 36*(2).
- Voskoboinik, E., et al. (2025). Enhancing second language speaking assessment: Integrating large language models. *Language Testing, 42*(4).
- Can, T., & Şimşek, İ. (2015). The use of 3D virtual learning environments in training foreign language pre-service teachers. *TOJDE, 16*(4).
- Torres, P. J., & Kahveci, Y. E. (2025). Effectiveness of artificial intelligence (AI) in language teaching. *Computers and Education: Artificial Intelligence*, 9, 100522.
- Kundu, A., & Bej, T. (2025). Transforming EFL teaching with AI: A systematic review of empirical studies. *International Journal of Artificial Intelligence in Education*, 35(4), 2281–2314.

Chapter 6. Teacher Roles, Beliefs, and AI Pedagogical Literacy

6.1 Introduction

The integration of Artificial Intelligence (AI) into educational settings has shifted from a theoretical possibility to a tangible reality, fundamentally altering the landscape of language teaching, learning, and assessment. While technology has long played a supplementary role through *Computer-Assisted Language Learning (CALL)* and *Mobile-Assisted Language Learning (MALL)*, the emergence of *Generative AI (GAI)* and *Large Language Models (LLMs)* like *ChatGPT* and *Gemini* has introduced a level of agency and interactivity that challenges traditional instructional boundaries. As these technologies become more pervasive, the role of the teacher is being redefined rather than replaced. For this technological evolution to be effective, it is imperative that teachers transition from being primary "knowledge providers" to becoming *"content curators," "facilitators," and "learning designers*" who can navigate the complex interplay between human intuition and machine intelligence. The successful implementation of these technologies does not rely solely on the sophistication of the algorithms but heavily depends on the *beliefs, readiness, and literacy* of the educators who wield them. This chapter explores the human dimension of AI in language education (AIED), examining how teacher perceptions shape technology adoption, the necessity of defining and developing AI pedagogical literacy, models for professional development, and the complex interplay of resistance, anxiety, and ethical responsibility.

6.2 Teacher Perceptions and Readiness for AI Integration

6.2.1 Beliefs as Predictors of AI Adoption

Teacher perceptions have emerged as one of the most decisive and consistently reported factors influencing whether AI technologies are meaningfully integrated into language classrooms. Across diverse educational contexts, empirical studies demonstrate that teachers' beliefs about AI fundamentally shape not only *whether* AI is adopted, but *how* it is pedagogically framed, positioned, and sustained over time. Teachers who conceptualize AI as a pedagogical enhancement tool—capable of supporting feedback,

differentiation, learner autonomy, and reflective practice—are markedly more likely to engage in exploratory, experimental, and iterative forms of AI-supported instruction (Dinh, 2025). In contrast, teachers who interpret AI primarily through a lens of professional threat or epistemic risk tend to either avoid AI altogether or restrict its use to low-impact, peripheral functions.

This belief-driven divergence aligns with long-standing research in educational technology adoption, which emphasizes that teachers' subjective interpretations of technology often outweigh objective measures of institutional readiness or tool availability. In the context of language education, these beliefs are particularly influential because teaching practices are closely tied to deeply held convictions about interaction, communicative competence, assessment validity, and the social nature of language learning. When AI is perceived as compatible with these pedagogical values, teachers are more inclined to integrate it as a complementary resource. However, when AI is perceived as reductive—encouraging shortcut learning, surface-level production, or automation of cognitive effort—it is frequently framed as pedagogically misaligned, regardless of its technical sophistication.

Deficit-oriented beliefs constitute a recurring theme in studies examining teacher resistance to AI. Such beliefs often frame AI as undermining language learning rigor by promoting dependency, diminishing learner struggle, or weakening critical thinking skills. In these cases, AI is positioned as a threat to academic integrity, assessment authenticity, and the teacher's evaluative authority. Research indicates that these perceptions commonly result in either symbolic adoption—where AI tools are acknowledged but rarely used—or tightly controlled implementation, limited to grammar correction or mechanical practice (Tran et al., 2025). These findings suggest that resistance to AI is less about technological incompetence and more about unresolved pedagogical and epistemological concerns.

Empirical studies demonstrate that teachers' beliefs about AI fundamentally shape not only *whether* AI is adopted, but *how* it is pedagogically framed, positioned, and sustained over time. Research utilizing the *Technology Acceptance Model (TAM)* indicates that teachers' behavioral intentions are substantially influenced by self-efficacy, expected benefits, and ease of use (Al Darayseh, 2023). Teachers who conceptualize AI as a

pedagogical enhancement tool—capable of supporting feedback, differentiation, learner autonomy, and reflective practice—are significantly more likely to engage in exploratory and iterative forms of AI-supported instruction (Dinh, 2025).

Quantitative technology-acceptance research sharpens this picture by identifying which predictors matter most, and how much variance they explain. Using an integrated UTAUT2–TAM framework with 326 surveyed faculty and 22 interviewees at Shanghai University, Gao and Yang (2026) found that performance expectancy, perceived usefulness, institutional support, and AI self-efficacy each predicted teachers' adoption attitude, while perceived risk suppressed it—evidence that beliefs operate alongside, not instead of, structural support. A UTAUT2 study of 425 teachers at an Ecuadorian university reported high acceptance of educational AI that correlated positively with constructivist pedagogical beliefs and with teaching modality, directly linking technology acceptance to the pedagogical commitments discussed above (Cabero-Almenara et al., 2024). Among 498 pre-service teachers surveyed across eight education departments, an extended UTAUT model found that AI self-efficacy and digital literacy jointly predicted 51.4% of the variance in generative-AI acceptance (Bulut, 2026), while a TAM2-based study of Moroccan preservice teachers evaluating AI for teaching students with ADHD identified perceived usefulness as the single strongest predictor of acceptability (Amouri et al., 2025). These studies do not converge on one dominant predictor, however: a survey of 164 teachers across school levels in Northern Cyprus found that practical knowledge of AI tools, not beliefs or demographic variables, best predicted actual classroom AI use (Güneyli et al., 2024), suggesting that the relative weight of belief-based versus competence-based predictors is itself context-dependent.

Self-efficacy recurs as a predictor across otherwise disparate contexts, though rarely as a sufficient explanation on its own. A mixed-methods survey of 416 Thai K-12 teachers found that self-efficacy strongly predicted generative AI adoption ($r = .68$, $p < .001$), even as the same teachers raised ethical and critical-thinking concerns and cited infrastructure gaps as a constraint on follow-through (Tongchai & Malakul, 2025). Model-testing studies also caution against assuming that every theoretically posited predictor performs equally well across settings: a combined technology-acceptance and planned-behavior model applied to Indian university teachers and

students found empirical support for only nine of thirteen hypothesized paths predicting intention to adopt AI-based robots in education (Roy et al., 2022). Belief, in short, remains necessary but not sufficient: self-efficacy, practical knowledge, and institutional conditions each carry independent explanatory weight, and their relative contribution shifts from one educational system to the next.

Conversely, teachers who interpret AI through a lens of professional threat or epistemic risk tend to restrict its use to low-impact, peripheral functions. Interestingly, many language instructors possess an *"unconscious familiarity"* with AI through daily tools like Siri or Netflix, yet they often fail to equate these experiences with the AI competencies required for teaching (Kohnke et al., 2023). This lack of confidence often stems from a *"black box" phenomenon*, where teachers use AI tools without understanding their underlying mechanisms, leading to a diminished sense of control and self-efficacy (Weng & Chiu, 2023). This belief-driven divergence often results in a *dichotomy of trust and suspicion*; for instance, a study of instructors in Lebanon found that 67% expressed distrust toward AI, viewing it as a vehicle for plagiarism and *"intellectual laziness"* (Dakakni & Safa, 2023)

Recent bibliometric analyzes further substantiate the centrality of teacher beliefs within the broader AI-in-language-education research landscape. Large-scale mapping studies identify *teacher acceptance, readiness, and professional identity* as dominant research clusters, closely interconnected with themes of pedagogical alignment, ethical governance, and sustainable implementation (Yang et al., 2025; Zhang et al., 2025). Importantly, these analyzes reveal a shift in scholarly focus: earlier research tended to emphasize technological affordances, whereas more recent studies foreground the human and interpretive dimensions of AI integration. This shift reflects growing recognition that AI adoption is not a linear or purely technical process but a socially mediated practice shaped by beliefs, values, and institutional cultures.

Essentially, teacher readiness cannot be reduced to technical proficiency or digital skills alone. While operational competence is necessary, research increasingly emphasizes that readiness also encompasses epistemological beliefs about how languages are learned, what constitutes valid evidence of proficiency, and how knowledge should be constructed and evaluated in AI-mediated environments. Teachers' conceptions of assessment, for

instance, strongly influence their willingness to integrate generative AI into writing or speaking instruction. Educators who view *assessment as a process-oriented*, *formative practice* are more likely to see AI feedback as a scaffold for revision and reflection, whereas those who prioritize *product-oriented* or *summative assessment* may perceive AI as compromising evaluative fairness.

Beyond this, teachers' beliefs about their own professional role play a pivotal mediating function. Educators who conceptualize themselves as facilitators, designers, and guides within learning ecosystems tend to interpret AI as a co-pedagogical resource that can redistribute routine tasks while preserving human judgment. By contrast, teachers whose professional identity is closely tied to content authority or error correction may experience AI as an encroachment on their expertise. This tension highlights the need to understand AI adoption as an identity negotiation process rather than a simple skills-acquisition challenge.

Taken together, existing research highlights that teacher beliefs operate as both enablers and constraints in AI integration. They shape not only adoption decisions but also the depth, quality, and ethical framing of AI use in language classrooms. For this reason, any effort to promote sustainable and pedagogically sound AI integration must address teachers' beliefs explicitly—through reflective dialogue, theory-informed professional development, and opportunities to critically examine how AI aligns with core principles of language education. Without such engagement, AI risks remaining either underutilized or pedagogically misapplied, regardless of its technological potential.

This pattern of qualified optimism recurs outside language education specifically, suggesting that the belief-adoption link identified here reflects a general feature of teacher cognition under technological change rather than a peculiarity of language teaching. Among 136 Jordanian science teachers, readiness to adopt AI was tempered by data-privacy concerns and limited resources or professional development, even where optimism about the technology itself was present (Alshorman, 2024). Similarly, 580 mathematics teachers in Abu Dhabi reported that AI tools increased student motivation and engagement, but also required more instructional effort, with perceived challenges varying by teachers' own qualifications (Tashtoush et al., 2024). Both findings echo the claim developed above: readiness is a joint function of belief and circumstance, and neither predicts adoption reliably on its own.

6.2.2 Contextual and Institutional Influences

While individual beliefs play a decisive role in shaping teachers' engagement with AI, these beliefs are neither formed nor enacted in isolation. Teacher perceptions and readiness for AI integration are deeply embedded within broader *contextual and institutional ecosystems* that mediate access, legitimacy, and pedagogical risk-taking. Research in language education consistently demonstrates that even teachers who hold positive beliefs about AI may experience constrained agency due to *institutional structures, policy environments, assessment regimes*, and *cultural norms* surrounding teaching and learning (Huang et al., 2025).

This claim is not merely theoretical. In a structural equation model of 385 Pakistani teachers and administrators, colleague influence and institutional support significantly shaped beliefs about integrating AI into education, indicating that even attitudinal predictors are themselves partly downstream of the institutional environment (Naeem et al., 2023). A comparative study makes this mechanism visible by holding language-teaching context constant and varying national setting: semi-structured interviews with 40 Iranian and Italian EFL teachers found that Italian teachers felt ready to integrate AI tools into their teaching, while Iranian teachers reported readiness barriers rooted not in professional attitude but in economic and infrastructural factors (Ghiasvand & Kogani, 2024). Because the gap in readiness here tracks national economic and infrastructural conditions rather than a gap in teachers' orientation toward AI as a pedagogical resource, the comparison isolates context, not belief, as the operative variable—complicating any account of readiness built primarily on belief change.

From a *teacher cognition* perspective, beliefs are understood as dynamic, experience-based constructs that evolve through interaction with contextual affordances and constraints. Borg's (2005) model of teacher cognition emphasizes that teachers' instructional decisions emerge from the interplay between prior learning experiences, professional education, classroom practice, and institutional expectations. In AI-mediated language education, this interplay becomes particularly salient, as teachers must reconcile emerging technologies with established curricular mandates and accountability systems. When institutional contexts implicitly discourage experimentation—through rigid syllabi, high-stakes testing, or ambiguous AI

policies—teachers' willingness to adopt AI should tend to diminish, regardless of their personal openness to innovation.

Institutional policy clarity has emerged as a critical variable in shaping teacher readiness. Studies conducted across higher education and secondary EFL contexts reveal that uncertainty regarding acceptable AI use, assessment integrity, and authorship often leads teachers to adopt defensive or risk-averse stances. In the absence of explicit institutional guidelines, teachers may perceive AI integration as professionally hazardous, particularly in relation to plagiarism, data privacy, and academic misconduct. As a result, AI use is either avoided or relegated to informal, unacknowledged practices that lack pedagogical coherence (Tran et al., 2025; Zhang et al., 2025).

Infrastructure and access further mediate institutional influence. Unequal access to reliable digital tools, learning management systems, and AI-enabled platforms continues to shape teachers' perceptions of feasibility and fairness. Teachers who adopt generative AI often frame their own practice in explicitly equity-oriented terms, even as they continue to face systemic barriers—unequal infrastructure, inadequate training, and restrictive institutional norms—that individual effort alone cannot overcome. In Moroccan higher education, an exploratory study of 15 ESP teachers found that integrating a 6Cs framework for inclusive AI learning design was hindered by low student motivation, material limitations, and time constraints (Belkbir, 2024). Research from under-resourced and rural contexts indicates that teachers often resist AI integration not due to ideological opposition but because of concerns about exacerbating educational inequities. When AI tools are available only to certain learners or depend on personal devices and paid subscriptions, teachers may view their use as ethically problematic and pedagogically exclusionary. These findings underline that readiness must be understood not just as individual competence but as institutionally enabled capacity.

The role of assessment culture is particularly pronounced in language education. In contexts where summative assessment and standardized testing dominate, teachers frequently report tension between AI-supported learning processes and institutional evaluation frameworks. AI tools that support drafting, feedback, or oral rehearsal may align well with communicative language teaching principles, yet appear incompatible with assessment

systems that prioritize product-based outcomes. This misalignment often leads teachers to compartmentalize AI use—allowing it for practice but excluding it from assessed tasks—thereby limiting its pedagogical impact.

These contextual dynamics can be further illuminated through *Diffusion of Innovations Theory (*Rogers, 2003), which provides a useful framework for understanding differential patterns of technology adoption among language teachers. According to this theory, adoption is influenced by perceived relative advantage, compatibility with existing practices, complexity, trialability, and observability. Institutional environments strongly shape all five attributes. For example, when institutions frame AI as aligned with curricular goals and provide structured opportunities for experimentation, AI is perceived as both compatible and trialable. Conversely, when AI is introduced without pedagogical framing or institutional endorsement, it is often perceived as complex, risky, and misaligned with professional norms.

Diffusion theory also highlights the social dimension of adoption. Teachers are more likely to integrate AI when they observe respected colleagues doing so successfully and when professional communities normalize reflective experimentation. Conversely, in institutional cultures characterized by surveillance, compliance, or punitive accountability, early adopters may be discouraged, slowing diffusion across the teaching community. This finding reinforces the importance of *organizational culture* in shaping not only adoption rates but also the depth and sustainability of AI integration.

Cultural and national educational traditions further mediate institutional influence. Comparative studies indicate that teachers working in centralized education systems may experience more top-down pressure but less pedagogical autonomy, whereas those in decentralized systems may enjoy greater freedom but lack systematic support. In both cases, institutional framing considerably affects whether AI is interpreted as a mandated tool, an optional resource, or an ethically ambiguous intrusion. These differences suggest that teacher readiness must be analyzed through *context-sensitive lenses*, rather than universal adoption models.

Cross-cultural interview evidence corroborates this pattern beyond the Iranian-Italian case. Six language teachers drawn from Eastern (Vietnam, Iran, Indonesia) and Western (Spain, Germany, UK) educational systems were uniformly positive about AI-assisted language learning in principle, yet expressed anxiety and diverging equity concerns tied to their respective

national contexts (Huertas-Abril & Palacios-Hidalgo, 2023). The convergence of positive orientation with divergent contextual concern reinforces this section's broader argument: national educational tradition shapes not whether AI is welcomed in principle, but what teachers judge to be safe, equitable, and feasible in practice.

Taken together, contextual and institutional influences act as powerful mediators between teacher beliefs and classroom practice. They shape what forms of AI use are perceived as legitimate, safe, and pedagogically worthwhile. Importantly, these influences do more than merely constrain teachers; they also have the potential to *enable belief transformation*. Institutions that invest in coherent AI policies, align assessment practices with process-oriented learning, and encourage professional dialogue around ethics and pedagogy can actively support shifts in teacher cognition. In such environments, AI is more likely to be perceived not as an external imposition, but as a shared pedagogical resource integrated into the collective professional identity of language educators.

6.3 AI Pedagogical Literacy for Language Teachers

6.3.1 Defining AI Pedagogical Literacy

AI pedagogical literacy extends well beyond functional digital competence or surface-level familiarity with AI tools. Rather, it refers to teachers' capacity to *critically evaluate, pedagogically align,* and *ethically govern* the use of AI within language education contexts. This form of literacy requires educators to move from instrumental engagement with AI toward reflective, theory-informed, and context-sensitive instructional decision-making. Mateos-Blanco et al. (2024) conceptualize AI pedagogical literacy as a *multidimensional professional competence* comprising technological awareness, pedagogical judgment, linguistic expertise, and ethical reasoning—dimensions that are inseparable in language teaching practice. Utilizing the *Technological Pedagogical Content Knowledge (TPACK)* framework, research identified that educators often feel least confident in their *Content Knowledge (CK)* regarding AI concepts, such as machine learning or data ethics (Kim & Kwon, 2023)

Recent work has begun to operationalize this construct rather than only defining it conceptually. Using exploratory factor analysis with 385 Chinese

EFL university students (α = 0.960) followed by confirmatory factor analysis with an independent sample of 420, Huang et al. (2025) validated a 22-item, four-dimension scale measuring AI literacy in second-language writing—comprising understanding, application, ethics and risk awareness, and self-efficacy—providing the field with a psychometrically tested instrument rather than a purely conceptual taxonomy. A complementary review of 19 empirical studies on language-teacher–ChatGPT collaboration, organized around UNESCO's AI competency framework, found that the construct empirically spans five recurring dimensions: technological foundations, ethical awareness, a human-centered mindset, pedagogical application, and ongoing professional development (Liu, 2025). These dimensions map closely onto the technological, pedagogical, ethical, and linguistic components proposed conceptually above, but ground them in coded evidence from actual teacher-AI interaction rather than definitional argument alone. A further measurement instrument developed for an adjacent purpose reinforces this shift toward operationalization: Rahimi (2025) developed and validated a scale of language teachers' AI-mediated instructional design-thinking competency, extending psychometric measurement from literacy itself to the design competencies that literacy is meant to enable.

At its core, AI pedagogical literacy involves understanding *what AI does*, *how it does it*, and *why it should—or should not—be used* for particular learning purposes. Unlike traditional educational technologies, generative AI systems actively participate in meaning-making processes by producing language, offering feedback, and simulating interaction. It follows that teachers must possess a foundational understanding of how AI systems generate linguistic output, including their reliance on probabilistic models, large-scale training data, and pattern recognition rather than communicative intent or experiential understanding. Without such awareness, there is a risk that AI-generated language may be misinterpreted as authoritative or pedagogically neutral.

A core component of this literacy for language educators is *prompt engineering*—the skill of formulating precise instructions to elicit high-quality linguistic output from an LLM. Literacy also involves understanding the distinction between human and artificial cognition, enabling teachers to guide students in **"***human-AI collaborations***"** rather than viewing the machine as a replacement. A literate teacher acts as a *validator*, judging when

AI feedback is helpful versus when it is *"hallucinated"* or *factually incorrect*. This literacy is an extension of pedagogical content knowledge, allowing teachers to determine the *"pedagogical appropriateness"* of a tool—ensuring it supports development rather than undermining valuable human skills like introspection or resilience (Adams et al., 2023)

From a language education perspective, AI pedagogical literacy also entails recognizing the *limitations of AI systems*, particularly in areas central to communicative competence. While AI tools often demonstrate high accuracy in grammatical form and lexical choice, research consistently highlights their shortcomings in pragmatic appropriateness, discourse sensitivity, intercultural nuance, and sociolinguistic variation. Teachers who are AI-literate are therefore better positioned to contextualize AI output, frame it as provisional rather than definitive, and guide learners in critically interrogating its suitability for specific communicative purposes.

Beyond this, AI pedagogical literacy encompasses the ability to design learning tasks that leverage AI as a *cognitive and metacognitive scaffold*, rather than as a substitute for learner effort. This requires teachers to align AI use with established principles of second language acquisition, such as meaningful interaction, noticing, feedback uptake, and reflective revision. In this sense, AI literacy is not an additional technical skill layered onto teaching practice but an extension of pedagogical content knowledge adapted to AI-mediated environments.

Ethical reasoning constitutes a central pillar of AI pedagogical literacy. Teachers must be able to navigate issues related to data privacy, authorship, transparency, and learner agency, particularly when AI systems blur boundaries between human and machine-generated language. Ethical AI use in language education demands that teachers articulate clear expectations, model responsible practices, and foster learners' critical awareness of AI's affordances and constraints. Thus, AI pedagogical literacy functions not only as a professional competency but also as a form of *pedagogical leadership* within digitally mediated learning spaces.

6.3.2 From Tool Awareness to Pedagogical Design

While initial engagement with AI in language education often begins at the level of tool awareness, research indicates that meaningful pedagogical transformation occurs only when teachers progress beyond instrumental

use toward *intentional instructional design.* Systematic reviews demonstrate that teachers with higher levels of AI pedagogical literacy are markedly more likely to integrate AI at *transformative levels*, where technology reshapes learning processes rather than merely automating existing practices (Mohebbi, 2025). In contrast, teachers with limited AI literacy tend to employ AI in substitutive or augmentative ways, such as grammar checking or content generation, with minimal impact on learning depth.

AI-literate teachers conceptualize AI not as an answer-generating mechanism but as a *pedagogical mediator* capable of supporting scaffolding, formative feedback, and reflective engagement. For instance, rather than using AI solely to correct linguistic errors, teachers design tasks that require learners to compare AI feedback with peer or teacher input, justify revision choices, and reflect on patterns of error or improvement. Such practices align AI use with metacognitive development and promote learner autonomy, a key objective in contemporary language education.

In speaking instruction, AI chatbots exemplify the shift from tool awareness to pedagogical design. When deployed without instructional framing, chatbots often function as informal conversational partners with limited pedagogical value. However, when embedded within task-based or interactionist frameworks, they can serve as dialogic tools that support rehearsal, negotiation of meaning, and fluency development. Research indicates that AI-supported speaking tasks are most effective when teachers structure interactions around communicative goals, provide post-task reflection, and integrate human feedback alongside AI-generated responses (Wang et al., 2025). In such designs, AI becomes part of a broader interactional ecology rather than a stand-alone substitute for human communication.

The functional stakes of this literacy are visible at the level of classroom interaction structure itself. A social network analysis of 50 Chinese university students in AI-assisted L2 courses found a dual-centric interactional structure, in which the teacher remained authoritative while AI functioned as a passive advisor; critically, students' AI literacy moderated the relationship between their interactional centrality and their achievement and enjoyment, meaning that the same pattern of AI-mediated interaction produced different learning benefits depending on how literately students engaged with it (Xun & Wang, 2026). If literacy moderates outcomes for learners interacting with AI, the corollary for teachers is direct: a teacher's own AI

pedagogical literacy conditions not only how AI is used, but how much of its potential benefit is realized in the interactional structure of the classroom.

Similarly, automated writing feedback systems acquire pedagogical significance only when teachers actively mediate their use. AI-literate educators guide learners in evaluating the quality, relevance, and appropriateness of AI feedback, encouraging them to accept, modify, or reject suggestions based on rhetorical purpose and audience awareness. This evaluative stance transforms AI from an authoritative corrector into a *dialogic partner* in the writing process, nurturing critical literacy and decision-making skills.

This evaluative competency has itself been theorized as a specific literacy. Jiang et al. (2024) propose '*multimodal feedback literacy'* as a dimension of L2 writing teachers' feedback literacy in the digital age, comprising feedback design, affordance recognition, orchestration across modalities, and ensemble competence in combining automated and human feedback sources. This framework gives the evaluative stance described above—teachers guiding learners to accept, modify, or reject AI suggestions—a more precise vocabulary: what is required is not generic critical thinking about AI, but a specific, teachable competency in orchestrating automated and human feedback as a coherent ensemble.

Feedback fatigue, not only feedback quality, motivates much teacher interest in automated tools. Hyland (2025) critically examines whether automated writing evaluation and generative AI feedback can relieve EFL teachers' well-documented feedback fatigue while still genuinely improving student writers rather than merely their texts—a question that this chapter's emphasis on active teacher mediation, rather than passive delegation, is directly relevant to answering.

In particular, the transition from tool awareness to pedagogical design also requires teachers to anticipate and manage potential risks, including overreliance, reduced learner effort, and inequitable access. AI pedagogical literacy enables teachers to establish boundaries for AI use, differentiate tasks based on learner needs, and maintain the primacy of human judgment in assessment and feedback. In this way, AI integration becomes a deliberate pedagogical choice rather than a reactive response to technological availability.

Concerns about eroded evaluative agency are not confined to formative feedback on drafts; they extend to summative scoring itself. Barrot (2026) argues that generative AI-driven automated essay scoring risks diminishing teachers' evaluative agency, and proposes an ecological framework in which teachers critically mediate AI-generated scores to mitigate bias and protect data privacy, rather than accepting them as authoritative—a concern this chapter's account of AI pedagogical literacy anticipates, and one taken up more fully in the assessment-theoretic treatment of Chapters 10 and 11.

Overall, the movement from *tool awareness* to *pedagogical design* represents a qualitative shift in how AI is positioned within foreign language education. It reflects a maturation of teacher practice in which AI is integrated not for efficiency alone but for its potential to deepen learning, support reflection, and enhance communicative competence. This shift makes clear the central argument of this chapter: that *AI's educational value is realized not through technology itself, but through teachers' informed, critical, and pedagogically grounded engagement with it*.

6.3.3 A Conceptual Model Linking AI Pedagogical Literacy, Task Design, and Learner Outcomes

To synthesize the preceding discussion, this section proposes a *conceptual model* that positions AI pedagogical literacy as the mediating construct between technological affordances and meaningful learner outcomes in language education. The model conceptualizes AI integration not as a linear input–output process but as a *pedagogically mediated cycle*, in which teacher expertise determines how AI is transformed into learning value.

At the foundation of the model lies *AI pedagogical literacy*, encompassing teachers' technological understanding, linguistic expertise, pedagogical judgment, and ethical awareness. This literacy functions as an enabling condition that shapes how teachers interpret AI capabilities and limitations. Rather than responding to AI tools reactively, AI-literate teachers engage in intentional pedagogical reasoning, asking not what AI *can* do, but what it *should* do in relation to learning objectives.

AI pedagogical literacy directly informs *instructional task design*, which constitutes the central operational layer of the model. At this stage, teachers translate AI affordances into structured learning tasks aligned with principles of second language acquisition, such as interaction, feedback, noticing,

and reflection. Task design mediates whether AI is used at a substitutive level (e.g., automated correction) or at a transformative level (e.g., dialogic interaction, scaffolded revision, metacognitive reflection). Crucially, this design phase embeds constraints, prompts, and reflection mechanisms that prevent uncritical reliance on AI output.

Well-designed AI-mediated tasks, in turn, shape *learner engagement and outcomes*. When AI is pedagogically framed, learners are more likely to demonstrate higher-order language use, strategic decision-making, and reflective awareness of language form and function. Empirical evidence suggests that such designs support not only linguistic accuracy and fluency but also learner autonomy, self-regulation, and critical digital literacy (Mohebbi, 2025; Wang et al., 2025).

Significantly, the model is recursive rather than static. Learner outcomes feed back into teachers' pedagogical knowledge, informing subsequent refinement of AI use and task design. In this way, AI pedagogical literacy evolves through reflective practice, classroom experience, and professional dialogue. The model thus reinforces the central premise of this chapter: *AI does not directly produce learning outcomes; teachers' pedagogical mediation does.*

This recursive framing is consistent with recent theoretical work on teacher-AI interaction more broadly. Drawing on a systematic literature review, Cukurova et al. (2025) propose a framework for synergistic teacher-AI interaction explicitly designed to guard against reduced teacher agency and deprofessionalization—risks that a purely input-output model of AI integration, treating teachers as conduits rather than mediators, would otherwise invite. The conceptual model proposed here can be read as a language-education-specific instantiation of that broader synergistic principle.

6.3.4 AI Pedagogical Literacy in Teacher Education Policy and International Frameworks

The concept of AI pedagogical literacy articulated in this chapter aligns closely with emerging international policy frameworks that emphasize human-centered, ethical, and pedagogically grounded AI integration in education. From a teacher education perspective, this alignment is particularly evident in European initiatives and *UNESCO*-led instruments such as the Beijing Consensus on Artificial Intelligence and Education (UNESCO,

2019), which increasingly frame AI competence as an extension of professional judgment rather than technical proficiency alone.

Within the *European Union's DigCompEdu framework* (Redecker, 2017), teachers' digital competence is defined across dimensions such as professional engagement, pedagogical integration, assessment, learner empowerment, and ethical responsibility. AI pedagogical literacy directly corresponds to these dimensions by foregrounding teachers' ability to select, adapt, and critically evaluate digital tools in relation to learning goals. Specifically, AI literacy extends DigCompEdu's emphasis on digital pedagogy by requiring teachers to understand algorithmic processes, data implications, and the epistemic status of AI-generated language. In this sense, AI pedagogical literacy can be understood as a *next-generation expansion* of digital competence, responsive to generative and autonomous technologies.

Similarly, *UNESCO's guidance on AI in education* (Miao et al., 2021) emphasizes the preservation of human agency, equity, transparency, and inclusivity in AI adoption. The teacher-centered model proposed in this chapter resonates strongly with UNESCO's assertion that educators must remain central actors in AI-mediated learning environments. AI pedagogical literacy operationalizes this principle by equipping teachers to act as ethical stewards, ensuring that AI supports—not replaces—human judgment, relational pedagogy, and cultural sensitivity in language education.

From a policy standpoint, these alignments suggest that AI pedagogical literacy should be embedded explicitly within *pre-service and in-service teacher education curricula*. Rather than treating AI as an optional technological add-on, teacher education programs are increasingly called upon to integrate AI literacy into coursework on methodology, assessment, ethics, and professional identity. Such integration ensures coherence between institutional policy, classroom practice, and international standards.

Equally, framing AI competence through pedagogical literacy rather than technical mastery addresses a critical policy challenge: scalability. Teachers are not expected to become AI engineers, but reflective professionals capable of making informed decisions about AI use. This framing supports sustainable professional development models that prioritize critical engagement, collaborative inquiry, and contextual adaptation over tool-specific training.

In policy terms, therefore, AI pedagogical literacy represents a *bridge concept*—linking classroom-level practice with institutional governance and international educational values. It allows policymakers, teacher educators, and practitioners to articulate shared expectations for AI use that are pedagogically grounded, ethically defensible, and aligned with broader goals of quality and inclusive education. Embedding this concept within teacher education frameworks is thus not merely desirable but essential for ensuring that AI integration in language education remains both innovative and human-centered.

In addition to European and UNESCO frameworks, the notion of AI pedagogical literacy articulated in this chapter also aligns closely with the *OECD's educational policy frameworks* (OECD, 2019a, 2019b), particularly those addressing teacher professionalism, future competencies, and the integration of AI into learning systems. The OECD's work on *Education 2030* and the *Future of Education and Skills* agenda emphasizes the development of teacher capacities that support agency, reflective judgment, and ethical decision-making in increasingly technology-mediated environments. Within this perspective, teachers are positioned not as implementers of prescribed technologies but as *professional designers of learning*, capable of evaluating digital tools in relation to disciplinary knowledge, learner needs, and societal values. AI pedagogical literacy resonates with the OECD's emphasis on *transformative competencies*, including critical thinking, responsibility, and the ability to navigate uncertainty, by framing AI use as a context for pedagogical reasoning rather than technical compliance. Moreover, OECD guidance on AI in education highlights the necessity of human-in-the-loop approaches, transparency, and equitable access—principles that are operationalized through teachers' informed mediation of AI tools in language classrooms. By embedding AI pedagogical literacy within teacher education and professional development, educational systems can respond to OECD priorities for resilient, future-ready teaching professions capable of leveraging AI while safeguarding educational quality, equity, and trust.

6.3.5 Synthesizing EU, UNESCO, and OECD Frameworks: Toward a Coherent Policy Vision for AI Pedagogical Literacy

Although developed within distinct institutional and geopolitical contexts, the policy frameworks articulated by the European Union, UNESCO, and the OECD reveal a striking convergence in how they conceptualize teachers' roles in AI-mediated education. Collectively, these frameworks

advance a shared vision in which AI integration is not framed as a technical modernization project, but as a *human-centered pedagogical transformation* that depends fundamentally on teachers' professional judgment, ethical responsibility, and instructional design capacity. When read together, they provide a coherent policy foundation for the concept of *AI pedagogical literacy* developed in this chapter.

Across EU policy discourse, particularly within the DigCompEdu framework and related ethical guidelines, teachers are positioned as reflective professionals responsible for aligning digital technologies with pedagogical intent, learner empowerment, and assessment integrity. AI competence is thus framed not as mastery of tools but as the ability to make informed, context-sensitive decisions about their educational use. This emphasis resonates strongly with the core premise of AI pedagogical literacy: that teachers must understand both the affordances and epistemic limitations of AI systems in order to integrate them meaningfully into language education.

UNESCO's contributions further strengthen this vision by foregrounding ethical governance, equity, and human agency as non-negotiable principles in AI adoption. UNESCO explicitly cautions against technologized determinism and advocates for teacher-centered AI ecosystems in which educators retain authority over curricular decisions, learner assessment, and cultural mediation. Within this framework, AI pedagogical literacy becomes a mechanism for operationalizing ethical principles in classroom practice, enabling teachers to critically mediate AI use while promoting learners' awareness of algorithmic influence, authorship, and bias.

The OECD complements these perspectives by embedding AI integration within a broader agenda of future-ready education systems and professionalized teaching work. Through initiatives such as Education 2030 and the Learning Compass, the OECD conceptualizes teachers as designers of learning environments capable of navigating uncertainty, complexity, and technological change. AI competence is thus linked to transformative professional capacities, including critical thinking, responsibility, and adaptive expertise. From this standpoint, AI pedagogical literacy is not a discrete skill set but an extension of teacher professionalism in digitally saturated educational landscapes.

Taken together, these frameworks converge on three key principles that directly inform AI pedagogical literacy in language education.

- First, *teacher agency* is consistently emphasized as central to AI integration, rejecting models that marginalize educators as mere implementers of technological systems.
- Second, *pedagogical alignment* is prioritized over technological novelty, highlighting the necessity of grounding AI use in disciplinary knowledge, learning theory, and instructional design.
- Third, *ethical and social responsibility* is foregrounded as an integral component of teacher competence, particularly in relation to equity, transparency, and learner autonomy.

This convergence has important implications for teacher education and professional development. Rather than introducing AI as an isolated technological topic, policy-aligned teacher education programs are increasingly called upon to embed AI pedagogical literacy across methodology, assessment, ethics, and reflective practice. Such an approach ensures coherence between international policy expectations and classroom-level implementation, while preserving the centrality of human judgment in language education. In this sense, the synthesis of EU, UNESCO, and OECD frameworks does not simply legitimize AI pedagogical literacy as a policy construct; it positions it as a foundational component of sustainable, ethical, and pedagogically sound AI integration in language teaching and learning.

6.4 Professional Development Models for AI Integration

6.4.1 Limitations of Traditional Training Approaches

Despite increasing institutional investment in AI-related training, traditional professional development models have proven largely insufficient for cultivating sustainable AI competence among language teachers. One-off workshops, short-term seminars, and tool-focused demonstrations continue to dominate institutional responses to AI integration. However, empirical evidence consistently indicates that such approaches rarely lead to enduring pedagogical change or deep professional learning (Dinh, 2025; Zhang et al., 2025).

This is not entirely a new problem. Kohnke and Green-Eneix (2025) situate ChatGPT within a longer lineage of 'teaching machines' dating to the 1960s—consistent with the CALL-to-MALL historical arc traced in Chapter 1—arguing that each wave of instructional technology has provoked a similar cycle of training that emphasizes operating the machine over

reasoning about its pedagogical role, a pattern this section's critique of tool-focused workshops reproduces empirically.

A key limitation of these models lies in their *instrumental orientation*. Training sessions frequently prioritize procedural knowledge—how to operate a specific AI tool or platform—while neglecting the pedagogical, linguistic, and ethical dimensions that determine meaningful classroom use. As a result, teachers may acquire surface-level familiarity with AI applications without developing the conceptual frameworks necessary to evaluate their relevance, limitations, or alignment with language learning objectives. This often leads to fragmented adoption, where AI tools are used sporadically or experimentally rather than systematically embedded within instructional design.

At the same time, traditional training models tend to be *decontextualized from teachers' everyday practice.* Generic workshops rarely address the specific curricular demands, learner profiles, assessment regimes, or institutional constraints faced by language teachers. Accordingly, teachers struggle to transfer newly acquired technical knowledge into their own classrooms, particularly when AI use raises unresolved questions about academic integrity, assessment validity, or learner autonomy. In the absence of contextual adaptation, AI integration remains abstract and peripheral.

Another critical limitation concerns the lack of *continuity and reflective support.* Research indicates that professional learning is most effective when teachers are given opportunities to revisit, refine, and critically reflect on their practices over time. However, one-off AI training initiatives typically offer little follow-up, peer collaboration, or feedback mechanisms. Teachers report that without sustained dialogue and mentorship, initial enthusiasm often fades, and AI tools are gradually abandoned or reduced to low-impact uses (Dinh, 2025).

Traditional models also tend to position teachers as *passive recipients of expert knowledge*, rather than as active agents capable of shaping AI integration according to pedagogical principles. This transmission-oriented approach conflicts with the complex decision-making demands of AI-mediated language education, where teachers must continuously interpret learner needs, evaluate AI output, and negotiate ethical boundaries. As a result, such models risk reinforcing dependency on commercial tools and

external expertise, rather than fostering professional autonomy and critical judgment.

Taken together, these limitations highlight a central insight: professional development for AI integration cannot be treated as a discrete technical intervention. Instead, it must be understood as a *long-term professional learning process* that engages teachers' beliefs, identities, and pedagogical reasoning. Without this shift, AI-related training is unlikely to produce meaningful or sustainable transformation in language education.

6.4.2 Toward Sustainable Professional Learning Frameworks

In response to the shortcomings of traditional approaches, contemporary research increasingly advocates for *sustainable professional learning frameworks* that reconceptualize AI integration as an ongoing, collaborative, and practice-oriented endeavor. Rather than focusing on isolated skill acquisition, these frameworks emphasize teachers' capacity to engage in continuous inquiry, experimentation, and reflection within their own instructional contexts. PD must address the "*black box"* issue by explaining the basic principles of AI operation. Understanding the limitations and data-driven nature of these systems empowers teachers to answer student questions and use the technology more creatively (Weng & Chiu, 2023).

To navigate the AI-integrated classroom, teachers require a specific set of competencies collectively known as *AI literacy*. This goes beyond mere technical skill; it involves the ability to critically evaluate AI technologies, communicate effectively with them, and use them as tools for problem-solving (Long & Magerko, 2020, as cited in Laupichler et al., 2023). For language teachers, this literacy is *multidimensional.* Kim and Kwon (2023) propose that AI competencies for teachers should be viewed through the Technological Pedagogical Content Knowledge (TPACK) framework. Their research with elementary teachers identified that educators felt least confident in their Content Knowledge (CK) regarding AI concepts (e.g., understanding machine learning or data ethics) compared to their pedagogical or technological knowledge. This suggests that AI literacy for teachers must include a foundational understanding of *what* AI is and *how* it processes data, not just how to operate a specific software interface (Kim & Kwon, 2023).

One prominent approach involves *communities of practice*, where language teachers collaboratively explore AI tools, share experiences, and negotiate pedagogical meaning. Within such communities, AI integration becomes a collective professional activity rather than an individual burden. Teachers benefit from peer modeling, mutual support, and the normalization of trial-and-error practices, which are particularly important when working with rapidly evolving technologies. Research suggests that communities of practice cultivate deeper pedagogical integration by enabling teachers to observe how AI can be adapted across different language skills, proficiency levels, and assessment contexts (Cheng & Wang, 2023; Kohnke et al., 2023; Mohebbi, 2025).

Another influential model is *design-based professional learning*, which positions teachers as designers of AI-enhanced learning tasks rather than consumers of pre-packaged solutions. In this model, teachers iteratively design, implement, evaluate, and refine AI-mediated instructional activities, drawing on learner feedback and reflective analysis. Design-based approaches align closely with AI pedagogical literacy, as they require teachers to critically assess how AI supports—or constrains—language learning processes such as interaction, feedback uptake, and metacognitive awareness. Importantly, this model supports the development of transferable pedagogical principles rather than tool-specific expertise. Teachers need training on how to redesign curricula to integrate AI. This involves not just using the tool, but understanding how to restructure learning objectives and assessments to accommodate AI's presence, such as moving from rote tasks to those requiring critical evaluation of AI outputs (Kim & Kwon, 2023; Dakakni & Safa, 2023).

Reflective inquiry also plays a central role in sustainable professional learning. Teachers engaged in reflective practice examine not only the effectiveness of AI tools but also their underlying assumptions about language learning, assessment, and teacher–learner roles. Such reflection is particularly vital in addressing ethical concerns, including data privacy, learner dependency, and equity of access. By embedding ethical reflection into professional learning, teachers are better equipped to make principled decisions about when and how AI should be used.

A growing body of literature further emphasizes the importance of *integrating AI pedagogical literacy into pre-service teacher education*. Rather

than treating AI as an optional or advanced topic, scholars argue that future language teachers should encounter AI-related issues as part of their core methodological and assessment training (Mateos-Blanco et al., 2024). Early exposure enables novice teachers to develop critical awareness, ethical sensitivity, and confidence in human–AI collaboration before entering the profession. This approach also supports coherence between pre-service preparation and in-service professional development, reducing fragmentation across career stages.

Emerging design-science work illustrates what context-aware pre-service preparation can look like in practice. Using design science research methodology, Nyaaba et al. (2026) developed and evaluated GenAITEd Ghana, a curriculum-aligned conversational AI agent built specifically to support responsible, culturally responsive teacher education within the Ghanaian curriculum context, rather than adapting a generic AI tool to local conditions after the fact. Such context-first design responds directly to the critique, raised in Section 6.4.1, that generic AI training is decontextualized from teachers' actual curricular and cultural circumstances; building cultural and curricular alignment into the tool itself, rather than leaving it to individual teachers to supply, represents one concrete answer to that critique.

Crucially, sustainable professional learning frameworks reposition teachers as *co-designers and knowledge producers* in AI-enhanced education. Teachers are encouraged to contribute to institutional dialogue, shape policy interpretation, and participate in the development of locally relevant AI practices. This participatory orientation not only strengthens professional agency but also aligns with international policy frameworks that emphasize teacher leadership, ethical governance, and pedagogical alignment.

To conclude, effective professional development for AI integration requires a paradigmatic shift—from episodic training toward *embedded, reflective, and collaborative professional learning*. Such frameworks acknowledge the complexity of AI-mediated language education and recognize teachers as central actors in transforming technological potential into meaningful, ethical, and sustainable learning experiences.

6.4.3 Distinguishing Productive from Superficial Adoption

A recurring pattern across the studies reviewed in this chapter is that positive attitudes toward AI and actual productive classroom use are only loosely coupled. In a professional-development intervention involving 452 K-12 educational professionals, 88% expressed willingness to adopt AI chatbots, yet only 31% had actually used one—a fifty-seven-point gap between stated openness and enacted practice that belief alone cannot explain (Bergdahl & Sjöberg, 2025). This gap raises the question of what, beyond willingness, separates teachers who integrate AI in ways that transform their instruction from those whose engagement remains nominal. A Q-methodological study addresses this question directly: forty teachers sorted forty-two statements about AI-generated feedback in second-language writing, yielding four distinct viewpoints ranging from instructor-led trust, through strategic resistance, to transformative embrace (Jinowat et al., 2026). The existence of qualitatively distinct postures, rather than a single continuum from skeptical to enthusiastic, suggests that adoption style is a matter of professional orientation as much as attitude strength.

Productive adopters, across the available evidence, share a structural feature: their engagement with AI unfolds within a sustained social or reflective practice rather than isolated trial and error. A qualitative study of nine Chinese EFL teachers examined how their digital literacy around large language models developed through sustained participation in a university professional learning community (Xue, 2024), consistent with the broader argument that collective, iterative engagement—rather than one-off exposure—is what converts awareness into literacy. Structured reflection appears to function similarly. In a four-week quasi-experiment, sixty in-service teachers who used a generative-AI agent to support self- or peer-directed reflection on their teaching showed significantly greater gains in reflection self-efficacy and instructional-design reflection than a conventional-technology control group, with no difference between the self- and peer-directed conditions (Cao et al., 2026). Similarly, twenty preservice elementary science teachers who used ChatGPT-assisted re-reflection on their practice produced deeper, clearer reflections and greater professional-identity development than a traditional-journaling comparison group (Abualrob, 2025); although drawn from outside language education, the finding generalizes a mechanism directly relevant to this chapter's argument—that

reflective mediation, not mere exposure to the tool, is what converts AI use into pedagogical learning.

The payoff of productive adoption is measurable in classroom relationships and outcomes, not only in teacher self-report. Semi-structured interviews with 35 Chinese EFL teachers found that AI-mediated L2 instruction improved classroom rapport between teachers and students, via increased student agency, reduced distance, and greater trust (Zhang, 2025). At a more granular level, a survey of 454 Chinese EFL university students found that teachers' AI behavior-management competence—their skill in structuring and moderating AI-mediated classroom activity—predicted students' willingness to communicate in English both directly and indirectly, via reduced extraneous cognitive load, with the effect strongest when AI nudges were personalized (Chen et al., 2026). These findings sharpen the contrast introduced in Section 6.2.1 between symbolic adoption, where AI is acknowledged but rarely used, and pedagogically mediated adoption: the difference is not simply whether AI is used, but whether the teacher actively manages the interaction it creates.

6.5 Resistance, Anxiety, and Ethical Concerns

6.5.1 Teacher Anxiety and Professional Identity

Despite growing scholarly and institutional interest in Artificial Intelligence for language education, teacher resistance remains a persistent and structurally significant barrier to meaningful adoption. Large-scale bibliometric and systematic reviews consistently report that teachers express anxiety related to potential job displacement, erosion of pedagogical authority, and the perceived devaluation of human expertise in AI-mediated learning environments (Yang et al., 2025; Zhang et al., 2025). These concerns are particularly salient in language education, where affective interaction, cultural mediation, and discourse sensitivity constitute core dimensions of teaching practice.

Empirical evidence at the level of individual teachers corroborates what the bibliometric mappings suggest in aggregate. Teachers frequently hold hindering and bolstering concerns about generative AI adoption at the same time, rather than resolving into a single anxious or enthusiastic profile, which suggests that professional development needs to be tailored to where

a given teacher actually stands rather than designed to move every teacher uniformly from resistance to acceptance. A phenomenographic study of 15 Iranian EFL teachers offers a finer-grained picture of this concurrent experience: AI reshaped participants' professional identity across six interrelated themes, including role transformation, pedagogical alignment, and a shift toward more self-reflective, technologically confident practice (Ghiasvand & Seyri, 2025).

Empirical studies examining teachers' responses to generative AI tools such as ChatGPT indicate that anxiety is frequently linked to fears of professional redundancy and diminished instructional legitimacy. Reviews of AI use in ESL and EFL contexts show that teachers often perceive AI-generated language as competing with their role as linguistic models and feedback providers, especially when institutional discourse emphasizes efficiency, automation, or productivity gains (Faiz & Hazilan, 2024). In such contexts, AI is not just experienced as a technological innovation but as a symbolic disruption of long-established professional norms.

It bears emphasizing that recent research cautions against interpreting resistance as technophobia or lack of digital competence. Instead, resistance is increasingly understood as an expression of *professional identity negotiation* in response to disruptive innovation (Yang et al., 2025; Zhang et al., 2025). Teachers' apprehension reflects deeper epistemological concerns about what constitutes legitimate language learning, authentic assessment, and meaningful interaction. For many educators, particularly those working within communicative or interactionist paradigms, AI raises unresolved questions about the role of struggle, feedback, and human judgment in language development.

Studies focusing on higher education and vocational EFL contexts further demonstrate that resistance intensifies when AI implementation is perceived as top-down, policy-driven, or inadequately supported by pedagogical dialogue (Zhang et al., 2025). In contrast, when AI adoption is framed as a teacher-guided, exploratory process and embedded within professional learning communities, anxiety tends to shift toward cautious engagement rather than outright rejection (Mohebbi, 2025). These findings underscore that resistance is not a fixed disposition but a contextually mediated response shaped by institutional culture and professional trust.

Not all apparent resistance reflects anxiety in the clinical sense; some reflects considered disagreement among stakeholders about AI's educational value. A survey of 329 students and 14 faculty at a Portuguese university found genuinely divergent views on generative AI's effect on skill development—roughly half of each group disagreed that AI harms competency development—yet both groups converged on the need for clear institutional guidelines regardless of their position on AI's net effect (Fernandes et al., 2025). This divergence-with-consensus pattern is consistent with the Q-methodological finding, discussed in Section 6.4.3, that one of the empirically distinct teacher viewpoints on AI-generated feedback is best described as strategic resistance rather than technophobic avoidance—teachers holding this viewpoint accept AI's technical competence but withhold trust for specific, articulable pedagogical reasons (Jinowat et al., 2026).

From this perspective, teacher anxiety functions as a diagnostic signal rather than a deficit. It highlights the need for professional spaces in which educators can critically examine how AI reshapes authority, expertise, and relational pedagogy. Addressing resistance therefore requires institutional recognition of teachers' emotional and ethical labor, alongside opportunities for reflective dialogue that reaffirm the irreplaceable human dimensions of language education.

6.5.2 Ethical Governance and Responsible AI Use

Ethical concerns constitute one of the most consistently reported dimensions of teacher apprehension toward AI integration in language education. Across systematic reviews, bibliometric mappings, and empirical syntheses, recurring ethical issues include data privacy, algorithmic bias, transparency of AI-generated output, authorship, and accountability for learning outcomes (Mateos-Blanco et al., 2024; Zhang et al., 2025; Yang et al., 2025). These concerns are not peripheral but central to teachers' willingness to engage with AI in instructional and assessment contexts.

Beyond documenting that ethical concerns are salient, recent research has begun to measure teachers' actual AI-ethics knowledge and to identify what predicts it. A survey of 116 Chinese educators found significant gaps in AI-ethics literacy overall, with university teachers scoring higher than their K-12 counterparts and professional development emerging as a strong predictor of ethical awareness (Zhang & Rhee, 2025). This finding reframes ethical concern not as a fixed disposition but as a trainable

competency: teachers most anxious about AI ethics may be precisely those who have not yet had the professional development that would resolve the anxiety into informed judgment. A PRISMA-guided systematic review of 34 studies mapped AI-ethics tensions across K-12 and higher-education stakeholders more broadly, proposing four recurring ethical- and literacy-framework domains for responsible AI integration (Yan & Chau, 2025), while a separate systematic review of 39 studies spanning K-12, higher education, and language learning identified training, equity, and ethics as the three persistent gap areas limiting responsible AI integration, despite documented gains in personalization and feedback (Adil, 2025).

In language education, ethical complexity is heightened by the generative nature of contemporary AI systems. Reviews focusing on ChatGPT and similar tools indicate that teachers frequently struggle with questions of authorship and originality when learners submit AI-assisted writing or speaking outputs (Faiz & Hazilan, 2024). Without clear pedagogical frameworks, AI-generated language risks undermining assessment validity and blurring boundaries between learner competence and machine assistance. Teachers thus assume a critical mediating role in defining acceptable AI use and maintaining academic integrity.

Data privacy and *surveillance concerns* further complicate ethical governance. Systematic reviews highlight that many AI tools collect and process sensitive learner data, including written texts, voice recordings, and interaction histories, often through opaque commercial platforms (Mateos-Blanco et al., 2024; Mohebbi, 2025). Teachers operating in the absence of clear institutional data governance policies report uncertainty regarding their ethical and legal responsibilities, which in turn discourages experimentation with AI-supported pedagogy.

Algorithmic bias and *lack of transparency* represent additional ethical challenges. Bibliometric analyzes show increasing scholarly attention to the ways AI systems may reproduce dominant linguistic norms, marginalize non-standard varieties, or reinforce cultural stereotypes embedded in training data (Yang et al., 2025; Zhang et al., 2025). Teachers with higher levels of AI pedagogical literacy are better equipped to contextualize AI output, encourage critical evaluation, and address instances where AI-generated language conflicts with communicative appropriateness or intercultural sensitivity. Teachers often worry that AI models trained on *"WEIRD"*

(Western, Educated, Industrial, Rich, and Democratic) values may penalize learners using non-standard linguistic varieties. There is also a fear that *over-reliance on AI* could erode critical thinking skills and student cognition (Segbenya et al., 2023)

This concern is not speculative. Using design-and-development research to test three distinct ChatGPT configurations for English language teaching, Lee et al. (2025) found consistent pedagogical limitations in representing pluralistic Global Englishes, with outputs defaulting toward standardized forms rather than the varieties learners actually encounter. Because such defaults are largely invisible unless a teacher is specifically attentive to them, the finding underscores why algorithmic bias is properly treated as an ethics-literacy problem rather than a purely technical one: recognizing the bias requires exactly the sociolinguistic awareness that AI pedagogical literacy, as defined in Section 6.3, is meant to cultivate.

Rather than advocating prohibition or strict exclusion, the literature increasingly supports *guided, transparent, and ethically framed integration* of AI tools in language education, where teachers act as ethical role models and stewards of a **"***humanized posthuman ethic"* that protects both student rights and the "*human touch*" of education (Adams et al., 2023; Mohebbi, 2025; Mateos-Blanco et al., 2024). In this model, teachers act as ethical role models who openly discuss the affordances and limitations of AI, establish explicit usage guidelines, and promote learners' critical AI awareness. Such practices align with broader policy recommendations emphasizing human-in-the-loop approaches and responsible innovation.

Ultimately, ethical governance in AI-mediated language education is enacted at the classroom level through teachers' pedagogical decisions and professional judgment. While institutional frameworks and policy guidelines are essential, the ethical quality of AI use depends on teachers' capacity to interpret, mediate, and humanize AI-supported learning. Recognizing and supporting this ethical labor is therefore fundamental to addressing resistance, alleviating anxiety, and ensuring that AI integration remains aligned with the core values of language education: equity, authenticity, and human connection.

6.6 Conclusion

This chapter has demonstrated that teachers remain central to the pedagogical, ethical, and emotional dimensions of Artificial Intelligence integration in language education. Across empirical studies, systematic reviews, and bibliometric analyzes, teacher beliefs, professional identity, and AI pedagogical literacy emerge as decisive factors shaping how AI tools are interpreted, adopted, adapted, or resisted. Far from being neutral conduits of technological innovation, teachers actively mediate AI through their pedagogical judgments, ethical reasoning, and contextual understanding of language learning. As such, AI integration in language education must be understood as a *human-centered, professionally situated process*, rather than a technologically deterministic transformation.

The chapter has further shown that *teacher readiness* extends well beyond technical proficiency. Readiness encompasses epistemological beliefs about language learning, assessment validity, learner autonomy, and the role of human interaction in communicative development. *Resistance* and *anxiety*, often framed as obstacles to innovation, have been reinterpreted here as meaningful expressions of professional identity negotiation in response to disruptive change. When acknowledged and addressed through reflective dialogue and institutional support, such responses can become catalysts for more thoughtful, ethical, and pedagogically grounded AI adoption.

The readiness challenge documented throughout this chapter is unlikely to diminish as AI systems become more autonomous. A survey of 75 university educators examining perceptions of agentic AI—systems capable of planning and executing multi-step tasks with reduced human prompting at each stage—found that enthusiasm about potential applications coexisted with explicit concern about governance mechanisms and institutional readiness for adoption (Brohi & Jhanjhi, 2026). As Chapter 15 explores in more detail, the shift from generative to agentic AI does not resolve the readiness questions raised here; it raises their stakes, since a teacher's beliefs, literacy, and institutional support must now extend to systems that act with less continuous human oversight at each step.

Professional development models play a pivotal role in this process. Traditional, tool-focused training approaches are insufficient for preparing teachers to navigate the complexities of AI-mediated language education.

Instead, *sustainable professional learning frameworks*—characterized by *continuity, collaboration, design-based inquiry,* and *ethical reflection*—are essential for cultivating AI pedagogical literacy. These frameworks position teachers as co-designers of AI-enhanced pedagogy, capable of translating technological affordances into learning tasks that support higher-order language use, metacognitive awareness, and learner agency.

It should be noted that the arguments advanced in this chapter align closely with and are reinforced by major international policy frameworks developed by the European Union, UNESCO, and the OECD. Across these frameworks, a shared policy vision emerges: AI integration in education must preserve human agency, pedagogical autonomy, and ethical responsibility. The EU's DigCompEdu framework emphasizes teachers' role in pedagogical alignment, learner empowerment, and ethical digital practice. UNESCO foregrounds human-centered AI, equity, transparency, and the safeguarding of cultural and linguistic diversity. The OECD, through its Education 2030 agenda and work on teacher professionalism, positions educators as designers of learning environments equipped to navigate complexity, uncertainty, and technological change. Together, these frameworks converge on a clear principle: *AI competence in education is fundamentally a pedagogical and ethical competence, more than merely a technical one.*

By synthesizing research evidence with these international policy perspectives, this chapter argues that AI pedagogical literacy should be recognized as a core component of contemporary language teacher professionalism. Embedding this literacy within pre-service education, in-service professional development, and institutional policy is essential for ensuring coherence between classroom practice and global educational values. Such alignment enables teachers to harness AI's potential while critically managing its risks, thereby safeguarding educational quality, equity, and trust.

At its core, AI does not redefine language education on its own. It is teachers' informed, critical, and value-driven engagement with AI that determines whether these technologies contribute to meaningful, inclusive, and human-centered learning experiences. Recognizing and supporting teachers as pedagogical leaders, ethical stewards, and designers of AI-mediated language learning is therefore not a peripheral concern, but a foundational condition for the responsible future of AI in language teaching, learning, and assessment.

Case Box 6.1—Preparing language teachers for technology-rich instruction: from augmented reality to digital competence assessment

The professional development challenges analyzed in this chapter have a concrete counterpart in teacher education practice in Türkiye. In the context of English-medium instruction (EMI), Can and Rey (2022) examined the training of language teachers through augmented reality applications, illustrating how emerging technologies can be embedded in teacher preparation not as objects of instruction but as environments for it—the same reframing this chapter argues for with respect to AI. Complementing the design strand, Can and Karacan (2025) validated a DigCompEdu-based scale for assessing pre-service language teachers' digital competences, providing the kind of measurement infrastructure that AI-era professional development requires: if AI pedagogical literacy is to be developed systematically (Section 6.3), teacher educators need validated instruments to diagnose where candidate teachers stand and to evaluate whether interventions work. Together, these strands exemplify the position developed in this chapter: teacher preparation for AI-rich classrooms is continuous with two decades of technology-in-teacher-education research, and its success is measurable, not merely aspirational.

Pedagogical Companion for Chapter 6

Key Concepts: teacher beliefs; readiness and adoption; AI pedagogical literacy; DigCompEdu; professional development models; teacher anxiety; professional identity; ethical governance.

Discussion Questions

1. The chapter argues that beliefs predict AI adoption better than access does. What experiences formed your own beliefs about AI, and how revisable are they?
2. Distinguish AI *tool awareness* from AI *pedagogical literacy* using two concrete teacher behaviors for each.
3. Why do one-shot training workshops fail to produce durable change in AI practice, and what structural features would a sustainable alternative require?

4. Teacher anxiety about AI is often framed as a deficit to be remedied. When might such anxiety be a professionally appropriate response?
5. Case Box 6.1 links AI-era teacher preparation to two decades of technology-in-teacher-education research. What, if anything, is genuinely unprecedented about preparing teachers for AI?

Further Reading

- Hubbard, P. (2008). CALL and the future of language teacher education. *CALICO Journal, 25*(2).
- Hubbard, P., & Schulze, M. (2025). AI and the future of language teaching. *IJCALLT, 15*(1).
- Miao, F., & Cukurova, M. (2024). *AI competency framework for teachers*. UNESCO.

PART III

AI in Language Learning

Chapter 7. AI-Mediated Learning Environments: Intelligent Tutors, Conversational Agents, and Adaptive Systems

7.1 Introduction

The rapid expansion of Artificial Intelligence (AI) in foreign language education has transformed learning environments from static, curriculum-driven spaces into *dynamic, adaptive ecosystems*. Unlike earlier generations of Computer-Assisted Language Learning (CALL), which primarily functioned as supplementary drill-and-practice tools, contemporary AI-supported environments are characterized by personalization, interactivity, and data-driven adaptation. These environments integrate *intelligent tutoring systems, conversational agents, mobile applications,* and *immersive technologies* to support learners' linguistic development across skills, proficiency levels, and sociocultural contexts (Kartal & Yeşilyurt, 2024; Jaleniauskienė et al., 2023).

This transformation reflects a broader paradigmatic shift from technology as a *delivery mechanism* toward technology as a *learning mediator*. AI-supported environments no longer merely present linguistic input; rather, they analyze learner behavior, model developmental trajectories, and dynamically shape instructional pathways in response to learner needs. From an applied linguistics perspective, this shift aligns with long-standing calls for individualized input, meaningful interaction, and timely feedback—core principles of second language acquisition (SLA) theory that were historically difficult to operationalize at scale (Wu, 2024).

Recent research syntheses and bibliometric mappings consistently identify *intelligent tutoring systems (ITS), AI-driven conversational agents, mobile-assisted language learning (MALL),* and *adaptive personalization* as dominant clusters in AI-mediated language learning research (Kartal & Yeşilyurt, 2024; Yang et al., 2025). These clusters illustrate the field's movement away from isolated tools toward *integrated learning environments* that combine instructional scaffolding, communicative practice, and learner analytics within unified systems. Importantly, this convergence also reflects increasing interdisciplinary integration, drawing on insights from SLA, learning sciences, human–computer interaction, and educational data mining.

A key finding is that, however, contemporary scholarship emphasizes that the pedagogical effectiveness of AI-supported learning environments does not stem from automation alone. Rather, effectiveness is contingent upon alignment with SLA principles, the promotion of learner autonomy, and the presence of human-guided instructional design. Systematic reviews demonstrate that AI environments are most beneficial when they function as *augmentative systems*—enhancing learner engagement, feedback quality, and self-regulation—rather than as autonomous instructional authorities (Mohebbi, 2025; Wu, 2024). This perspective challenges deterministic narratives of AI-driven efficiency and foregrounds the continued centrality of pedagogical intent and teacher mediation.

Beyond this, AI-supported language learning environments raise important questions concerning agency, ethics, and educational responsibility. As systems increasingly rely on learner data to drive adaptation, issues such as algorithmic bias, data privacy, and over-reliance on automated feedback become pedagogically salient. These concerns necessitate a shift toward *human-in-the-loop models*, where teachers remain responsible for interpretive judgment, ethical oversight, and the orchestration of learning experiences (Yang et al., 2025).

Against this backdrop, this chapter examines four core dimensions of AI-supported language learning environments that have emerged as particularly influential in both research and practice:

1) *Intelligent tutoring systems*, which operationalize individualized instruction and feedback through learner modeling;
2) *Conversational agents and chatbots*, which simulate interactional opportunities and support communicative practice;
3) *Mobile and immersive AI environments (VR/AR)*, which situate language use within embodied and contextualized experiences; and
4) *Adaptive and personalized learning pathways*, which integrate data-driven decision-making to tailor instruction over time.

By synthesizing empirical findings, theoretical perspectives, and critical debates, this chapter aims to clarify not only how AI-supported environments function, but also under what conditions they meaningfully contribute to language learning. In doing so, it positions AI not as a replacement for pedagogical expertise, but as a powerful—yet limited—mediator whose educational value depends on informed, ethical, and theoretically grounded integration.

7.2 Intelligent Tutoring Systems in Language Learning

Intelligent Tutoring Systems (ITS) constitute one of the most mature and theoretically grounded applications of Artificial Intelligence in language learning. Originating from early work in cognitive science and expert systems, ITS were designed to emulate the adaptive guidance of a human tutor by diagnosing learner performance, modeling knowledge states, and delivering individualized feedback. In contemporary language education, ITS integrate machine learning, natural language processing (NLP), and learner analytics to provide real-time, adaptive instruction across linguistic domains such as grammar, vocabulary, pronunciation, and writing (Kartal & Yeşilyurt, 2024).

From a *Second Language Acquisition (SLA)* perspective, the pedagogical value of ITS lies in their capacity to operationalize several foundational theoretical principles that are otherwise difficult to implement at scale. In particular, ITS align strongly with *cognitive and interactionist SLA frameworks*, where learning is understood as a gradual restructuring of linguistic knowledge through input, feedback, and practice.

7.2.1 ITS and Cognitive–Skill Acquisition Models

Within *skill acquisition theory* (DeKeyser, 2007), language learning is conceptualized as a *progression from declarative knowledge (explicit rules) to proceduralized and automatized performance through repeated, meaningful practice*. ITS are particularly well-suited to this model, as they can deliver sequenced exercises, monitor response accuracy and speed, and adjust task difficulty based on learner performance.

Intelligent Tutoring Systems (ITS) represent one of the most established applications of AI in education, designed to simulate the pedagogical strategies of a human tutor. Structurally, an ITS typically comprises four components: a *student model* that tracks the learner's knowledge state; a *knowledge model* containing the domain expertise; a *tutor model* that decides pedagogical strategies; and a *user interface* for interaction (Graesser et al., 2005).

In language learning, ITSs provide bespoke instruction and immediate feedback, allowing learners to bridge the gap between theoretical knowledge and practical application (Gligorea et al., 2023). Unlike traditional computer-assisted instruction, modern ITSs utilize deep learning to

analyze complex learner behaviors. For instance, systems like *AutoTutor* have long demonstrated efficacy by engaging students in natural language dialogues that scaffold learning through questioning and analyzing student responses (Graesser et al., 2005, as cited in Li et al., 2025).

Recent advancements have seen ITSs evolve from rule-based architectures—which often struggle with scalability and flexibility—to generative-based models powered by Large Language Models (LLMs) (Li et al., 2025). These advanced systems can generate personalized feedback on open-ended tasks, such as essays or spoken responses, rather than relying solely on pre-programd answers. For example, ITS platforms like *ALEKS* (Assessment and Learning in Knowledge Spaces) utilize adaptive questioning to map a student's knowledge state precisely, offering a personalized path to mastery that has been shown to improve outcomes in STEM and language-related logic tasks (Hagerty & Smith, 2005, as cited in Li et al., 2025). Further, researchers are now integrating computer vision into ITS interfaces to detect non-verbal cues, such as gaze and facial expressions, to predict learner engagement and help-seeking behaviors, thereby mimicking a human teacher's ability to "read the room" (Wang, Hatori, & Satō, 2025).

Early evidence on such multimodal designs is encouraging, even if still preliminary. A review of 46 studies published between 2020 and 2025, combined with a small-scale experiment (n = 20), found that a multimodal AI pedagogical agent functioning simultaneously as tutor, coach, and evaluator improved performance, reduced errors, and boosted engagement (Jansone & Cīrule, 2026). A parallel PRISMA-based review of multi-agent cognitive architectures for adaptive tutoring reached a similar conclusion, finding superior individualized feedback compared with conventional methods and recommending neuroadaptive and emotion-responsive tutoring agents as directions for future work (Gupta & Heggond, 2026).

Empirical studies indicate that ITS-supported practice facilitates the automatization of grammatical structures and lexical retrieval, especially in form-focused domains such as verb morphology, syntactic patterns, and collocations (Xu & Ismail, 2025). Unlike static CALL exercises, ITS dynamically adapt practice schedules, thereby supporting spaced repetition and individualized pacing—both critical mechanisms in cognitive learning theories.

The magnitude of such gains is not merely theoretical. In a six-week controlled experiment with 100 Chinese university students, an adaptive AI

smart-classroom framework built on a hidden Markov model raised English translation accuracy by 8.3 percent and engagement by 12.7 percent relative to traditional instruction, with 89.2 percent of students reporting satisfaction with the system (Chen & Hu, 2026). At a broader scale, a PRISMA-guided synthesis of 22 empirical studies drawn from 189 screened records found that AI tools produced significant gains across reading, writing, listening, speaking, and vocabulary in school-age EFL settings—though the same review cautioned that digital-divide constraints and uneven teacher readiness limited how consistently those gains were realized across school contexts (Kundu & Bej, 2025). Automatization gains of this kind, in other words, are conditional on infrastructure and implementation quality, not guaranteed by the technology itself.

However, SLA scholars caution that skill acquisition alone does not account for communicative competence. As such, ITS are most effective when embedded within broader instructional designs that connect form-focused practice to communicative use such as *Apilingua App*, rather than functioning as isolated drill systems.

7.2.2 ITS, Feedback, and the Noticing Hypothesis

Another central theoretical alignment emerges with *Schmidt's* (1990) *Noticing Hypothesis* which posits that conscious attention to linguistic form is a necessary condition for acquisition. ITS excel in this domain by providing *immediate, explicit, and targeted feedback*, drawing learners' attention to discrepancies between their interlanguage and target language norms.

Automated feedback in ITS may include error highlighting, metalinguistic explanations, reformulations, or guided prompts. Systematic reviews show that such feedback enhances accuracy in writing and grammar tasks, particularly when it is adaptive rather than generic (Wu, 2024). From an SLA standpoint, this supports the idea that noticing is facilitated not merely by exposure, but by *diagnostic feedback that is contingent on learner output*.

Nevertheless, scholars emphasize that excessive explicit feedback may encourage surface-level correction rather than deep restructuring of interlanguage. This reinforces the need for pedagogically calibrated feedback strategies, where ITS complement—rather than replace—teacher-led reflection and interaction.

7.2.3 ITS and Interactionist SLA Perspectives

From an *interactionist perspective*, language acquisition is driven by meaning negotiation, feedback, and modified input during interaction (Long, 1996). While ITS do not typically support open-ended human interaction, advanced systems increasingly incorporate dialogic elements, such as scaffolded prompts, clarification requests, and adaptive task sequences that simulate interactional contingencies.

Research suggests that ITS can partially approximate interactional feedback—particularly in pronunciation training and writing—by responding to learner output with context-sensitive prompts (Xu & Ismail, 2025). However, interactionist SLA theory also highlights the importance of social presence, pragmatic nuance, and shared meaning-making, which remain difficult for ITS to fully replicate.

As a result, contemporary research frames ITS not as substitutes for communicative interaction, but as *preparatory and supportive environments* that strengthen linguistic accuracy and confidence prior to human interaction. This positioning aligns with hybrid pedagogical models that integrate ITS with classroom discussion, peer collaboration, and teacher mediation.

7.2.4 Learner Modeling, Autonomy, and Self-Regulated Learning

A defining feature of ITS is *learner modeling*—the system's ability to construct evolving representations of a learner's knowledge state, error patterns, and progress. From the perspective of *self-regulated learning (SRL)* and autonomy-oriented SLA frameworks, this capability holds significant pedagogical potential.

By visualizing progress, recommending next steps, and adjusting instructional pathways, ITS can support metacognitive awareness and learner agency. Systematic reviews indicate that AI-supported environments, including ITS, contribute positively to *goal-setting, self-monitoring,* and *persistence*—key components of autonomous language learning (Mohebbi, 2025).

The relationship between system authority and learner agency has itself become an object of empirical study. Analyzing more than 7,000 discussion threads and over 117,000 utterances generated through interaction with a generative AI teachable agent, one study found that the degree of authority the agent assumed in a dialogue was significantly associated with students'

own sense of agency and with their learning outcomes (Xing et al., 2026). The implication is that learner modeling can either crowd out or cultivate self-regulation, depending on how much interpretive control the system claims for itself. Population matters as much as design here: in a three-semester comparative study, an AI reflection agent produced higher-quality, more critically connective reflections than conventional reflection reports or short-answer prompts, but the gain was concentrated among low-performing teams, which showed the largest improvements (Zheng et al., 2026). Learner modeling, in other words, does not raise all learners equally.

However, autonomy is not an inherent outcome of adaptivity. Scholars caution that opaque algorithmic decisions may reduce learners' sense of control or encourage passive compliance. Therefore, effective ITS design should prioritize *transparency, learner choice, and reflective prompts*, ensuring that adaptivity supports—not undermines—self-directed learning.

7.2.5 Pedagogical Limitations and the Role of the Teacher

Despite their strengths, ITS exhibit notable limitations from an SLA and educational standpoint. They remain largely constrained in addressing pragmatic competence, discourse organization, intercultural meaning, and affective dimensions of language use. Equally, ethical concerns related to data privacy, bias in learner modeling, and over-reliance on automated correction necessitate critical oversight (Yang et al., 2025).

A review of *agentic architectures* for *autonomous tutoring* corroborates this picture from the design side: even as multi-agent architectures and emotion-responsive instruction have advanced, transparency, data protection, and equitable access remain persistent, unresolved gaps rather than solved engineering problems (Meti, 2026). Among child second-language learners specifically, a mixed-methods study found that AI tools such as speech recognition and interactive environments enhanced L2 skills through immediate feedback, though ethical and overdependence risks remained (Alkhutaba & Alnajjar, 2026). Young learners are arguably least equipped to evaluate such feedback critically for themselves, which is one reason these limitations do not fall evenly across the age range that ITS and related tools serve.

Consequently, current scholarship converges on *human-in-the-loop models*, where ITS function as pedagogical instruments under teacher

supervision. In such models, teachers interpret system feedback, contextualize practice, and integrate ITS outputs into broader communicative and curricular goals. This approach reaffirms the teacher's role as instructional designer and ethical gatekeeper, rather than positioning AI as an autonomous authority.

In sum, Intelligent Tutoring Systems occupy a pivotal position within AI-supported language learning environments due to their strong alignment with cognitive, interactionist, and autonomy-oriented SLA frameworks. When theoretically grounded and pedagogically integrated, ITS can enhance accuracy, efficiency, and learner awareness. However, their educational value depends on thoughtful design, transparent adaptivity, and sustained human mediation.

Rather than representing a replacement for language teachers or communicative pedagogy, ITS are best understood as *precision tools*—powerful in supporting individualized practice and feedback, yet fundamentally reliant on human judgment to ensure meaningful, ethical, and communicative language learning.

7.3 Conversational Agents and Chatbots: Interactionist and Sociocultural Perspectives

The emergence of Generative AI (GenAI) has revolutionized conversational agents, moving them from simple command-response tools to sophisticated dialogue partners capable of human-like interaction. Chatbots like OpenAI's ChatGPT, Google's Gemini, and Microsoft's Copilot function as virtual interlocutors that provide a low-anxiety environment for language practice (Crompton & Burke, 2023; Okonkwo & Ade-Ibijola, 2021, as cited in Agyare et al., 2025).

Conversational agents and AI-powered chatbots have become one of the most visible and pedagogically debated manifestations of Artificial Intelligence in language learning. Enabled by advances in natural language processing (NLP) and large language models (LLMs), these systems simulate dialogic interaction, allowing learners to engage in written or spoken exchanges that approximate human conversation. Their rapid adoption in language education reflects a growing demand for scalable, low-anxiety, and readily accessible opportunities for communicative practice, particularly in

contexts where authentic interaction is limited (Faiz & Hazilan, 2024; Wang et al., 2025).

From a *Second Language Acquisition (SLA)* standpoint, conversational agents occupy a unique position because they directly target *interaction*, a construct that lies at the heart of both *interactionist* and *sociocultural theories* of language learning. Unlike intelligent tutoring systems, which primarily emphasize accuracy and form-focused practice, chatbots foreground dialogic engagement, negotiation of meaning, and learner participation in simulated communicative events.

7.3.1 Conversational Agents and Interactionist SLA Theory

Within *interactionist SLA theory* (Long, 1996), language acquisition is facilitated through meaning-focused interaction, particularly when learners encounter communication breakdowns that require negotiation of meaning. Conversational agents operationalize aspects of this theory by providing learners with opportunities to produce output, receive feedback, and adjust their language use in response to system-generated prompts.

Empirical research indicates that chatbot-mediated interaction can enhance learners' speaking fluency, lexical diversity, and willingness to communicate, especially among learners who experience anxiety in face-to-face interaction (Wang et al., 2025; Xu & Ismail, 2025). Chatbots enable repeated interaction without social risk, thereby increasing output quantity—a key variable in interactionist accounts of acquisition.

The strength and consistency of this effect have since been quantified directly. A meta-analysis pooling 70 effect sizes from 28 studies found that chatbot-assisted conversation practice, including practice mediated through ChatGPT, produces an overall positive effect on language-learning performance, though the size of that effect is moderated rather than uniform across studies (Wang et al., 2024). Individual controlled studies illustrate what that moderation looks like in practice. In a twelve-week study, 120 ESL undergraduates split evenly between an AI-chatbot condition and a control condition showed the chatbot group significantly outperforming controls on fluency, pronunciation, grammar, vocabulary, and speaking confidence (Grab, 2025). In a quasi-experimental comparison of 200 language learners against a conventional-instruction group, chatbot and virtual-assistant use improved vocabulary retention, fluency, and motivation

while reducing language anxiety (Jara Chiriboga et al., 2025). And in a pretest–posttest design with 203 EFL students, conversational agents that delivered adaptive rather than fixed feedback raised self-regulated-learning scores by 133 percent relative to baseline and outperformed a non-adaptive comparison group by %16 on measures of self-regulation and retention (Du, 2025).

A significant advantage of AI chatbots in language learning is their ability to reduce the affective filter. Learners often experience anxiety when speaking with human peers or instructors due to fear of negative evaluation. Chatbots offering a non-judgmental space where students can practice freely, which has been found to increase willingness to communicate and oral proficiency (Yuan, 2023, as cited in Tan et al., 2025). For example, studies involving tools like *Apilingua*, *Duolingo* and *Speeko* demonstrate that AI-driven conversation practice substantially enhances speaking skills by providing real-time corrections on pronunciation and grammar without the social pressure of a classroom (Qiao & Zhao, 2023; Shafiee Rad, 2024, as cited in Tan et al., 2025).

Beyond speaking, conversational agents serve as dynamic writing assistants. In English for Academic Purposes (EAP) contexts, students utilize tools like *Grammarly* and *Quillbot* alongside generative chatbots to brainstorm ideas, refine sentence structures, and expand vocabulary (Kohnke et al., 2025). GenAI tools can act as "Socratic" tutors, prompting students to think critically about their arguments rather than simply correcting errors (Dan et al., 2023, as cited in Li et al., 2025). However, the effectiveness of these agents depends on "*AI literacy*"—the ability of learners to critically evaluate AI outputs—as over-reliance can lead to diminished critical thinking or academic integrity issues (Kohnke et al., 2025).

However, from a theoretical perspective, chatbot interaction differs fundamentally from human interaction. While systems can simulate clarification requests, reformulations, or confirmation checks, they rarely engage in *genuine negotiation of meaning* driven by shared communicative intent. On this basis, interactionist scholars caution that chatbot interaction supports *interactional practice*, but only partially replicates the cognitive and social conditions under which negotiation-driven acquisition occurs.

Not every controlled comparison favors the chatbot, however. In a pre-post study of 125 EFL undergraduates (87 fully analyzed), an AI chatbot built

for dynamic assessment of speech acts did not outperform either classroom-based dynamic assessment or direct corrective feedback; classroom dynamic assessment in fact outperformed direct feedback on both immediate and delayed measures leaving the AI condition the weakest of the three (Saeedi & Soltani, 2025). Reviewers of the wider chatbot literature counsel similar caution against overgeneralizing from favorable studies: a systematic synthesis of chatbot-usefulness research evaluates the evidentiary quality behind claimed learning-outcome impacts rather than accepting them at face value (Şahin Kızıl et al., 2025), and a content analysis of dialogues from three AI language-learning apps—Duolingo, ImmerseMe, and MakesYouFluent—found that none of them exploited the natural-language-processing capabilities available to them for developing learners' emotional or intercultural competence (Sriganesh & Eamer, 2025).

7.3.2 Output, Feedback, and Modified Interaction

Conversational agents also intersect with *Swain's (1995) Output Hypothesis* which emphasizes the role of pushed output in promoting syntactic processing and noticing gaps in interlanguage. Chatbots encourage learners to articulate meanings, test hypotheses, and reformulate language when prompted by system feedback or follow-up questions.

Studies demonstrate that chatbot feedback—ranging from implicit reformulations to explicit corrections—can draw learners' attention to linguistic form while maintaining a communicative focus (Wu, 2024). From an interactionist lens, such feedback functions as *modified interaction*, supporting both fluency development and form–meaning mapping.

Nevertheless, automated feedback raises theoretical and pedagogical concerns. Chatbots may prioritize surface-level correctness or generate linguistically plausible but pragmatically inappropriate responses. Without human mediation, learners may struggle to distinguish between communicatively effective language and statistically probable language, highlighting the limits of interactionist alignment in fully automated systems.

7.3.3 Sociocultural SLA and Chatbots as Mediational Tools

From a *sociocultural SLA perspective*, language learning is understood as a socially mediated process, shaped by interaction, tools, and participation within culturally situated activity systems (Lantolf & Thorne, 2006). In this framework, conversational agents can be conceptualized as *mediational*

artifacts rather than interlocutors, supporting learners' participation in language use through scaffolded dialogue.

Chatbots function as symbolic tools that mediate cognition by prompting reflection, modeling discourse patterns, and sustaining interaction over time. Research suggests that chatbot-supported environments can promote learner confidence, identity experimentation, and exploratory language use—particularly for novice or marginalized learners who may feel inhibited in traditional classroom settings (Mohebbi, 2025).

However, sociocultural theory also foregrounds the importance of *intentionality, shared meaning, and social agency*, dimensions that chatbots cannot authentically reproduce. While systems may simulate politeness strategies or cultural norms, they lack lived experience and ethical accountability. As a result, sociocultural scholars emphasize that chatbots should be positioned as *supportive mediators within broader social learning ecologies,* not as replacements for human interaction.

The magnitude of these mediational benefits is not constant across age groups or instructional contexts. A longitudinal study following Swedish lower-secondary students using a spoken dialogue system with an embodied conversational agent in English lessons found sustained engagement, alongside only a slightly positive trend in students' social and emotional learning experience over time (Ericsson & Johansson, 2023). Set beside the double-digit gains reported above for university-age learners, this is a comparatively modest result, suggesting that chatbot mediation may pay larger dividends for older or more proficient learners than for younger adolescents. In a five-week quasi-experiment with 70 Chinese university students, social-network-based interaction with an AI speaking app improved English speaking skills more than non-interactive use of the same app, based on surveys, interviews, and pre/post-tests (Zou et al., 2023); the comparison suggests that how learners are structured to use a chatbot matters as much as whether they use one at all. Evidence on combining conversational AI with immersive technology is also emerging: in a four-month intervention pairing a chatbot with virtual-reality tasks, 320 Chinese EFL undergraduates showed significant pre-post gains in intercultural competence, particularly in cultural knowledge and communication skills (Ma & Shi, 2026).

7.3.4 Scaffolding, the Zone of Proximal Development, and Adaptivity

One of the most theoretically compelling applications of chatbots emerges in relation to *Vygotsky's Zone of Proximal Development (ZPD)*. Adaptive conversational agents can adjust linguistic complexity, prompt depth, and task demands based on learner responses, approximating contingent scaffolding.

When designed transparently and pedagogically, such adaptivity can support learners at transitional stages of development, particularly in vocabulary expansion, pragmatic routines, and discourse initiation. Systematic reviews indicate that learners benefit most when chatbot scaffolding is gradual, responsive, and paired with reflective tasks (Mohebbi, 2025).

Yet, ZPD-oriented scaffolding presupposes intentional guidance and withdrawal of support—processes that remain difficult to operationalize algorithmically. Without human oversight, chatbot scaffolding risks becoming static or misaligned with learners' evolving needs, underlining the necessity of teacher-guided integration.

7.3.5 Learner Agency, Autonomy, and Ethical Considerations

Conversational agents are frequently associated with enhanced learner autonomy due to their availability, responsiveness, and non-judgmental interaction style. Learners can initiate conversations, control topics, and practice at their own pace, aligning with autonomy-oriented SLA perspectives.

At the same time, scholars warn against *over-reliance and epistemic authority displacement*, where learners treat chatbot output as linguistically or culturally authoritative. Ethical concerns related to data privacy, bias, and hallucinated content further complicate their pedagogical deployment (Yang et al., 2025).

The youngest learners raise this concern most acutely, and the evidence base here is only beginning to catch up. A randomized controlled trial directly compared parent-led dialogic reading with LLM-powered conversational-agent dialogic reading using children's e-books, testing effects on English-as-a-foreign-language comprehension and conversational skill (Xiao et al., 2025)—a design that, unusually for this literature, treats the human interlocutor as the standard against which the AI is measured, rather than the reverse. Designs of this kind matter because younger learners are

least able to judge for themselves when a conversational agent's confident tone outruns its actual reliability, reinforcing the broader point that chatbot mediation needs framing calibrated to learner age rather than a single deployment model applied uniformly from childhood through university.

Accordingly, current research converges on *human-in-the-loop pedagogies*, where chatbots serve as interactional practice tools embedded within teacher-curated tasks, peer discussion, and critical reflection. In such models, educators frame chatbot interaction as exploratory, provisional, and subject to evaluation—thereby preserving learner agency and critical language awareness.

Conversational agents and chatbots occupy a theoretically rich yet pedagogically complex position within AI-supported language learning environments. Anchored in interactionist and sociocultural SLA frameworks, these tools offer valuable opportunities for increased output, reduced anxiety, and scaffolded interaction. However, their effectiveness depends on careful pedagogical framing, transparency of function, and sustained human mediation.

Rather than functioning as conversational substitutes, chatbots are best understood as *mediational instruments*—tools that expand access to practice and reflection while remaining embedded within socially grounded, ethically guided language learning ecologies.

7.4 Mobile and Immersive AI Environments (VR/AR): Embodied and Sociocultural Dimensions of Language Learning

Mobile-assisted language learning (MALL) and immersive technologies—particularly Virtual Reality (VR) and Augmented Reality (AR)—represent a significant expansion of AI-supported language learning environments beyond traditional screen-based interaction. The convergence of AI with Virtual Reality (VR) and Augmented Reality (AR) has given rise to immersive learning environments, often conceptualized within the "*Metaverse*." These environments offer situated learning experiences that are difficult to replicate in a traditional classroom. These technologies enable learners to engage with language through movement, spatial orientation, and contextualized action, thereby shifting language learning from abstract symbol manipulation toward *situated, embodied experience*. When combined with

AI-driven adaptivity and analytics, mobile and immersive environments offer new affordances for contextualized input, experiential interaction, and socially meaningful language use (Mohebbi, 2025; Zhang et al., 2025).

From a *Second Language Acquisition (SLA)* perspective, the pedagogical relevance of mobile and immersive AI environments lies in their alignment with both *embodied cognition* and *sociocultural theories of learning*, which conceptualize language acquisition as fundamentally grounded in action, perception, and social participation rather than decontextualized mental procedure.

Recent systematic and bibliometric reviews identify VR/AR-enhanced environments as an emerging but rapidly growing cluster within AI-mediated language learning research, particularly in relation to speaking, pronunciation, and learner engagement (Wang et al., 2025; Yang et al., 2025). These environments are increasingly framed not simply as technological enhancements, but as *pedagogical spaces* in which cognition, interaction, and social participation are intersecting.

AI-enhanced VR systems create embodied learning experiences where learners can interact with AI-driven avatars in realistic scenarios. For instance, the *Interview Training and Education Module (ITEM)* utilizes VR and LLMs to simulate job interviews, providing students with a realistic setting to practice professional communication skills while receiving personalized feedback from an AI scorer (Nofal et al., 2025). Such multimodal applications allow for the assessment of not just linguistic accuracy, but also paralinguistic features like tone and pacing.

How such AI-driven avatars are positioned within the virtual space is not a cosmetic detail. A between-subjects experiment with 64 students found that whether a pedagogical agent appeared inside or outside the video frame in a desktop virtual-reality task measurably affected gains in English academic presentation proficiency (Li et al., 2024).The finding is suggestive rather than settled in any case—a sample of 64 is enough to detect a design effect, not to establish which positioning generalizes across tasks or proficiency levels. Comparable caution applies to mixed-reality group conversation: a system pairing multiple AI agents with human learners in a shared mixed-reality space increased engagement, willingness to communicate, and eased speaking-related fear among 21 second-language learners (Fernandez-Espinosa et al., 2025)—a promising result, but one drawn from a single small trial rather than a body of replicated findings.

In language education, VR can transport students to context-rich environments—such as a virtual café or a foreign city—where they must navigate social interactions in the target language (Can, 2020). Research indicates that students using Edu-Metaverse models outperform peers in traditional settings across writing, reading, and speaking domains due to the high level of engagement and "deep learning" facilitated by the immersive context (Shu & Gu, 2023, as cited in Tan et al., 2025). Additionally, systems designed for younger children, such as *Kid Space*, blend physical and digital play to support collaborative problem-solving, demonstrating how AI can mediate social interaction in physical spaces through projected virtual environments (Aslan et al., 2025).

7.4.1 Embodied Cognition and Language Learning

Embodied cognition theories challenge traditional views of language learning as a purely cognitive process, arguing instead that meaning emerges from bodily interaction with the physical and social world. From this perspective, linguistic knowledge is not stored independently of sensory and motor systems, but is shaped through *sensorimotor experience, spatial awareness, and affective engagement.*

Immersive AI environments operationalize embodied cognition by situating learners within simulated contexts where language is tied to action—such as navigating a virtual city, ordering food in a simulated café, or interacting with virtual objects and characters. Research indicates that such environments enhance vocabulary retention, pragmatic awareness, and discourse comprehension by anchoring linguistic forms to perceptual and motor experiences (Mohebbi, 2025).

Unlike traditional multimedia input, VR and AR environments allow learners to *do things with language* rather than merely observe or respond. From an SLA standpoint, this supports deeper form–meaning mappings and reduces cognitive load by distributing processing across perceptual channels. Embodied engagement thus functions as a cognitive scaffold, reinforcing learning through multimodal integration.

Empirical evidence synthesized in recent reviews indicates that VR-supported language learning enhances vocabulary retention, pronunciation accuracy, and pragmatic awareness by anchoring linguistic forms in perceptual and motor experience (Mohebbi, 2025). Similarly, AI-assisted

immersive scenarios reduce cognitive overload by distributing processing across visual, auditory, and kinesthetic channels, supporting sustained engagement and deeper learning (Xu & Ismail, 2025).

7.4.2 Situated Learning and Contextualized Input

A key contribution of immersive environments is their capacity to provide *situated input*, a concept central to both embodied and sociocultural learning theories. In immersive settings, language input is embedded within meaningful contexts that mirror real-world communicative situations. This contrasts sharply with decontextualized textbook dialogues or isolated practice items. From a sociocognitive and interactionist SLA perspective, immersive environments are particularly valuable because they provide *situated and task-relevant input.* Unlike traditional CALL materials, VR/AR systems embed language use within meaningful scenarios—such as service encounters, navigation tasks, or collaborative problem-solving—thus approximating real-world communicative conditions (Wang et al., 2025).

AI-enhanced VR systems can dynamically adjust linguistic complexity, task demands, and interactional prompts based on learner behavior, enabling *context-sensitive scaffolding.* For example, learners may receive adaptive prompts, visual cues, or linguistic support when navigating unfamiliar situations, approximating authentic communicative pressure while maintaining pedagogical safety. Bibliometric analyzes highlight this adaptive contextualization as a defining feature distinguishing immersive AI environments from earlier multimedia approaches (Kartal & Yeşilyurt, 2024; Yang et al., 2025).

From an SLA perspective, such environments promote *comprehensible input* that is both meaningful and task-relevant, supporting acquisition through engagement rather than rote exposure. Importantly, this form of contextualization also enhances pragmatic competence, an area often neglected in form-focused AI tools.

7.4.3 Sociocultural SLA: Participation, Mediation, and Social Presence

From a *sociocultural SLA* perspective, language learning is fundamentally a process of *participation in socially organized activity*, mediated by tools, symbols, and interaction with others. Mobile and immersive AI environments can be conceptualized as *mediational spaces* that structure learners'

participation in language use through guided action and simulated social interaction.

In multi-user VR environments or AI-driven role-play scenarios, learners engage in collaborative tasks, negotiate roles, and co-construct meaning. Even in single-user immersive systems, virtual agents and environments function as symbolic others, shaping learners' discourse choices and interactional strategies. These experiences align with sociocultural notions of learning as *legitimate peripheral participation*, where learners gradually move toward fuller communicative competence through supported engagement.

However, sociocultural theory also foregrounds the importance of *authentic social relations, cultural norms, and ethical accountability*, dimensions that immersive AI environments can only partially approximate. While virtual agents may simulate politeness strategies or cultural scripts, they lack lived cultural experience. Seen in this light, immersive learning environments are most effective when integrated with reflective discussion, peer interaction, and teacher mediation that contextualize virtual experiences within real sociocultural frameworks. Scholars also caution that simulated social interaction cannot fully replicate authentic cultural participation. As emphasized in bibliometric mappings, immersive AI environments require *human mediation* to contextualize cultural norms, interpret pragmatic appropriateness, and connect virtual experiences to real-world communicative practices (Zhang et al., 2025).

Research reviewed by Mohebbi (2025) demonstrates that immersive AI tools promote learner confidence, identity experimentation, and willingness to communicate—particularly when learners interact with virtual agents or participate in role-based simulations. These environments support gradual movement from peripheral participation toward more autonomous language use, consistent with sociocultural models of development.

7.4.4 The Zone of Proximal Development and AI-Driven Scaffolding

Immersive AI environments also intersect meaningfully with *Vygotsky's Zone of Proximal Development (ZPD).* Adaptive VR/AR systems can modulate task complexity, linguistic input, and interactional support in response to learner performance, thereby approximating contingent scaffolding.

For instance, an AI system may simplify discourse, provide visual supports, or introduce prompts when learners struggle, and gradually withdraw assistance as competence increases. Research suggests that such adaptive scaffolding supports confidence, persistence, and risk-taking—key affective conditions for language development (Mohebbi, 2025).

Nevertheless, ZPD-based scaffolding presupposes intentional guidance and sensitivity to learners' evolving needs. Algorithmic scaffolding, while responsive, lacks pedagogical intentionality unless explicitly designed and supervised. As such, immersive AI environments require *teacher orchestration* to ensure that scaffolding aligns with curricular goals and learners' developmental trajectories.

Systematic reviews report that such adaptive scaffolding enhances learner persistence, reduces anxiety, and supports gradual skill development—particularly in speaking and listening tasks (Mohebbi, 2025; Wang et al., 2025). However, research also highlights that algorithmic scaffolding must remain pedagogically transparent and teacher-supervised to avoid misalignment with learners' developmental needs.

7.4.5 Mobility, Informality, and Language Use Beyond the Classroom

Mobile AI environments extend immersive learning beyond fixed instructional spaces, enabling *context-aware, informal language learning* (Liu & Zhao, 2025), supporting "anytime-anywhere" accessibility. AI-assisted audio-learning modules, for example, allow students to engage with content while multitasking, catering to diverse learning needs including those of neurodiverse students (Jafarian & Kramer, 2025). These mobile platforms often integrate adaptive algorithms to push content that matches the learner's immediate context and proficiency level. Smartphones equipped with AI-driven AR applications can overlay linguistic information onto real-world environments, transforming everyday spaces into learning opportunities.

From a *sociocultural perspective*, this blurring of formal and informal learning spaces is pedagogically significant. Language learning becomes embedded in daily practices—navigation, shopping, social interaction—supporting identity formation and learner agency. However, researchers caution that without reflective framing, mobile immersion risks fragmenting learning into episodic experiences rather than sustained development (Yang et al., 2025).

7.4.6 Pedagogical and Ethical Considerations

Despite their promise, mobile and immersive AI environments raise practical and ethical concerns. *High costs, accessibility barriers*, and *uneven teacher readiness* may exacerbate educational inequities. Moreover, immersive realism can obscure the pedagogical framing of activities, leading learners to conflate simulation with authentic cultural competence (Zhang et al., 2025).

The evidence base for these ethical concerns is in fact more developed than the evidence base for the pedagogical claims that motivate immersive adoption in the first place. A PRISMA-based synthesis of 63 higher-education studies found that human-centered AI, blended learning, augmented-reality, and gamification approaches outperformed technology-centric implementations, though digital-divide access barriers and uneven faculty readiness persisted (Mahyiyuddin & Erlita, 2026). Read against this section's claims about embodied cognition and situated input, the picture is genuinely mixed: the theoretical case for VR and AR in language learning is more fully worked out than the empirical case for it, and what empirical case exists is small in scale, concentrated in a handful of national contexts, and consistently qualified by access and readiness constraints rather than confirmed as a general effect.

Ethically, immersive environments intensify data collection through tracking movement, gaze, and interaction patterns, necessitating transparent governance and informed consent. These considerations reinforce the need for *human-centered design and oversight*, ensuring that immersive technologies serve pedagogical and ethical objectives rather than technological novelty.

Mobile and immersive AI environments represent a powerful evolution in language learning, aligning closely with embodied cognition and sociocultural SLA frameworks. By situating language within action, space, and social participation, these environments offer rich opportunities for contextualized, meaningful learning. However, their effectiveness depends on intentional pedagogical integration, reflective mediation, and ethical stewardship.

Rather than functioning as autonomous learning spaces, immersive AI environments are best understood as *experiential extensions of socially grounded pedagogy*, where embodiment, interaction, and human guidance converge to support sustainable language development.

7.5 Adaptive and Personalized Learning Pathways

The ultimate promise of AI in education lies in its ability to offer *adaptive* and *personalized learning pathways*. By analyzing vast amounts of learner data, AI systems can tailor the sequence, difficulty, and type of content to individual needs, a process often referred to as adaptive learning. Also, adaptive and personalized learning pathways represent the integrative core of AI-supported language learning environments. Unlike discrete tools (e.g., chatbots or ITS) that target specific skills, adaptive systems operate at the *curricular and longitudinal level*, continuously adjusting content, pacing, feedback, and task sequencing based on learner data. These systems synthesize inputs from performance metrics, interaction histories, and engagement indicators to generate individualized learning trajectories that evolve over time (Kartal & Yeşilyurt, 2024; Yang et al., 2025).

Adaptive platforms like *Knewton* and *Smart Sparrow* analyze student performance in real-time to adjust learning trajectories (Cukurova et al., 2023, as cited in Agyare et al., 2025). If a student struggles with a specific grammatical concept, the system can automatically reroute them to remedial exercises or alternative explanations before allowing progression. This aligns with differentiated instruction theories, ensuring inclusivity by preventing learners from falling behind (Agyare et al., 2025).

From a *Second Language Acquisition (SLA)* standpoint, adaptive learning pathways resonate strongly with *usage-based theories of language learning*, which conceptualize acquisition as an emergent process driven by frequency, distribution, and meaningful use. Simultaneously, from an educational psychology perspective, adaptive pathways intersect with *self-regulated learning (SRL)* frameworks, emphasizing learner agency, metacognitive control, and sustained engagement. The convergence of these perspectives positions adaptive AI systems as both linguistic and motivational infrastructures for language development.

Meta-analytical evidence confirms that personalization significantly enhances learning outcomes, particularly when combined with formative feedback and learner reflection (Wu, 2024). Adaptive systems are especially effective in listening and speaking development, where individualized pacing and repeated exposure are critical (Xu & Ismail, 2025).

However, scholars emphasize that personalization must not be conflated with pedagogical neutrality. Algorithmic decision-making reflects

embedded assumptions about language, learning, and proficiency. Ethical concerns related to data privacy, bias, and learner surveillance highlight the need for transparent design and institutional oversight (Edmett et al., 2024; Yang et al., 2025).

7.5.1 Usage-Based SLA: Frequency, Input Optimization, and Emergent Learning

Usage-based SLA models argue that language knowledge emerges from repeated exposure to linguistic constructions in meaningful contexts, with frequency and variability shaping entrenchment and generalization. From this perspective, acquisition is not the internalization of abstract rules but the gradual strengthening of form–meaning mappings through use.

Adaptive AI systems operationalize usage-based principles by *tracking learner exposure, usage patterns, and error distributions*, thereby optimizing input frequency and sequencing. For example, systems may recycle lexical bundles, morphosyntactic constructions, or discourse routines that learners underuse or misuse, while gradually introducing variability to promote generalization. Bibliometric analyzes identify such adaptive recycling and sequencing as a defining feature of contemporary AI-mediated language learning environments (Kartal & Yeşilyurt, 2024; Zhang et al., 2025).

Significantly, usage-based theory emphasizes *meaningful use over isolated practice*. Adaptive pathways that merely increase exposure without communicative relevance risk reinforcing surface familiarity rather than functional competence. As a result, effective adaptive systems integrate frequency optimization with contextualized tasks, aligning usage-based learning with communicative goals.

7.5.2 Input, Output, and Adaptive Task Sequencing

Usage-based SLA also foregrounds the role of *output* as a mechanism for consolidating constructions and testing hypotheses about language use. Adaptive learning pathways support this process by adjusting task demands to encourage repeated, varied production across modalities (speaking, writing, interaction).

Systematic reviews indicate that adaptive systems are particularly effective in speaking and listening development when they balance receptive exposure with productive tasks and feedback (Wang et al., 2025; Xu & Ismail,

2025). By analyzing learner output, AI systems can detect stagnation, overgeneralization, or avoidance patterns and respond with targeted prompts or task redesign.

However, scholars caution that algorithmic adaptivity must avoid *over-optimization*, where learners are confined to narrow linguistic patterns that maximize short-term accuracy at the expense of exploratory language use. Usage-based theory underlines the importance of variability and risk-taking, suggesting that adaptive pathways should deliberately include opportunities for productive struggle and creative recombination.

7.5.3 Self-Regulated Learning and Learner Agency

From a *self-regulated learning (SRL) perspective*, adaptive and personalized learning pathways influence not only what learners learn, but *how they manage their learning processes*. SRL frameworks conceptualize learning as a cyclical process involving goal setting, strategic action, monitoring, and reflection.

AI-driven Competency Recommendation (CR) systems utilize machine learning to recommend specific competencies and content based on students' academic performance and behavioral patterns, such as attendance (Junaidi & Wahyono, 2025). What is more, these pathways support Self-Regulated Learning (SRL). By visualizing progress and providing personalized nudges, AI tools help learners set goals, monitor their understanding, and adjust their strategies (Zimmerman, 1990, as cited in Heung & Chiu, 2025). For example, generative AI can create differentiated learning materials—such as simplifying a text for a novice reader or generating complex critical thinking questions for an advanced learner—thereby maintaining optimal challenge levels for all students (Pesovski et al., 2025).

A major contribution of AI-driven adaptivity is its capacity to support SRL by making learning processes visible. Dashboards, progress indicators, and adaptive recommendations can enhance learners' metacognitive awareness, enabling them to plan, monitor, and evaluate their learning more effectively. Mohebbi's (2025) systematic review provides strong evidence that AI-supported personalization positively affects learner autonomy, engagement, and persistence, particularly when learners retain some control over learning decisions.

Nevertheless, SRL research also warns that excessive system control may undermine agency. When adaptive pathways operate opaquely—making decisions without explanation—learners may become passive recipients rather than active regulators of their learning. Therefore, personalization must be *negotiable and transparent*, allowing learners to understand, question, and override system recommendations when appropriate.

7.5.4 Motivation, Affect, and Persistence

Adaptive learning pathways also intersect with affective dimensions of language learning. By adjusting task difficulty and feedback timing, AI systems can help maintain learners within optimal challenge zones, reducing anxiety and disengagement. Reviews in Asian EFL contexts report that adaptive AI environments contribute to increased motivation and willingness to communicate, particularly among lower-proficiency learners (Wang et al., 2025).

However, motivation is not solely a function of difficulty calibration. Usage-based and sociocultural perspectives both emphasize *purposeful participation* and identity investment. Adaptive systems that prioritize efficiency over meaningful engagement risk reducing learning to performance optimization. As such, adaptive pathways should be embedded within broader pedagogical designs that emphasize goals, relevance, and social connection.

7.5.5 Ethical and Pedagogical Constraints of Personalization

Despite their promise, adaptive learning pathways raise significant ethical and pedagogical concerns. *Algorithmic personalization* relies heavily on learner data, including performance histories and behavioral traces, intensifying issues of privacy, consent, and bias (Yang et al., 2025). At the same time, bibliometric analyzes highlight the risk of *normative modeling*, where learners are measured against idealized trajectories that may marginalize diverse learning styles or linguistic backgrounds (Zhang et al., 2025).

Empirical work on who actually benefits from personalization substantiates this concern rather than merely theorizing it. A social-network analysis of 50 Chinese university students across AI-assisted L2 courses found a persistently dual-centric structure in which teachers remained the authoritative node and AI occupied a passive, advisory position, with a learner's own AI literacy moderating whether greater interaction with the system translated

into higher achievement and enjoyment (Xun & Wang, 2026). Personalization, on this evidence, does not flatten the playing field so much as reproduce it: learners already equipped to interrogate and direct an AI system extract more from it, while those who are not are left with a system that behaves, functionally, the same for everyone—a pattern Chapter 12 examines further in relation to digital equity. A comparable pattern emerged among 98 university students grouped by AI self-efficacy while using a knowledge-graph-integrated generative AI feedback agent: high-self-efficacy learners showed greater self-monitoring, lower cognitive load, and higher self-evaluation on complex tasks than low-self-efficacy peers using the identical system (Tu et al., 2026).

From an *SLA perspective*, such normalization conflicts with usage-based views of language development as nonlinear and individualized. In turn, scholars advocate for *human-centered adaptive design*, where teachers interpret system outputs, contextualize recommendations, and ensure that personalization supports inclusion rather than standardization.

Reliance patterns show the same asymmetry. In collaborative problem-solving mediated by a ChatGPT-based AI assistant, students grouped by prior-knowledge level did not differ significantly in final performance, but learners with lower prior knowledge interacted with the assistant far more frequently and depended more heavily on its guidance (Tang et al., 2026)—precisely the profile of learner for whom over-reliance carries the greatest long-term cost to independent competence. Findings of this kind sharpen, rather than resolve, the case for human-centered adaptive design already argued above. A framework for ethical, inclusive, and pedagogically driven AI-assisted language education—spanning intelligent tutoring, NLP-based chatbots, and AI-generated content—was validated against the judgment of 48 practicing language educators, who expressed moderate-to-strong agreement across all five of its proposed dimensions (Paul et al., 2026). The result suggests that teachers themselves are well positioned to specify what responsible personalization should look like, provided they are asked.

Adaptive and personalized learning pathways represent the culmination of AI-supported language learning environments, integrating usage-based SLA principles with self-regulated learning frameworks. By optimizing input frequency, sequencing output opportunities, and supporting metacognitive

control, these systems offer powerful mechanisms for individualized language development.

Yet, their effectiveness depends on transparency, learner agency, and pedagogical oversight. Rather than functioning as autonomous curriculum engines, adaptive pathways are most educationally valuable when embedded within *human-guided, theoretically grounded learning ecosystems* that respect the emergent, variable, and socially situated nature of language acquisition.

7.6 Conclusion

AI-supported language learning environments represent a fundamental shift in how languages are learned, practiced, and experienced. Intelligent tutoring systems, conversational agents, immersive technologies, and adaptive pathways collectively enable personalized, interactive, and scalable learning opportunities. Yet, research consistently demonstrates that *AI's educational value is contingent upon human pedagogical judgment, ethical governance, and thoughtful instructional design*.

Rather than replacing teachers, AI-supported environments reconfigure their roles—as designers of learning experiences, mediators of meaning, and ethical stewards of technology-enhanced education. The future of AI in language learning therefore lies not in automation, but in *collaborative human–AI ecosystems* that balance technological innovation with pedagogical responsibility.

Pedagogical Companion for Chapter 7

Key Concepts: intelligent tutoring systems; noticing and feedback; conversational agents; mediational tools; scaffolding and adaptivity; embodied cognition; situated learning; VR/AR environments; adaptive pathways; personalization.

Discussion Questions

1. ITS research predates LLMs by decades (see Graesser et al., 2005). Which ITS design principles survive the transition to LLM-based tutors, and which are obsolete?

2. Evaluate a chatbot conversation you have had against the interactionist criteria of Section 7.3: was there genuine negotiation of meaning, or its simulation?
3. What does embodied cognition predict about the difference between learning a language through text chat versus through an avatar in a virtual environment?
4. Adaptive pathways optimize the sequence of input. What aspects of language learning resist sequencing altogether?
5. Design the "handover points" for a tutoring system in your context: at which moments must the system route the learner to a human teacher?

Further Reading

- Graesser, A. C., Chipman, P., Haynes, B. C., & Olney, A. (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education, 48*(4).
- Okonkwo, C. W., & Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review. *CAEAI, 2*.
- Liu, G., & Zhao, X. (2025). A scoping review of AI-mediated informal language learning. *ReCALL, 38*(1).

Chapter 8. Learner Autonomy, Self-Regulation, and Agency in AI-Mediated Language Education

8.1 Introduction: Autonomy at a Technological Turning Point

The integration of Artificial Intelligence (AI) into language education has fundamentally redefined the relationship between the learner and the learning process, shifting the locus of control from a teacher-directed model toward a dynamic, interactive ecosystem. In this new landscape, AI acts as a "cognitive partner," empowering students to take charge of their acquisition process by establishing clear objectives and managing their own progress. This shift is characterized by *distributed agency*, where technological tools and human learners collaboratively construct knowledge. By providing individualized paths and real-time support, AI supports *learner autonomy*, which is increasingly recognized as a fundamental prerequisite for successful technology-enhanced language acquisition. Thus, this chapter explores how AI facilitates self-regulation and motivation while critically examining the systemic risks of intellectual passivity and dependency.

Unlike traditional computer-assisted language learning (CALL), which relied on pre-programmed content, contemporary generative AI (GenAI) and Intelligent Tutoring Systems (ITS) facilitate iterative interactions that empower students to co-regulate their linguistic development. While these technologies offer unprecedented opportunities to encourage learner agency and Self-Regulated Learning (SRL), their ubiquitous presence introduces significant paradoxes regarding cognitive engagement and potential dependency on automation. This chapter examines the interplay between AI and the psychological dimensions of learning, focusing on feedback mechanisms, self-directed pathways, affective factors, and the critical risks associated with the automation of language acquisition. While earlier educational technologies primarily enhanced access and efficiency, contemporary AI tools directly intervene in the cognitive, motivational, and metacognitive processes that underlie language acquisition.

Within this evolving ecology, *learner autonomy, self-regulation,* and *agency* have re-emerged as central theoretical and pedagogical concerns. Empirical studies document improvements in vocabulary retention (Ngo, 2024), pronunciation accuracy (Xu & Ismail, 2025), speaking fluency (Huang et al., 2025), and engagement (Mohebbi, 2025). Bibliometric

analyzes further reveal that "learner autonomy," "engagement," and "AI-supported instruction" are now dominant thematic clusters in AI-language research (Yang et al., 2025; Zhang et al., 2025).

These individual findings converge in aggregate. A multilevel meta-analysis pooling 117 effect sizes from 46 EFL studies published between 2022 and 2025 found a medium-to-large overall effect of AI on language learning with the strongest effects for vocabulary acquisition and the weakest for speaking (Torres & Kahveci, 2025)—a pattern consistent with the skill-specific findings above and a caution against treating AI's benefits as uniform across language domains.

Yet, this apparent empowerment coexists with a growing body of cautionary scholarship warning of over-reliance, passive consumption of AI outputs, diminished critical thinking, and erosion of human-mediated interaction (Xu & Ismail, 2025; Zhang et al., 2025). AI simultaneously promises autonomy and threatens it.

Drawing on systematic reviews (Mohebbi, 2025), empirical investigations (Huang et al., 2025; Xu & Ismail, 2025; Ngo, 2024), bibliometric mappings (Yang et al., 2025; Zhang et al., 2025), and foundational theories (Zimmerman, 2002; Deci & Ryan, 2000; Bandura, 1997; Vygotsky, 1978; Sweller, 1988; Benson, 2013), this chapter develops an integrated framework for understanding AI-mediated autonomy.

Rather than framing AI as either emancipatory or detrimental, the chapter argues that autonomy in AI environments is *contingent, scaffolded, and socially distributed*. AI can amplify agency—but only when embedded within reflective, ethically governed pedagogical structures.

8.2 AI Feedback and Metacognition

A central contribution of AI to language learning is its ability to serve as a *virtual "more capable peer"* that stimulates metacognitive awareness (Zhang et al., 2025). Metacognition, defined as the capacity to monitor, evaluate, and regulate one's own cognitive processes, is a cornerstone of successful language acquisition (Yan et al., 2025). Effective learners do not just produce language; they reflect on how they produce it. They anticipate difficulties, monitor comprehension, evaluate outcomes, and adjust strategies. In this sense, metacognition underlies both autonomy and agency.

Recent advancements in Artificial Intelligence (AI) have considerably transformed the architecture of feedback in language learning environments. Unlike static instructional materials or delayed teacher commentary, AI-driven systems provide *adaptive, response-contingent scaffolding* that dynamically interacts with learner cognition. These systems are capable of guiding learners through the phases of *self-regulation—forethought, performance,* and *reflection*—in real time (Heung & Chiu, 2025). AI feedback, therefore, does not simply correct; it structures cognitive awareness.

Research indicates that *adaptive feedback* generated by LLMs, such as ChatGPT, can markedly improve the quality of student output, particularly in complex tasks like diagnostic reasoning and argumentation (Kinder et al., 2025). By providing specific justifications and suggestions for improvement, rather than merely correcting errors, AI tools encourage learners to engage in deeper processing of information. For instance, in writing tasks, AI-generated feedback has been found to be more effective than static expert feedback in improving the quality of justifications, as it reduces the cognitive load required to bridge the gap between current performance and desired outcomes (Kinder et al., 2025).

Furthermore, the interactive nature of Generative AI (GenAI) shifts the pedagogical dynamic from passive reception to active inquiry. A key affordance of GenAI is its ability to serve as a partner in "question asking," a critical help-seeking behavior. Studies utilizing quantitative ethnographic methods have shown that learners engaging with GenAI tend to ask more pragmatic, direct questions regarding conceptual and procedural knowledge compared to those interacting with human tutors, who often elicit more indirect requests for feedback (Cheng et al., 2025). This suggests that AI can create a *safe epistemic space* where learners are less hesitant to admit knowledge deficits, thereby facilitating immediate remedial action and metacognitive repair (Cheng et al., 2025). However, the effectiveness of this feedback loop is contingent upon the learner's ability to formulate meaningful prompts, placing a premium on "*prompt engineering*" as a new metacognitive skill required for effective agency (Garg & Sood, 2025).

The *DRIVE (Directive Reasoning Interaction and Visible Expertise) framework* provides a practice-oriented model for assessing how students exercise agency through AI. According to Oliveira et al. (2025), learners who

exhibit high "*Directive Reasoning Interaction*" (DRI) do more than merely accept AI outputs but actively steer the dialogue, critically questioning suggestions and integrating their own domain knowledge (*Visible Expertise*). This active steering functions as a *cognitive safeguard* against "*automation bias*"—the tendency to uncritically follow machine-generated advice (Oliveira et al., 2025).

This safeguard is not automatically available to every learner; it appears to depend on prior AI literacy. A social network analysis of 50 Chinese university students across AI-assisted L2 courses found a structurally dual-centric pattern, with teachers occupying the authoritative position and AI functioning as a passive advisor rather than a co-equal interlocutor, and found that AI literacy moderated the relationship between a student's centrality in this interaction network and their achievement and enjoyment (Xun & Wang, 2026). Directive Reasoning Interaction, in other words, is conditioned by a literacy that itself requires cultivation, not a disposition learners either have or lack.

8.2.1 From Corrective Feedback to Metacognitive Scaffolding

Traditional corrective feedback in language classrooms has often been delayed, selective, and teacher-dependent. Written feedback typically arrives after task completion, and oral correction may be constrained by classroom time and distribution of attention. As a result, learners may struggle to connect correction to the cognitive processes that produced the error.

AI systems fundamentally alter this temporal and epistemic structure. Speech-recognition engines flag pronunciation deviations instantly (Xu & Ismail, 2025). Writing assistants highlight syntactic and lexical inconsistencies in real time. Generative AI platforms provide reformulations, stylistic variations, and contextual expansions interactively.

In vocabulary learning contexts, ChatGPT-mediated input has been shown to reduce cognitive load and facilitate retention when aligned with Krashen's *Input Hypothesis* and *Cognitive Load Theory* (Ngo, 2024). By organizing explanations and examples responsively, AI reduces extraneous cognitive load and strengthens germane processing. Yet the deeper contribution of AI lies not in efficiency but in cognitive scaffolding.

Research indicates that adaptive feedback generated by large language models (LLMs) can substantially improve the quality of student output in

complex tasks such as diagnostic reasoning and argumentative writing (Kinder et al., 2025). Unlike static expert feedback that merely identifies deficiencies, AI-generated feedback often includes justifications, elaborations, and improvement pathways. This explanatory dimension reduces the cognitive effort required to bridge the gap between current performance and desired standards (Kinder et al., 2025).

The magnitude of such gains is measurable directly against non-adaptive alternatives. In a pretest–posttest design with 203 EFL students, Du (2025) found that AI conversational agents providing adaptive feedback raised self-regulated-learning scores by 133 percent from baseline and outperformed a non-adaptive comparison group by 16% on combined measures of self-regulation and retention. The scale of this difference suggests that the metacognitive scaffolding described here depends on responsiveness to the individual learner's state, not on automation as such.

In writing tasks, for instance, AI feedback has been shown to improve the quality of justificatory reasoning more effectively than static commentary because it externalizes intermediate reasoning steps, thereby supporting metacognitive restructuring.

For example, from a *Self-Regulated Learning perspective* (Zimmerman, 2002), such adaptive scaffolding strengthens all three regulatory phases: *Forethought, Performance, and Self-reflection.*

8.2.1.1 AI and the Forethought Phase

AI platforms generate personalized study plans, diagnostic feedback, and CEFR-aligned progression maps. These features may assist learners in setting realistic goals. However, autonomy requires volitional ownership of goals. If learners merely follow AI-generated pathways without reflection, regulation becomes system-directed rather than self-directed. Effective AI integration must therefore allow learners to negotiate goals rather than accept them passively.

8.2.1.2 AI and the Performance Phase

Real-time feedback in speech recognition tools (Xu & Ismail, 2025) and conversational AI platforms (Huang et al., 2025) enhances self-monitoring. Writing assistants enable recursive drafting and revision (Mohebbi, 2025).

However, excessive reliance on AI detection mechanisms may weaken internal monitoring. Pedagogically, learners should be encouraged to anticipate corrections before revealing AI feedback, preserving metacognitive engagement.

8.2.1.3 AI and the Reflection Phase

Revision histories and analytics dashboards support reflective evaluation. Yet analytics alone do not ensure reflection. Structured prompts, comparison tasks, and reflective journals are necessary to activate SRL cycles. AI can scaffold SRL—but cannot replace it.

Mohebbi's (2025) systematic review confirms that AI tools significantly promote self-regulated learning by enabling recursive feedback loops that connect planning, monitoring, and evaluation. Automated writing systems and conversational agents further enable recursive drafting cycles that allow learners to compare successive versions of texts, reflect on lexical precision and syntactic complexity, evaluate stylistic tone shifts, monitor progress longitudinally. Revision history becomes a visible trace of cognitive development. The learner's thinking is externalized, making metacognitive growth observable.

However, immediacy does not guarantee depth. Without structured engagement, AI feedback may devolve into surface-level editing rather than conceptual restructuring (Zhang et al., 2025). The shift from correction to metacognition depends on how learners interact with the feedback.

8.2.2 AI as a Metacognitive Mirror

Metacognition involves awareness of one's own cognitive processes—thinking about thinking, monitoring comprehension, evaluating strategy use, and regulating effort. In language learning, metacognition manifests when learners recognize patterns in their errors, anticipate difficulties, and adjust strategies accordingly.

AI systems, when used dialogically rather than passively, function as what may be termed a *metacognitive mirror*. They reflect linguistic production back to the learner in augmented form, enabling cognitive comparison.

Instead of merely accepting correction, learners can ask: "*Why is this incorrect?", "What rule governs this structure?", "What alternative structures could convey this meaning?", "How can I adjust tone or register?", "What pattern do you see in my recurring errors?".* This kind of questioning shifts the learner from feedback recipient to *epistemic agent*.

Xu and Ismail (2025) note that AI-driven speech recognition tools enhance pronunciation awareness. Yet their findings suggest that effectiveness increases considerably when learners actively compare AI-generated corrections with their own recorded output. The cognitive benefit lies not in the correction itself but in the *contrastive awareness* it produces.

This shift—from receiving correction to interrogating correction—marks a transition from feedback consumption to metacognitive engagement. The learner is not merely correcting an utterance but examining the cognitive processes that generated it.

Similarly, studies using quantitative ethnographic methods reveal that learners interacting with Generative AI tend to ask more direct, pragmatic questions about conceptual and procedural knowledge than those interacting with human tutors (Cheng et al., 2025). AI appears to create a psychologically safe epistemic space in which learners are less hesitant to admit knowledge gaps. This openness facilitates immediate remedial action and metacognitive repair (Cheng et al., 2025).

The interactive nature of GenAI shifts pedagogy from passive reception to *active inquiry*. One of its most significant affordances is its role in stimulating help-seeking behaviors. Unlike traditional classrooms where students may avoid asking questions due to social evaluation concerns, AI platforms reduce perceived risk.

However, this affordance introduces a new metacognitive demand, that is effective prompt formulation. As Garg and Sood (2025) argue, "prompt engineering" emerges as a metacognitive skill in itself. Learners must translate implicit cognitive uncertainty into explicit, structured queries. This translation requires planning, monitoring, and evaluative judgment. Thus, the ability to ask productive questions becomes central to agency in AI-mediated environments.

From a *sociocultural perspective* (Vygotsky, 1978), AI can be understood as a *mediational tool* operating within the learner's Zone of Proximal Development. When learners use AI to explore linguistic alternatives, they engage in guided discovery. However, mediation must remain contingent. If AI over-scaffolds by providing complete reformulations without prompting reflection, internalization may weaken.

Beyond this, from a critical perspective, the metacognitive mirror is not neutral. AI feedback reflects linguistic norms embedded in training data. Learners must therefore be encouraged to critically evaluate suggestions:

- Does this reformulation align with my communicative intent?
- Does it privilege a particular dialect or register?
- Is this stylistic shift necessary or normative?

Without such interrogation, AI feedback risks becoming invisible authority rather than reflective scaffold.

This is not a hypothetical concern. A qualitative case study evaluating ChatGPT's responses across EFL writing, grammar, vocabulary, and conversation tasks found it linguistically accurate and broadly useful, but limited in pragmatic appropriateness and cultural sensitivity (Islomova, 2026)—precisely the register- and context-dependent dimensions of language that the interrogative questions above are designed to surface.

Zhang et al. (2025) caution that without explicit training in reflective AI use, feedback may devolve into superficial editing rather than cognitive restructuring. Learners may correct surface features while leaving underlying conceptual misunderstandings intact. Metacognition requires deeper engagement: recognition of patterns, strategy adjustment, and awareness of linguistic systems. Therefore, effective AI-mediated metacognition depends on pedagogical framing. Such practices convert AI from an automated corrector into a reflective instrument.

AI-supported writing instruction has also been shown to influence cognitive emotion regulation. Heydarnejad (2025) demonstrates that embedding structured self-evaluation within AI-assisted writing tasks improves self-competence and emotional regulation.

When learners use ChatGPT to refine writing, they engage in a recursive metacognitive loop:

1) Translate implicit thought into explicit prompt
2) Evaluate AI response
3) Identify gaps in logic or clarity
4) Revise either the prompt or the argument

This loop requires constant comparison between internal cognition and external representation (Yan et al., 2025). AI thus becomes both mirror and mediator.

This four-part loop now has a validated measurement counterpart. Drawing on focus groups with 12 students and a subsequent survey of 542 university L2 writers, Dong (2025) developed and validated a 22-item scale measuring cognitive, behavioral, emotional, and ethical engagement with ChatGPT feedback, operationalizing across four empirically distinguishable dimensions much of what the reflective loop above requires of learners.

In addition, AI-generated reflective prompts integrated into multimodal learning environments—such as educational videos—have been shown to trigger high levels of generative activity (Şimşek et al., 2025). Learners verbalize reasoning, test hypotheses, and articulate understanding in real time. Metacognition here becomes *multimodal* and *dialogic*.

8.2.3 Cognitive Depth Versus Cognitive Outsourcing

The expansion of AI feedback creates a paradox. As correction becomes more accessible, the risk of cognitive outsourcing increases. If learners habitually defer judgment to AI systems, internal monitoring mechanisms may weaken.

From a cognitive load perspective, AI reduces extraneous load but may inadvertently reduce germane processing if learners bypass reflective engagement.

Whether AI feedback increases germane load or forecloses it, moreover, is not uniform across learners. Tu et al. (2026), studying 98 university students grouped by their prior AI self-efficacy while using a knowledge-graph-integrated generative AI feedback agent, found that learners with high AI self-efficacy exhibited greater self-monitoring, lower cognitive load, and more accurate self-evaluation on complex revision tasks than their low-self-efficacy peers—the same tool appears to support cognitive depth for one group and something closer to outsourcing for the other. Tang et al. (2026) report a related asymmetry in collaborative problem-solving: students with lower prior knowledge interacted with an AI personal assistant more frequently and relied more heavily on its guidance than higher-knowledge peers, yet showed no significant difference in final performance—suggesting that for some learners, dependence on AI guidance can function as compensatory scaffolding rather than as outsourcing properly so called. The distinction between beneficial and detrimental offloading, in

other words, is not a property of the tool but of the interaction between the tool and the learner's existing self-regulatory capacity.

To mitigate the risks associated with AI feedback, a more active learning approach should be implemented. Instead of simply providing the correct answers immediately, AI systems should prompt learners to explain their reasoning and require them to predict necessary corrections first. Furthermore, educational platforms should encourage multiple alternative evaluations to foster critical thinking, while integrating reflection logs that explicitly document the student's interaction with the AI. Such strategies preserve cognitive effort while leveraging technological affordances.

AI feedback reaches its full pedagogical potential only when it is embedded within reflective ecologies that integrate human mediation, structured reflection, and critical AI literacy. Furthermore, this framework requires incorporating directive reasoning training alongside comprehensive prompt formulation instruction. Within these supportive and critical learning ecologies, AI shifts away from operating as an unquestioned, silent authority; instead, it serves as an active dialogic partner that directly fosters cognitive development.

Metacognition in AI-mediated language learning is therefore not simply enhanced by immediacy or adaptivity. It is cultivated through interrogation, comparison, intentional prompting, and ethical awareness.

AI feedback can transform corrective instruction into metacognitive scaffolding—but only when learners remain epistemic agents within the interaction.

8.2.4 Longitudinal Metacognitive Development

One of AI’s unique affordances is longitudinal tracking. Revision histories, performance analytics, and error logs allow learners to observe recurring patterns over time. Such temporal visibility strengthens metacognitive growth. Learners can identify persistent grammatical weaknesses, lexical over-reliance, pronunciation patterns and stylistic inconsistencies.

This longitudinal awareness supports higher-order self-regulation beyond immediate correction. It fosters strategic planning and intentional practice. However, the presence of analytics alone does not guarantee insight. Structured interpretation activities are required to transform data into metacognitive understanding.

Emerging longitudinal designs substantiate this claim rather than merely asserting it. Shi and Derakhshan (2026) surveyed 595 Chinese EFL students at three time points across an AI-powered course and used latent growth curve modeling to demonstrate genuine developmental trajectories, rather than a single cross-sectional snapshot, in autonomy, relatedness, and competence—the three needs specified by self-determination theory (Deci & Ryan, 2000). In a parallel 8-week quasi-experiment, Huang and Chen (2026) found that 53 intermediate-to-advanced Chinese-as-a-Second-Language undergraduates receiving ChatGPT-4.0 feedback showed greater gains in both writing performance and writing self-efficacy than a group receiving traditional teacher support. Wei (2023), in a controlled comparison of 60 EFL university students, similarly found that AI-mediated instruction raised self-regulated learning strategy use itself, and not merely its downstream outcomes, relative to traditional instruction. Together these studies suggest that AI-mediated metacognitive gains compound over instructional time rather than appearing only as an immediate, one-off effect of novelty.

AI feedback reaches its pedagogical potential when embedded within reflective feedback ecologies. Such ecologies include human mediation, structured reflection tasks, critical AI literacy and scaffolded autonomy. Within these ecologies, AI becomes a dialogic partner in cognitive development rather than a silent authority.

In summary, AI feedback can move language learning from delayed corrective intervention to continuous metacognitive scaffolding. It can transform learners from recipients of evaluation into agents of reflective inquiry. Yet this transformation is neither automatic nor technologically guaranteed. It depends on *pedagogical intentionality, cognitive framing,* and *critical awareness.*

Metacognition in AI-mediated language learning is not simply enhanced by immediacy. It is cultivated through interrogation, comparison, and reflective engagement.

8.3 Self-Paced and Self-Directed Learning

Artificial Intelligence has markedly transformed self-paced and self-directed learning (SDL) by enabling personalization, temporal flexibility, and adaptive scaffolding. Unlike traditional teacher-centered pacing models, AI systems allow learners to progress at individualized speeds, revisit

explanations, and explore content beyond curricular constraints. This temporal autonomy reduces performance anxiety while increasing rehearsal opportunities and sustained engagement.

Empirical evidence supports these affordances. Huang et al. (2025) found that AI chatbots and speech-based platforms in Asian tertiary EFL contexts increased learner participation and provided flexible rehearsal spaces for oral production. Self-paced AI interaction offers several pedagogical advantages, such as reduced performance pressure, increased rehearsal opportunities, adaptive task difficulty, continuous formative assessment.

Mateos-Blanco et al. (2024), in their catalog of 120 AI applications for language learning, emphasize that personalized learning pathways foster communicative competence by aligning instruction with individual learner profiles. The "anytime-anywhere" accessibility of AI technologies further expands these possibilities, making self-directed learning more scalable and inclusive.

From a *motivational perspective*, this flexibility aligns with Self-Determination Theory (SDT), which identifies autonomy as a fundamental psychological need essential for intrinsic motivation (Ode et al., 2025; Bognár & Khine, 2025). AI-supported environments enable learners to diagnose their own needs, set goals, and independently select learning resources (Kohnke et al., 2025). Such autonomy-supportive conditions enhance both engagement and persistence.

AI-assisted audio-learning modules provide a particularly compelling example. Jafarian and Kramer (2025) demonstrate that concise, AI-generated audio summaries increase reading engagement and academic achievement, especially among neurodiverse learners such as students with ADHD. By reducing extraneous cognitive load and supporting working memory management, these tools facilitate smoother self-paced progression through complex academic material.

AI integration within flipped classroom models further illustrates the synergy between individualized AI interaction and collaborative in-class learning. Adaptive systems can diagnose knowledge gaps prior to class sessions, helping students arrive better prepared and more confident for peer-based activities. This blending of independent AI-supported study with social constructivist classroom interaction strengthens learner agency while maintaining communal learning structures.

Beyond personalization, AI technologies increasingly support heutagogy—self-determined learning in which students retain ownership over their educational trajectories (Oliveira et al., 2025). Adaptive Learning Platforms (ALPs) exemplify this shift. By dynamically tailoring instructional sequences based on prior knowledge, motivation, and performance data, ALPs provide what Tan et al. (2025) describe as "step-loop adaptivity." Through continuous individualized feedback on each submission, these systems allow students in group settings to achieve learning gains comparable to one-on-one human tutoring.

To conceptualize effective human–AI collaboration in SDL contexts, Anders and Speltz (2025) propose an integrated Self-Regulated Learning (SRL) cycle consisting of three phases:

- *Plan:* Learners select appropriate AI tools, define learning goals, and establish ethical boundaries for AI use.
- *Iterate:* Students develop multi-step workflows, refine prompts, and engage in sustained dialogic interaction with AI systems.
- *Evaluate:* Learners critically assess both the quality of the final product and the effectiveness of the human–AI partnership.

This model reframes AI not as an autonomous authority but as a cognitive partner within a metacognitively guided process.

The design of that partnership matters empirically as well as theoretically. In a triple-blind randomized controlled trial with 456 EFL learners at CEFR B1–B2 level, Rui (2025) compared three AI dialogue modes—guided feedback, interactive question-and-answer, and inquiry-driven scaffolding—and found differential gains in computational thinking and language proficiency across the three conditions. The self-regulatory cycle AI is said to support, in other words, is an effect of specific dialogue-design choices, not an inherent property of conversational AI in general.

Significantly, evidence of AI-supported self-direction extends even to early childhood education. Xiao et al. (2025) observed that conversational agents embedded in interactive e-books cultivated task ownership among young learners. Children interacting with the agent "Mia" independently used multimedia affordances—such as replaying narration and rehearsing responses—to repair conversational breakdowns without adult intervention. These findings suggest that AI-mediated autonomy can develop foundational self-regulatory behaviors from early educational stages.

Broader evidence on child second-language learners points in a similar direction while sounding a similar caution. A mixed-methods study of AI tools in children's second-language learning found that speech recognition and interactive environments enhanced L2 skills through immediate feedback, while flagging ethical oversight and overdependence as unresolved concerns for this age group (Alkhutaba & Alnajjar, 2026).

Overall, AI-enhanced self-paced and self-directed learning represents a significant evolution in learner autonomy. By combining personalization, adaptive feedback, cognitive load optimization, and structured self-regulation cycles, AI systems expand the scope of learner agency. However, their educational value ultimately depends on intentional design, ethical guidance, and human pedagogical oversight.

8.4 Motivation, Affect, and Anxiety Reduction

8.4.1 AI as an Affective Buffer

The affective dimension of language learning—encompassing motivation, anxiety, trust, and self-efficacy—is profoundly reshaped by AI-mediated environments. In second language acquisition, the "*affective filter*" has long been recognized as a barrier to oral production and communicative risk-taking. AI systems, particularly conversational agents and speech-based platforms, function as non-judgmental interlocutors that lower this filter by offering psychologically safe rehearsal spaces.

Foreign language anxiety remains one of the most persistent inhibitors of oral participation. AI-mediated environments provide low-stakes opportunities for experimentation without fear of negative peer evaluation. Xu and Ismail (2025) report that AI speech systems improve fluency partly because they reduce evaluation-related anxiety. Similarly, Huang et al. (2025) demonstrate that AI chatbots increase learners' willingness to communicate (WTC), especially in speaking tasks requiring repeated attempts.

These findings are not confined to speech systems. In a six-week quasi-experimental study, Chen et al. (2025) found that AI-based instruction increased both positive achievement emotions and willingness to communicate among 204 Chinese EFL students relative to conventional instruction. The mechanism appears partly cognitive rather than purely affective: in a

survey of 454 Chinese EFL university students, Chen et al. (2026) found that teachers' AI behavior-management competence predicted willingness to communicate both directly and via reduced extraneous cognitive load, an effect that personalized AI nudges amplified further. This affective buffering effect is particularly powerful in pronunciation practice, role-play simulations, repetitive speaking drills, experimental lexical usage.

Because AI systems are perceived as patient, consistent, and emotionally neutral, learners are more willing to take risks and test emerging linguistic hypotheses. Tan et al. (2025) further confirm that learners often prefer practicing with AI before performing publicly, as it removes the social stakes associated with embarrassment or peer judgment.

However, anxiety reduction is not unidimensional. While AI lowers performance anxiety, it may simultaneously introduce forms of "*AI anxiety*" linked to privacy concerns, ethical uncertainty, or fear of technological replacement (Ode et al., 2025; Nazaretsky et al., 2025). *Trust* therefore becomes a crucial mediating variable. Students are more likely to engage deeply when AI systems are perceived as transparent, reliable, and pedagogically aligned. Interestingly, high-performing learners sometimes exhibit "*algorithm aversion*," doubting whether AI can accurately assess nuanced competencies (Jones-Jang et al., 2025). Thus, affective buffering depends not only on technological affordances but also on perceived credibility.

8.4.2 Engagement, Gamification, and Emotional Design

Beyond anxiety reduction, AI substantially influences motivation through *gamified interaction* and *emotionally responsive design*. Gamified AI environments—characterized by adaptive scoring systems, instant feedback loops, and performance dashboards—enhance intrinsic motivation when aligned with meaningful learning objectives (Mohebbi, 2025). Bibliometric analyzes reveal a growing convergence between AI research and affective engagement frameworks in language education (Yang et al., 2025).

Yet motivational gains are contingent upon task quality. If AI use is reduced to mechanical error correction or repetitive automation, engagement may decline over time (Zhang et al., 2025). Motivation thrives when AI facilitates meaningful, context-rich interaction rather than procedural substitution.

Several mechanisms explain AI's motivational power. For instance, in *immersive storytelling* Namaziandost and Çakmak (2025) demonstrate that AI-generated interactive narratives outperform traditional gamified platforms in promoting long-term vocabulary retention. By embedding lexical items in personalized, emotionally engaging storylines, AI enhances both cognitive depth and affective resonance. Second, *affective support features chatbots* incorporating humor, praise, emojis, and conversational warmth (e.g., "LogicalHamster") cultivate positive emotions such as enjoyment and comfort. These emotional cues are particularly effective in cognitively demanding tasks like argumentative writing or logical reasoning (Zhang et al., 2025). Also, *translanguaging scaffolds AI agents* that allow bilingual interaction support identity affirmation and meaning negotiation. Xiao et al. (2025) show that enabling home-language scaffolding reduces performance anxiety and enhances belonging, particularly in multilingual contexts.

From the perspective of the *Technology Acceptance Model (TAM)*, perceived enjoyment and trust strongly predict sustained AI adoption (Guo & Erdenebold, 2025). When learners view AI as a context-sensitive partner rather than a rigid evaluator, their self-efficacy and communicative confidence increase (Yan et al., 2025). Cross-cultural studies further suggest that autonomy and relatedness—central constructs in Self-Determination Theory—predict AI continuance intention (Ode et al., 2025). Notably, because AI tools scaffold performance effectively, perceived competence may become less of a barrier to initial adoption.

Controlled comparisons lend further weight to these mechanism-level claims. Hammami and Alomair (2026) tested an adaptive gamified multi-agent language-learning architecture and found significantly higher language proficiency and motivation in the experimental group than in a control group, suggesting that the motivational mechanisms described above translate into measurable proficiency gains rather than engagement alone.

8.4.3 Risks of Over-Reliance and Passive Learning

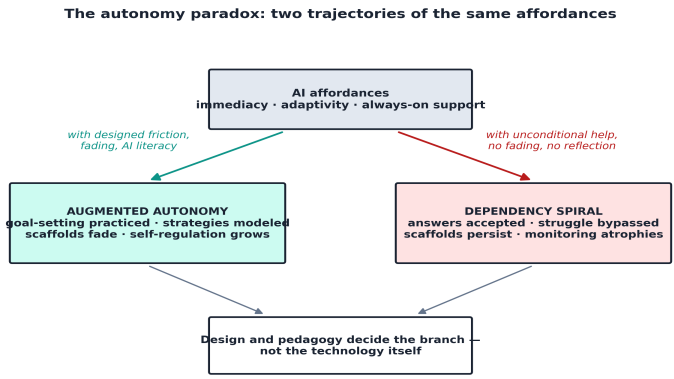


Figure 2. The autonomy paradox: identical affordances, divergent trajectories

Despite its affective and motivational benefits, AI introduces pedagogical risks. The convenience and immediacy of generative AI can encourage "*cognitive offloading*," where learners delegate thinking processes to the system (Radtke & Rummel, 2025). While such delegation may reduce short-term stress, it can undermine long-term skill development.

8.4.3.1 The Paradox of Dependency

Excessive reliance on AI is characterized by compulsive usage that interferes with independent learning. Academic stress may intensify this dependency; students under pressure sometimes use AI as an *avoidance mechanism*, bypassing productive cognitive effort (Acosta-Enriquez et al., 2025). Over time, self-regulatory skills and goal-setting capacity may weaken if AI scaffolding is not gradually faded (Bognár & Khine, 2025).

This dependency gradient is now measurable longitudinally rather than merely inferred. In a survey of 1,444 U.S. college-preparatory high school students, Xu et al. (2026) found that frequent AI use predicted lower self-confidence in academic work and a widening gap between reliance on the tool and belief in one's own unaided competence, a gap that was not significantly predicted by perceived school guidance in the same model. Complementing this, Franco (2026) identifies what he terms an "*attribution-efficacy paradox*" in a scoping review of higher-education studies: a paradox its name captures, since generative-AI assistance that measurably improves task performance coincides with erosion of learners' academic self-

efficacy rather than reinforcing it. Both findings converge on the mechanism named in Bandura's (1997) theory of self-efficacy: mastery beliefs require that learners attribute success to their own effort, and when improvement is externally supplied, that attributional link is weakened even as measured output improves.

Several interrelated risks emerge:

- *Skill Atrophy:* Overuse of AI for drafting or brainstorming may degrade independent writing and analytical skills (Ronksley-Pavia et al., 2025; Yan et al., 2025).
- *Homogenization:* GenAI-assisted outputs may become formulaic, reducing authorial voice and stylistic diversity (Ho & Lee, 2025).
- *Cognitive Offloading:* Frequent uncritical AI use negatively correlates with critical thinking and deep processing (Consoli & Petko, 2025).
- *Black-Box Opacity:* The opaque reasoning processes of AI models can obscure the rationale behind feedback, limiting students' ability to evaluate its validity (Nazaretsky et al., 2025).

Without structured guidance, learners may develop what researchers describe as "*metacognitive laziness*," outsourcing not only answers but also *evaluative judgment* (Fan et al., 2025).

These patterns have prompted a shift from anecdotal concern toward psychometric measurement. Hossain and Nawmi (2026) developed and validated a Generative AI Reliance Types Scale, indicating that undergraduate dependence on AI in academic writing differentiates into distinct reliance types rather than a single construct measurable by usage frequency alone—a distinction that matters pedagogically, since interventions calibrated to a blanket notion of "*over-reliance*" may miscalibrate for learners whose reliance pattern differs qualitatively. A survey of 1,580 Vietnamese university students lends corroborating support, confirming academic self-efficacy, academic stress, critical thinking, and AI dependency as reliable, discriminable constructs rather than a single undifferentiated syndrome (Viet et al., 2026).

8.4.4 Human-Centered Mitigation Strategies

To preserve learner agency, *Human-in-the-Loop* (HITL) frameworks are essential. Educators must transition from enforcement roles to "*support coaches*," emphasizing evaluative literacy, ethical reasoning, and critical interrogation of AI outputs (Anders & Speltz, 2025; Xu et al., 2025).

Assessment design should prioritize creation over replication, critical analysis over summary, process documentation over final product. As Cheah et al. (2025) argue, meaningful learning requires "active hands-on" engagement rather than passive consumption of AI-generated content.

The mitigation imperative is reinforced by the ambivalence found among stakeholders themselves. A survey of 329 students and 14 faculty at a Portuguese university found no consensus on whether generative AI erodes competency development—roughly half of each group disagreed that it does—but consensus on the need for institutions to provide explicit guidelines governing its academic use (Fernandes et al., 2025). This divergence in judgment, set against convergence on the need for governance, is itself evidence for the chapter's claim that the risks of AI reliance are contingent on instructional architecture rather than intrinsic to the technology.

AI-supported language learning environments hold transformative potential for reducing anxiety, increasing willingness to communicate, and enhancing intrinsic motivation. Through non-judgmental interaction, adaptive gamification, immersive storytelling, and translanguaging support, AI can lower affective barriers and strengthen learner self-efficacy.

However, this affective empowerment is accompanied by risks of dependency, cognitive offloading, and skill degradation. The central pedagogical challenge lies in maintaining a balance between support and substitution. AI must function as augmentation rather than replacement—an affective and cognitive scaffold that strengthens, rather than supplants, human agency.

This is particularly significant beacuse sustainable AI integration requires intentional design, ethical transparency, and the cultivation of critical AI literacy to ensure that learners remain the primary architects of their linguistic and intellectual development.

8.4.5 Automation vs. Agency

Despite its autonomy-enhancing potential, AI can inadvertently undermine learner agency when used as a shortcut rather than a scaffold. When learners rely on AI systems to generate ideas, correct errors, or construct entire texts without reflective engagement, agency shifts from the learner to the system. Multiple studies highlight risks including *over-reliance on AI-generated text* (Zhang et al., 2025), *reduced critical thinking in writing tasks*

(Acosta-Enriquez et al., 2025), *superficial revision without comprehension* (Xu & Ismail, 2025), and *diminished teacher-student interaction*.

Mohebbi (2025) cautions that AI-driven systems must be integrated cautiously to prevent dependency patterns that weaken self-regulated learning. While AI facilitates recursive feedback, it may also tempt learners to bypass *productive struggle*—a critical element in schema formation and long-term retention.

Set against these dependency risks, the aggregate picture is more encouraging than any single case study suggests. A meta-analysis pooling effect sizes from studies of AI-assisted learning's relationship to foreign-language learner agency found positive associations across every dimension examined, largest for engagement (r = .648) and smallest for enjoyment (r = .392), though flagging extreme heterogeneity across the included studies (Zai & Zhou, 2026). That heterogeneity may be the more important finding for this chapter's argument: a positive average effect coexisting with wide variance is what one would expect if the effect of AI on agency depends less on the technology than on how it is embedded pedagogically. At the level of individual outcomes, Qiao and Zhao (2023) similarly found that 93 Chinese EFL students randomly assigned to Duolingo-based AI instruction showed improved speaking self-regulation as well as speaking proficiency relative to a traditionally instructed control group, suggesting that self-regulatory gains and skill gains can move together rather than trading off.

8.4.6 Automation Bias and Epistemic Deference

A key cognitive risk in AI-mediated environments is automation bias—the tendency to uncritically accept machine-generated outputs. When learners perceive AI systems as authoritative, they may defer judgment rather than engage in evaluation.

From a metacognitive standpoint, this deference weakens monitoring processes. Instead of asking "*Why is this correction necessary?*", "*What principle governs this structure?*", "*Does this align with my communicative intention?*", learners may simply implement suggestions without analysis. Over time, such patterns reduce strategic flexibility and diminish internal regulatory mechanisms. The learner ceases to be an active decision-maker and becomes an executor of algorithmic suggestions.

Empirical work substantiates this risk while also complicating it. Analyzing 7,188 discussion threads comprising more than 117,000 utterances between students and a generative AI teachable agent, Xing et al. (2026) found that the AI's perceived authority within a dialogue was significantly associated with both students' agency and their learning outcomes—evidence that how much epistemic authority a system visibly claims is not a neutral design choice. Yet *automation bias* is not a fixed disposition: Le et al. (2025), in a 114-student experiment, found that learners preferred human-tutor feedback overall, but preference for ChatGPT's free-dialogue interface rose after use, complicating any simple narrative of uncritical deference. Mishra and Henriksen (2026) push the underlying critique further, arguing that "*agentic AI*" marketing borrows the moral weight of learner agency even as such platforms function as institutional "double agents," eroding the vocabulary available to defend student and teacher self-determination.

8.4.7 Cognitive Offloading and Productive Struggle

Cognitive psychology distinguishes between *beneficial cognitive offloading* (reducing unnecessary load) and *detrimental outsourcing* (bypassing essential processing). AI systems are highly efficient at reducing extraneous cognitive load, but they may inadvertently reduce germane load if learners skip analytical steps. For example, generating an entire essay outline through AI may prevent learners from engaging in conceptual organization, accepting AI-suggested vocabulary may bypass semantic network development, and allowing AI to restructure arguments may inhibit rhetorical reasoning.

Productive struggle—the cognitive effort involved in problem-solving and hypothesis testing—is essential for deep learning. If AI eliminates struggle entirely, it may also eliminate consolidation. Thus, the tension between automation and agency is not just technological; it is epistemological.

Recent work refines this distinction empirically rather than leaving it at the level of definition. Chow et al. (2026), analyzing the workflows of 188 postgraduate students, propose a four-tiered cognitive-offloading framework, finding that cognitive load is redistributed unevenly across the tiers of a task even as students themselves report that offloading freed capacity for higher-order critical thinking—a self-report finding worth reading alongside Yan et al.'s (2025) warning that generative AI can inflate performance without producing the deep processing genuine learning requires.

The risk is also unevenly distributed across learners, Seung and Basham (2026) argue that for students with learning disabilities, the same capacity to bypass effortful sub-processes that threatens productive struggle for some learners can remove a barrier that previously excluded others from a task altogether, making cognitive offloading simultaneously a risk and an accommodation depending on the learner's starting point. A scoping review of 22 empirical K-12 studies, meanwhile, found generative AI use associated with risks to *psychological wellbeing*, *cognitive dependency*, and *eroded creative authorship*, and *recommended process-focused task design* and explicit critical AI literacy curricula as correctives (Tao et al., 2026).

8.4.8 Passive Consumption vs. Active Construction

Language learning is fundamentally constructive. Learners build linguistic systems through active hypothesis testing, error correction, and strategic experimentation. When AI is used passively, this constructive process may weaken. Learners may accept AI corrections without analysis, submit AI-generated drafts with minimal revision, depend on AI for lexical retrieval, rely on AI to paraphrase rather than restructure ideas, avoid risk-taking by defaulting to AI suggestions. Such behaviors shift learning from active construction to passive consumption.

Yang et al. (2025) emphasize the need for equitable, teacher-inclusive AI frameworks to ensure sustainable adoption. AI should function within a "*human-in-the-loop*" model rather than as a replacement authority. AI-mediated writing ideally involves co-production: learners generate ideas, AI suggests refinements, and learners critically integrate revisions. However, when AI substitutes rather than scaffolds, authorship becomes blurred.

Instead of: *drafting → reflecting → revising → refining*

the process becomes: *prompting → accepting → submitting*

This procedural compression reduces cognitive engagement and weakens metacognitive loops.

Studies of emerging AI dependency likewise report reduced *critical engagement* (Acosta-Enriquez et al., 2025) in contexts where students relied heavily on AI-generated outputs without reflective integration. Similarly, Xu and Ismail (2025) identify superficial revisions where learners implemented corrections without understanding underlying principles.

This passive pattern, however, is not universal, and the contrast is instructive. In a study of 45 L2 students revising argumentative essays with ChatGPT feedback, Ziqi et al. (2024) found that learners rejected or ignored roughly 38% of the feedback offered—arguing against about 22% of it and disregarding a further 16% as inconvenient or unhelpful—and did so significantly more often for content-level suggestions than for form-level corrections. Far from uniform deference, this is selective, effortful engagement: learners exercised precisely the evaluative judgment automation bias is said to erode, at least regarding the substance of their arguments. Gyekye (2025), studying underserved U.S. high school students in a ChatGPT-integrated apprenticeship programme, similarly found that participants exhibited strong digital agency, critiquing AI-generated outputs rather than accepting them, despite starting from a position of unequal access to the technology. What distinguishes these cases from the dependency patterns reported elsewhere in this chapter appears to be less the tool itself than the task design: revision tasks that made rejection a visible, expected option, and apprenticeship structures that positioned students explicitly as critics rather than recipients, both cultivated the interrogative stance that Section 8.2.2 described as the metacognitive mirror in action. In such cases, AI ceases to function as a learning tool and becomes a performance optimizer detached from conceptual growth.

Another structural risk concerns the erosion of teacher–student interaction. If AI becomes the primary feedback provider, opportunities for dialogic meaning-making with human instructors may decline. Human feedback often includes *socio-emotional attunement*, *context-sensitive interpretation*, *discursive negotiation* and *ethical framing.*

While AI can simulate conversation, it lacks lived relational reciprocity. Over-reliance may reduce exposure to authentic communicative negotiation.Yang et al. (2025) therefore emphasize the necessity of *teacher-inclusive AI frameworks*. Sustainable AI adoption depends on human-in-the-loop models in which educators remain central mediators rather than peripheral supervisors.

Over-reliance also carries motivational risks. If learners attribute linguistic improvement to AI rather than to their own strategic effort, self-efficacy may weaken. According to *Social Cognitive Theory* (Bandura, 1997), efficacy beliefs depend on perceived personal mastery. If mastery is

externalized, persistence may decline once AI support is removed. Additionally, habitual AI reliance may shape learner identity. Students may begin to perceive themselves not as developing writers or speakers but as operators of intelligent systems. This shift subtly reconfigures epistemic ownership.

Autonomy is more than merely about completing tasks independently; it involves developing confidence in one's cognitive capacity. If AI replaces rather than reinforces this capacity, autonomy erodes. Over-reliance risks are not evenly distributed. Learners with weaker metacognitive skills or lower prior knowledge may be more vulnerable to passive patterns. Without explicit AI literacy training, these learners may default to acceptance rather than interrogation.

This differential vulnerability is precisely the concern Ramos-Benitez et al. (2026) address in proposing a "*critical alliance*" model of AI in education, in which cognitive independence, equitable access, and faculty development function jointly as the cornerstones of AI literacy rather than as separable add-ons. On this account, balanced autonomy is an equity issue as much as a cognitive one: learners who enter AI-mediated environments with weaker prior literacy or less institutional support are the same learners most exposed to the dependency risks this section has documented. Thus, dependency is not simply an individual choice; it is shaped by instructional design.

Yang et al. (2025) argue that equitable AI integration requires teacher-inclusive governance structures. Human oversight ensures that AI supports developmental trajectories rather than shortcuts them.

8.5 Designing for Balanced Autonomy

If AI-mediated environments redistribute agency across human and non-human actors, then autonomy cannot be left to emerge spontaneously. It must be deliberately designed.

Balanced autonomy refers to a pedagogical configuration in which AI enhances metacognitive development, strategic independence, and confidence without displacing cognitive effort, epistemic responsibility, or dialogic interaction. Such balance requires intentional instructional architecture rather than ad hoc tool adoption. To preserve agency while leveraging

AI's benefits, educators must embed AI within structured reflective ecosystems.

8.5.1 Teaching AI Literacy Explicitly

In contemporary language education, AI literacy has transitioned from being an optional skill to an absolute necessity. Rather than focusing merely on operational competence, or how to use a tool, instruction must advance toward critical and metacognitive awareness by exploring how these technologies actively shape human cognition. To achieve this, explicit AI literacy instruction must be intentionally designed to help learners understand how generative models produce output and recognize their inherent limitations, including potential hallucinations. Furthermore, a comprehensive framework requires students to identify algorithmic bias and normative linguistic standards, differentiate between a system's suggestion and actual authority, and consistently practice effective prompt construction.

Prompt formulation itself constitutes a metacognitive act. Learners must translate implicit uncertainty into explicit inquiry. Teaching prompt engineering as reflective questioning reinforces planning, monitoring, and evaluation processes central to self-regulated learning. AI literacy is therefore both technical and epistemic. It safeguards autonomy by preventing uncritical deference. Reflection logs, revision commentaries, and error pattern analyzes convert AI feedback into developmental insight. Hybrid models—combining AI personalization with teacher-guided metacognitive reflection—appear most promising (Mohebbi, 2025; Xu & Ismail, 2025).

Recent scholarship has moved beyond advocating AI literacy in principle toward defining and measuring it directly. Annapureddy et al. (2025) formalize the construct as twelve defining competencies, giving language educators a concrete curricular target for what AI-literacy instruction should cover rather than a general aspiration toward critical awareness. Jin et al. (2025) introduce the Generative AI Literacy Assessment Test (GLAT), a named instrument through which programs could measure whether explicit instruction actually raises learners' literacy rather than merely assuming it does. Within prompt formulation specifically—already identified in this chapter as a metacognitive act in its own right—Knoth et al. (2024) argue that AI literacy shapes the sophistication of the prompting strategies learners adopt, meaning that prompt-engineering instruction and broader AI-literacy instruction are not separable strands but the same developmental

process viewed at different grains. For *agentic AI systems* specifically, in which the assistant increasingly initiates rather than merely responds, Dandan et al. (2026) argue that EFL learner AI literacy must expand further still, to include agent orchestration, critical evaluation of autonomous suggestions, metacognitive monitoring of the collaboration itself, and ethical responsibility for outcomes the system helped produce—reframing self-regulated learning, in their account, as distributed human-AI collaboration rather than an individual competence AI merely supports.

The integration of generative artificial intelligence in language education can only reach its full pedagogical potential through a *hybrid structure* that carefully balances technological affordances with robust institutional governance and human-centric pedagogy. While AI excels at managing low-level automation—such as providing immediate feedback, facilitating individualized pacing, and guiding repetitive rehearsal—*human educators* remain profoundly irreplaceable for delivering *contextual interpretation, emotional attunement, ethical framing, and critical dialogue*. To successfully cultivate this *hybrid environment* and prevent the risks of passive learning or cognitive offloading, instruction must move beyond basic operational competence toward explicit AI literacy, encouraging students to actively analyze longitudinal error analytics, experiment with prompt construction, and differentiate between algorithmic suggestions and definitive authority. Ultimately, maintaining this delicate *pedagogical equilibrium* requires educational institutions to enforce a rigorous framework of ethical governance—*comprising transparent use policies, data protection safeguards, and academic integrity guidelines*—to ensure that technological autonomy is equitably distributed and purposefully leveraged to foster deeper metacognitive growth rather than tool dependency. Ethical governance reinforces pedagogical intentionality. Without it, autonomy becomes unevenly distributed and vulnerable to misuse.

8.6 Conclusion

Artificial Intelligence has not simply entered language education as a new tool; it has reconfigured the ecology within which autonomy, self-regulation, and agency are constructed. Throughout this chapter, we have traced how AI-mediated feedback, personalization systems, conversational agents, and generative drafting platforms reshape the cognitive and

sociotechnical architecture of learning. What emerges is neither a narrative of technological salvation nor one of pedagogical decline. Instead, AI in language education presents a structural paradox: it simultaneously expands and redistributes agency.

At the *cognitive level*, AI offers unprecedented scaffolding. Immediate, adaptive feedback enhances monitoring. Recursive drafting strengthens reflection. Analytics render developmental trajectories visible. When embedded within structured metacognitive practices, these affordances reinforce the cyclical processes described in Self-Regulated Learning theory (Zimmerman, 2002). AI can externalize cognition, making thought processes visible and therefore revisable. In such configurations, technology amplifies strategic awareness.

At the *motivational level*, AI-mediated environments can reduce anxiety, promote competence, and increase willingness to communicate. Speech-recognition tools provide low-stakes rehearsal spaces. Conversational agents reduce social evaluative pressure. Personalized feedback increases perceived efficacy. When aligned with Self-Determination Theory, AI can support autonomy and competence simultaneously.

At the *sociocultural level*, AI functions as a mediational tool operating within learners' Zones of Proximal Development. It scaffolds linguistic production while enabling iterative experimentation. However, mediation must remain contingent and dialogic. When scaffolding becomes substitution, internalization weakens.

At the *posthuman level*, agency no longer resides exclusively within the learner. It is distributed across human and non-human actants: algorithms, datasets, institutional policies, classroom practices, and assessment systems. From this perspective, autonomy is not lost; it is reassembled within sociotechnical networks. The question shifts from "Does AI reduce autonomy?" to "How is autonomy redistributed, and who retains epistemic authority?"

This redistribution introduces real risks. Empirical evidence reveals patterns of over-reliance, automation bias, superficial revision, and diminished dialogic interaction. AI can compress drafting processes, bypass productive struggle, and subtly displace epistemic responsibility. Without structured reflection and critical AI literacy, learners may internalize algorithmic authority rather than interrogate it.

Thus, the core insight of this chapter is that autonomy in AI-mediated language learning is not a technological outcome but a pedagogical construction, meaning that its success or failure depends entirely on the design of the instructional architecture. *Balanced autonomy* emerges when AI feedback is actively interrogated rather than implemented automatically, drafting processes remain highly visible and reflective, prompt construction is treated as a metacognitive exercise, teachers retain dynamic dialogic oversight, and a critical awareness of algorithmic bias is deliberately cultivated. Conversely, *unbalanced autonomy* occurs when artificial intelligence substitutes for human reasoning rather than scaffolding it, revision processes become purely mechanical rather than deeply analytical, epistemic authority shifts invisibly to the system, and product optimization completely replaces active process engagement.

AI-mediated language education therefore demands a reconceptualization of autonomy itself. Autonomy can no longer be understood as isolated individual control over learning processes. In technologically saturated environments, *autonomy* becomes *relational, reflective, and distributed.* It requires the capacity to navigate hybrid cognitive systems intentionally.

The educator's role correspondingly shifts. Rather than guarding against AI or surrendering to it, educators become *designers of reflective ecosystems.* They orchestrate the interaction between human cognition and machine mediation. They structure pauses for interpretation. They cultivate critical algorithmic awareness. They ensure that productive struggle remains part of learning trajectories.

With the integration of artificial intelligence, the learner's role undergoes a fundamental transformation; an *autonomous learner* in AI-mediated environments is neither one who completely avoids the technology nor one who relies on it uncritically. Instead, true autonomy is exhibited by learners who can *engage with AI dialogically, continuously question machine-generated suggestions,* and *explicitly attribute their academic progress to personal strategic effort.* Furthermore, these individuals successfully *balance operational efficiency with critical cognitive depth*, demonstrating the sophisticated judgment required to recognize exactly when to accept, revise, or reject algorithmic input. In the end, such learners do not relinquish their individual agency; rather, they actively exercise it within complex, distributed systems.

In the final assessment, AI in language education represents neither an autonomy revolution nor an agency crisis. It is both possibility and paradox. Its impact depends less on algorithmic sophistication than on pedagogical intentionality.

The future of AI-mediated autonomy in language education will be decisively shaped by *explicit AI literacy instruction, human-in-the-loop governance,* and *process-oriented assessment models*, alongside a critical examination of embedded linguistic norms and the establishment of ethical institutional frameworks. Ultimately, autonomy in the post-AI era does not mean preserving a traditional, pre-digital form of independence; rather, it demands the intentional cultivation of critically conscious, technologically literate agents who possess the reflective judgment necessary to successfully navigate complex, distributed sociotechnical assemblages.

In this reconfigured landscape, human judgment remains indispensable. Not because technology is insufficient, but because education is not merely about optimization of output. It is about the formation of thinking subjects. AI may assist in producing better sentences, clearer pronunciation, and more coherent arguments. Only pedagogy, however, can cultivate autonomous minds capable of deciding why those sentences matter, how those arguments function, and when those tools should be used—or resisted.

In this sense, the future of autonomy in language education is not determined by artificial intelligence. It is determined by how human intelligence designs its integration.

Pedagogical Companion for Chapter 8

Key Concepts: learner autonomy; self-regulated learning; metacognitive scaffolding; forethought–performance–reflection cycle; affective buffer; gamification; dependency paradox; automation bias; cognitive offloading; productive struggle; AI literacy.

Discussion Questions

1. The chapter's central paradox: the same automation that enables autonomy can undermine it. Trace this mechanism through one concrete learner scenario from your experience.

2. When is cognitive offloading to AI a rational learning strategy, and when is it self-defeating? Propose a rule of thumb a learner could actually apply.
3. What observable behaviors would distinguish a learner using AI as a "metacognitive mirror" from one using it as an answer machine?
4. Gamified AI systems reduce anxiety and increase persistence. At what point does emotional design become manipulation?
5. Holec defined autonomy as the capacity to take charge of one's own learning. Does delegating decisions to an adaptive system exercise that capacity or surrender it?

Further Reading

- Holec, H. (1981). *Autonomy and foreign language learning*. Pergamon.
- Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits. *Psychological Inquiry, 11*(4).
- Molenaar, I. (2022). Towards hybrid human-AI learning technologies. *European Journal of Education, 57*(4).

Chapter 9. Multimodality, Interaction, and Identity

9.1 Introduction: From Text to Multimodal Ecologies

Artificial Intelligence (AI) has precipitated a structural transformation in language education that extends far beyond increased efficiency or instructional automation. The current phase of AI integration—marked by the rapid development of generative language models, conversational agents, speech-recognition systems, image generators, and immersive multimodal platforms—signals not merely a technological shift, but a fundamental reconfiguration of interaction itself. Language learning is no longer confined to text-based Computer-Assisted Language Learning (CALL) exercises or isolated grammar drills. Instead, it now unfolds within complex, *multimodal ecologies* that combine text, voice, image, video, gesture, and immersive environments in dynamic and adaptive ways.

This transformation is not only technological; it is *epistemological* and *ideological*. AI-mediated environments reshape how interaction is structured, how communicative competence is conceptualized, how learner identities are constructed, and how linguistic norms are validated. They also influence how *global English* is imagined and circulated across *digital spaces*. In this sense, AI does not simply support language learning—it participates in redefining what it means to communicate, to belong linguistically, and to be recognized as a competent language user.

Recent bibliometric analyzes confirm that research in AI and language education increasingly clusters around themes such as *emotional engagement, multimodal communication, interactive feedback systems,* and *AI-supported pedagogy* (Yang et al., 2025; Zhang et al., 2025). Parallel systematic reviews document significant growth in AI-mediated speaking instruction, adaptive writing feedback, and environments designed to support self-regulated learning (Mohebbi, 2025; Wang et al., 2025; Xu & Ismail, 2025). These findings suggest that AI integration is not peripheral but central to contemporary discussions of communicative development.

At the *policy level*, international organizations have similarly recognized the transformative potential—and risks—of AI in education. UNESCO (2021), the European Commission (2022), and the OECD (2023) all emphasize the necessity of *human-centered, ethical, transparent, and inclusive AI deployment*. Their frameworks foreground *digital literacy,*

algorithmic accountability, equity of access, and *respect for cultural and linguistic diversity*. These concerns are particularly salient in language education, where issues of identity, normativity, and representation are deeply embedded in pedagogical practice.

This chapter examines AI-mediated language education through four interrelated dimensions: (1) multimodal interaction across text, voice, image, and immersive media; (2) identity construction within AI-mediated environments; (3) cultural and pragmatic competence in algorithmically shaped discourse; and (4) the tension between native-speaker norms and pluralistic Englishes. By synthesizing empirical research, sociocultural theory, multimodal frameworks, and international policy perspectives, the chapter argues that AI-mediated language education must be understood not simply as technological innovation, but as a sociolinguistic and ideological reconfiguration of communicative subjectivity.

Multimodal theory reminds us that communication extends beyond verbal language to include gesture, gaze, image, spatial organization, and sound. Contemporary AI tools integrate text generation, speech recognition, image creation, video synthesis, virtual avatars, and immersive virtual reality environments into unified learning systems. In the *Vademecum of Artificial Intelligence Tools Applied to the Teaching of Languages* (Mateos-Blanco et al., 2024) over 120 AI applications are identified relevant to the development of communicative competence, illustrating the expanding semiotic resources now available to learners.

Yet multimodality is not neutral. The combination of text, sound, and image shapes which identities are represented, which cultural narratives are normalized, and which communicative styles are privileged. AI-generated discourse and imagery reflect the data on which these systems are trained, thereby embedding cultural assumptions and normative frameworks into seemingly neutral technological outputs. For this reason, multimodal AI environments must be examined within sociocultural and ideological analysis rather than treated as purely functional instructional tools.

From a *sociocultural perspective*, learning is mediated by tools and artifacts that structure cognitive development (Lantolf & Thorne, 2006). AI systems function as powerful mediational artifacts: they scaffold interaction, externalize feedback, structure task progression, and regulate learner attention. Mohebbi's (2025) systematic review demonstrates that AI tools can

significantly enhance self-regulated learning, metacognitive awareness, and learner autonomy. However, mediation through AI is always shaped by algorithmic design choices, raising important questions about agency, epistemic authority, and the distribution of control between human educators and automated systems.

In this emerging landscape, AI is not simply assisting language learning—it is reshaping the conditions under which interaction, identity, and linguistic legitimacy are constructed. Understanding this transformation requires moving beyond instrumental perspectives and engaging critically with the multimodal, sociocultural, and policy dimensions of AI-mediated communication.

9.2 Modes of AI-Mediated Interaction: Text, Voice, and Vision

The integration of Artificial Intelligence into language education has fundamentally transformed the nature of interaction as well. Where earlier forms of CALL emphasized input delivery and discrete-item correction, contemporary AI systems simulate *dialogic engagement across text, speech, and increasingly, multimodal environments*. AI-mediated interaction is not simply a technical enhancement of communication; it reshapes the temporal, cognitive, and affective dynamics of language learning.

From a pedagogical standpoint, the central issue is no longer whether AI can provide feedback, but how algorithmically generated interaction mediates meaning-making differently from human interlocution. AI systems replicate certain features of conversational exchange—turn-taking, reformulation, clarification—but they do so without shared social context, lived experience, or embodied co-presence. This distinction has significant implications for cognition, agency, and identity in language learning.

9.2.1 Text-Based Interaction: Generative Dialogue and Writing Mediation

From an *interactionist Second Language Acquisition (SLA) perspective* (Long, 1996), language development emerges through negotiation of meaning, corrective feedback, and modified input. Contemporary generative AI systems increasingly simulate these interactional processes through clarification requests, adaptive prompts, recasts, scaffolded expansions,

and reformulations. Text-based models such as ChatGPT engage learners in ongoing dialogue, creating the appearance of negotiated interaction.

However, these exchanges differ structurally from human conversation. They are *algorithmically generated* rather than socially embedded. While AI can approximate interactional patterns, it does not participate in shared intentionality, mutual contingency, or sociocultural co-construction. The pedagogical question, therefore, is not whether AI replicates interaction, but how simulated dialogue influences cognitive processing, learner motivation, and affective engagement differently from human interlocution.

Generative AI platforms have transformed writing instruction in particular. Instead of functioning as static correction tools, they operate as *dialogic co-authors, engaging learners in iterative drafting, revision,* and *expansion*. Research demonstrates that AI-assisted writing can enhance lexical diversity (Ngo, 2024), improve structural coherence, and stimulate idea generation (Pham, 2025). Learners benefit from immediate feedback, contextual suggestions, and adaptive elaboration, often leading to increased textual complexity.

One of the most significant shifts introduced by AI-mediated writing is the transformation of feedback temporality. Traditional classroom feedback is delayed, allowing time for reflection and productive struggle. In contrast, AI provides instantaneous reformulation and correction. Mohebbi (2025) shows that such immediacy can enhance self-regulated learning, metacognitive monitoring, and strategic revision. Learners are able to experiment, receive feedback, and adjust in real time.

Yet immediacy carries risks. When corrections are delivered instantly and repeatedly, learners may bypass deeper cognitive processing, relying on AI suggestions rather than engaging in meaningful revision. Over-reliance may diminish productive struggle—the cognitive friction necessary for durable learning. As a result, AI-mediated writing environments must be accompanied by explicit metacognitive scaffolding to ensure that learners critically evaluate, rather than passively accept, automated feedback.

Text-based AI interaction also reconfigures authorship. Learners move fluidly between multiple roles: *author, editor, prompt designer, evaluator*, and sometimes even *curator of AI-generated content*. This role fluidity expands creative possibilities but complicates traditional notions of textual ownership. The OECD (2023) cautions that generative *AI can blur authorship*

boundaries and underscores the necessity of AI literacy education to preserve academic integrity and cognitive engagement.

In this sense, generative AI enhances productivity while simultaneously challenging established pedagogical assumptions about authorship, originality, and intellectual labor. Learners no longer operate solely as producers of text; they become managers of algorithmic collaboration. Agency is redistributed rather than eliminated.

As a result, text-based AI interaction represents both an expansion and a destabilization of traditional writing pedagogy. It enables dialogic co-construction, accelerates feedback cycles, and supports lexical and structural development. At the same time, it requires critical framing to prevent cognitive outsourcing and to maintain the learner's epistemic authority. AI does not replace interaction—it reframes it, shifting the locus of negotiation from social exchange to algorithmic mediation.

9.2.2 Voice-Based Interaction: Speech Recognition and Affective Buffering

Voice-based interaction represents one of the most transformative developments in AI-mediated language learning. Advances in Automatic Speech Recognition (ASR), speech synthesis (Text-to-Speech; TTS), and conversational AI have enabled learners to engage in spoken interaction with systems capable of *real-time transcription*, *pronunciation scoring*, *fluency analytics*, and *adaptive conversational feedback*. These developments shift speaking practice from occasional classroom performance to *continuous, on-demand rehearsal*.

AI-powered speech systems now provide real-time pronunciation scoring, fluency and pacing analytics, prosodic feedback on stress, rhythm, and intonation, and simulated conversational rehearsal environments. Systematic reviews demonstrate measurable improvements in pronunciation accuracy, fluency, and speaking confidence in AI-supported contexts (Xu & Ismail, 2025; Wang et al., 2025). Particularly in tertiary EFL settings, learners report increased willingness to communicate and greater perceived control over their speaking development.

One of the most significant findings across empirical studies is what may be termed the *affective buffering effect*. Learners frequently experience reduced anxiety when interacting with AI systems rather than human

interlocutors. In many Asian educational contexts, where public performance anxiety and fear of negative evaluation are well documented, AI-mediated rehearsal provides a psychologically safe space for experimentation (Wang et al., 2025). Unlike classroom speaking tasks, AI systems do not display visible judgment, impatience, or social comparison. Learners can repeat utterances, adjust pronunciation, and rehearse complex expressions without fear of embarrassment. This repeated, low-stakes engagement appears to support both confidence and automaticity.

This affordance is not confined to affect. In a five-week quasi-experiment with 70 Chinese university students, Zou et al. (2023) found that social network-based interaction with AI speaking applications improved English speaking skills more than non-interactive use of the same tools, a pattern corroborated across surveys, interviews, and pre/post-tests. The finding matters because it isolates interaction itself, rather than mere exposure to an AI system, as the active ingredient: rehearsal without a responsive interlocutor—human or artificial—appears to yield weaker gains than rehearsal within a structured exchange. Affective buffering, in other words, is necessary but not sufficient; it must be paired with genuinely interactive design.

From a *second language acquisition perspective*, this environment may increase opportunities for *comprehensible output* and *reduce affective filters*, potentially facilitating phonological development. However, the reduction of social pressure also alters the social dimension of language use. Speaking to AI is interactionally simulated but not socially reciprocal. The implications of prolonged AI-mediated rehearsal for authentic communicative competence therefore require careful pedagogical consideration.

Recent ASR systems, such as OpenAI's Whisper, have achieved near human-level transcription accuracy under controlled conditions and demonstrate robustness even in noisy environments (Radford et al., 2022). In educational settings, ASR has expanded beyond simple transcription to include reading speed analysis, error detection, prosodic evaluation, and automated fluency scoring. In reading fluency assessment, *ASR systems* can now collect and analyze data that previously required extensive teacher time, enabling scalable formative assessment and individualized tracking (da Silva et al., 2025).

However, important limitations remain. ASR models trained predominantly on adult speech corpora often struggle with the acoustic variability

of children's speech (Attia et al., 2024). Disfluencies may be automatically "corrected" by the system, masking genuine pronunciation or decoding difficulties. In multilingual contexts, accent variation further complicates recognition accuracy. Thus, while ASR enhances efficiency and data richness, it also introduces new layers of interpretive complexity.

The population gap is not confined to recognition accuracy alone. Mixed-methods evidence from child second-language learners finds that speech-recognition tools and interactive AI environments enhance L2 skills through immediate feedback, but that ethical concerns and overdependence risks persist alongside those gains (Alkhutaba & Alnajjar, 2026). Where an adult EFL learner can weigh an AI system's feedback against independent metalinguistic judgment, a child learner is, on the whole, less equipped to do so—which suggests that the accuracy gap documented above for children's speech compounds a second, developmental vulnerability rather than standing apart from it.

Despite technological sophistication, speech recognition systems frequently privilege *standardized accents*—most commonly *General American* or *Received Pronunciation* varieties. Xu and Ismail (2025) note that systems often misrecognize or penalize legitimate phonological variation. When AI scoring systems equate intelligibility with conformity to dominant accent norms, they risk reinforcing narrow definitions of "correct" English.

This concern intersects directly with UNESCO's (2021) ethical AI framework, which emphasizes inclusivity, cultural diversity, and avoidance of algorithmic bias. If speech systems systematically validate only certain phonological patterns, they may *reproduce global linguistic hierarchies* rather than democratize access. Language is not just a technical code but a marker of identity, belonging, and sociocultural positioning. When AI systems privilege standardized accents, they implicitly position those accents as linguistically superior, with implications for learner identity formation, confidence among non-native speakers, and perceptions of legitimacy in global communication. Voice-based AI therefore becomes a site where technological design intersects with *linguistic ideology.*

Technologies such as intelligent voice assistants and specialized language-learning applications enable interactive dialogue through ASR and TTS integration. These systems *translate speech to text, generate responses, and*

provide synthesized auditory feedback. Voice interaction offers a pressure-free environment for oral experimentation. However, *conversational AI lacks* genuine *pragmatic contingency, shared sociocultural experience,* and *embodied co-presence.* Learners engage in interactional simulation rather than fully reciprocal dialogue. The absence of human unpredictability may reduce anxiety but also limits exposure to authentic negotiation of meaning.

From a *pedagogical standpoint*, voice-based AI should therefore *complement* rather than replace human interaction. AI can scaffold phonological development and build confidence, but pragmatic competence and intercultural nuance remain deeply social phenomena.

At the *policy level*, both UNESCO (2021) and the OECD (2023) stress that AI systems in education must be human-centered and equity-oriented. In the context of voice-based language learning, this requires transparent explanation of scoring criteria, accent-inclusive model training, regular bias auditing, and teacher oversight in the interpretation of analytics. Without such safeguards, automated speech scoring may inadvertently entrench phonological hierarchies and reproduce global inequalities.

Voice-based AI considerably expands speaking opportunities. It increases rehearsal frequency, reduces anxiety, provides real-time analytics, and supports autonomous learning. These benefits are pedagogically substantial. Yet voice-based AI simultaneously raises a critical question: whose speech counts as correct? If AI systems validate only standardized phonological norms, they risk narrowing the spectrum of legitimate Englishes. If designed inclusively, however, they could promote intelligibility-based assessment and pluralistic pronunciation models aligned with global English usage.

Voice interaction therefore represents both *opportunity* and *ideological risk*. It enhances access to oral practice while foregrounding the politics of validation embedded in algorithmic design. In this sense, AI-mediated voice interaction does not simply teach pronunciation—it participates in defining phonological legitimacy in the digital age.

9.2.3 Visual and Multimodal AI: Image, VR, and Generative Media

AI tools increasingly integrate visual and multimodal elements, including image generation, virtual reality (VR), and avatar-based interaction. The comprehensive mapping of over 120 AI applications in language teaching

demonstrates that communicative competence is no longer developed solely through textual or auditory input, but through multimodal design that combines linguistic, visual, and interactional affordances (Mateos-Blanco et al., 2024). In this expanded ecosystem, language learning becomes an orchestration of words, images, movement, and digital embodiment.

Multimodality enhances meaning-making by combining linguistic, visual, and paralinguistic cues. When learners engage with AI-generated images, immersive simulations, or avatar-based environments, they do more than merely decode text; they interpret gesture, gaze, facial expression, spatial orientation, and visual symbolism. In VR-supported environments, learners participate in simulated intercultural encounters in which proxemics, posture, and turn-taking are integrated into communicative performance. Such environments approximate authentic communicative contexts more effectively than text-only tasks, supporting situated learning and pragmatic awareness.

Image generation tools extend this process by enabling learners to externalize imagination into visual artifacts. Large-scale text-to-image diffusion systems require learners to translate abstract ideas into precise prompts. Prompt engineering thus becomes a metalinguistic and creative act: students must articulate conceptual abstractions clearly enough for a generative system to co-construct an image. This process strengthens lexical precision, descriptive specificity, and pragmatic intentionality. In design-oriented contexts, the use of systems such as DALL·E allows immediate visualization of narrative scenes, cultural scenarios, or role-play contexts, enhancing motivation and creative engagement, although concerns remain regarding the potential erosion of traditional manual or artistic skills (Miralay, 2024).

Multimodal Large Language Models (MLLMs), such as LLAVA-7B, further bridge vision and language by enabling systems to interpret visual input alongside textual or spoken descriptions. These models can evaluate learners' interpretations of images, assess descriptive accuracy, and provide adaptive formative feedback. By integrating behavioral logs, interaction traces, and, in experimental contexts, even eye-tracking data, multimodal analytics produce a more granular understanding of the learning process. The result is not merely automated correction but dynamic insight into attention patterns, descriptive strategies, and semiotic interpretation.

Embodied conversational agents represent a further convergence of text, voice, and visual interaction. Systems such as "Kid Space" integrate multimodal sensing with a projected 3D environment and an animated agent to support collaborative problem-solving among children. By combining spoken dialogue with physical interaction and shared visual space, such environments demonstrate that AI can facilitate joint attention, social coordination, and embodied meaning-making rather than isolated individual cognition (Aslan et al., 2025). In these blended physical-digital spaces, language learning becomes *co-constructed* through gesture, movement, and shared perceptual focus.

Immersive AI environments offer several pedagogical benefits. They enable situated communicative practice, simulate culturally embedded scenarios, support rehearsal of gesture and non-verbal communication, and deepen role-play immersion. By embedding learners within context-rich scenarios—ordering food in a virtual café, negotiating in a simulated marketplace, or participating in intercultural dialogue—VR-based systems can scaffold pragmatic competence and sociolinguistic sensitivity.

These benefits should not be read as uncomplicated. A review of virtual-reality language-learning applications by Cowie and Alizadeh (2025) confirms gains in motivation, contextualized learning, and embodied cognition, but sets against them persistent barriers: teachers unfamiliar with VR platforms, institutions unable to resource headsets and technical support at scale, and cybersickness among learners. The last of these is easy to overlook in accounts that treat immersion as an unqualified pedagogical good—embodiment, the very feature meant to deepen situated practice, can instead produce physical discomfort for some users. VR's pedagogical promise, that is, is real but resource- and body-contingent.

However, multimodal AI also shapes representation. Visual outputs generated by AI systems are not neutral; they are products of training datasets that reflect historical, cultural, and geopolitical imbalances. Generative systems may reproduce stereotypical imagery, narrow beauty standards, or culturally homogenized depictions unless critically mediated. The European Commission (2022) emphasizes transparency, dataset diversity, and bias mitigation in AI systems to prevent discriminatory or reductive representations. In language education, this concern becomes particularly salient

when visual prompts shape learners' perceptions of cultures, identities, and social roles.

The pattern is not hypothetical. Wandera's (2024) case-study analysis of Alice and Sparkle, a children's book authored by an AI chatbot, identified embedded cultural tropes recurring across its depictions of ethno-racially diverse characters. The analysis was undertaken specifically to inform recommendations for inclusive AI-literacy pedagogy, underscoring that critical media literacy cannot be an afterthought bolted onto multimodal AI use; it is the mechanism by which learners and teachers detect representational harm that the generating system itself has no means of flagging.

Teachers therefore play a crucial role not only as facilitators of technology use but as mediators of semiotic interpretation. Visual literacy and critical multimodal analysis must become integral components of AI-supported curricula. Learners should be encouraged to question how images are generated, whose perspectives are represented, and how cultural narratives are encoded in visual form. Without such critical framing, multimodal AI risks reinforcing implicit hierarchies and unexamined cultural assumptions.

Multimodality, then, is not simply technological. It is *ideological* and *pedagogical*. It expands the semiotic repertoire available to learners, enabling embodied, situated, and creative engagement with language. At the same time, it introduces new responsibilities related to representation, bias, and digital literacy. Visual and immersive AI systems extend language learning into embodied interaction and collaborative cognition, but they also participate in constructing the cultural imaginaries within which language operates. As such, their integration into language education demands not only innovation but critical, reflective, and ethically informed pedagogy.

9.3 Identity Construction in AI Interaction

AI-mediated interaction profoundly reshapes how learners perceive themselves as language users. Identity in language learning is no longer a standalone achievement emerging solely from human-to-human interaction; rather, it is increasingly networked and co-constituted through engagement with technological artifacts. When learners interact with AI chatbots, multimodal agents, or generative systems, they are not merely practicing language—they are negotiating voice, agency, and belonging within digitally mediated communicative ecologies.

AI can function simultaneously as interlocutor and mirror. As an interlocutor, it provides a low-risk environment that encourages experimentation. Learners often report adopting more exploratory linguistic identities when interacting with AI, taking pragmatic risks or attempting complex formulations they might avoid in human interaction (Mohebbi, 2025). The absence of visible social judgment reduces anxiety and supports identity play. In such environments, learners may experiment with tone, register, discourse roles, and narrative positioning. AI-mediated spaces thus operate as identity sandboxes—arenas for rehearsing alternative voices and testing communicative selves.

Yet AI also functions as a normative mirror. Bibliometric mapping of AI-language research reveals a dominant emphasis on fluency, accuracy, and standardized performance metrics (Yang et al., 2025). These metrics implicitly define what counts as a "*competent speaker*." When learners repeatedly receive feedback aligned with native-like benchmarks or standardized norms, they may internalize these definitions of competence as aspirational identity models. AI thus simultaneously liberates experimentation and reflects back institutionalized linguistic standards. The tension between exploration and normalization is central to identity construction in AI-mediated learning.

Autonomy and *agency* are deeply implicated in this process. AI tools enable self-paced, personalized practice and promote learner independence (Mohebbi, 2025). Recommendation systems suggest tasks, highlight errors, and prioritize particular forms of revision. However, this autonomy is algorithmically mediated. What learners practice, what they see, and what they revise are partially shaped by opaque design decisions embedded in algorithmic systems. Zhang et al. (2025) caution that AI integration in language education often lacks strong theoretical grounding, leading to fragmented pedagogical implementation. Without explicit critical framing, learners may treat AI feedback as authoritative rather than provisional.

Algorithmic authority can subtly reshape *epistemic agency*. Because AI systems appear technologically sophisticated, learners may attribute to them a level of epistemic certainty that exceeds their actual reliability. UNESCO (2021) emphasizes that AI must remain human-centered and that educators must maintain pedagogical oversight. *Identity construction* in AI contexts therefore intersects with *AI literacy*. Learners need to understand

how generative systems produce outputs—how probabilities, pattern recognition, and training data shape responses—in order to maintain epistemic autonomy. Without such understanding, algorithmic authority may displace critical reflection, leading learners to align their linguistic identities uncritically with machine-generated norms.

Anthropomorphism further complicates identity negotiation. Children and adolescents frequently perceive AI chatbots not just as tools but as social actors or learning companions. This attribution of human characteristics influences communicative behavior: users often adopt politeness strategies, emotional disclosure, and socially oriented interaction styles when engaging with AI. In mobile-assisted contexts, AI systems can accompany learners continuously, nurturing what has been described as "lifelong mobility," where autonomous agents appear to "know" their users and support evolving identities as global language users. Such persistent companionship blurs the boundary between tool and partner, embedding AI more deeply into the learner's sense of self.

The integration of AI also generates *distributed agency*. Rather than being passive recipients of information, learners become evaluators and co-constructors of meaning alongside technological agents. *Knowledge production* becomes collaborative across human and machine. At the same time, excessive reliance on AI raises concerns about dehumanization. Some researchers warn that sustained interaction with non-sentient agents may attenuate the affective and relational dimensions that are foundational to language as a social practice. If language learning becomes overly mechanized, learners may experience a form of identity dissonance, particularly when authentic emotional reciprocity is absent.

Student *identity* is especially *dynamic* in AI-mediated environments. AI systems often act as "more knowledgeable others," scaffolding learning and increasing self-efficacy. Research indicates that interaction with AI chatbots can facilitate a transition from reticence to participation, reducing speaking anxiety and increasing willingness to communicate (Zhang et al., 2024). In high-stakes contexts such as interviews, AI tools that provide interpretable feedback on prosody, facial expression, or gesture empower learners to refine both verbal and non-verbal performance (Mawalim et al., 2025). These feedback loops can strengthen confidence and expand perceived communicative competence.

However, tensions emerge between technological assistance and authentic voice. Students have expressed concern that overreliance on generative AI for writing may homogenize linguistic output, diluting individual style and narrative distinctiveness (Kohnke et al., 2025). When AI is treated merely as a command-driven instrument rather than a dialogic partner, interaction tends to become instrumental and transactional, lacking the social coordination markers present in human-human tutoring (Cheng et al., 2025). Identity, in this case, risks becoming streamlined toward efficiency rather than expressive individuality.

Concern of this kind is not merely anticipatory. A comparative case study of two Saudi university English cohorts found that ChatGPT-assisted writing improved measurable linguistic accuracy, yet students actively resisted AI-suggested edits that diluted religious or gendered moral framing in their writing, overriding the tool's suggestions to preserve a voice they recognized as their own (Alzahrani & Rafi, 2026). The episode reframes homogenization as a contest rather than an inevitability: learners are not passive recipients of algorithmic streamlining but active editors of their own algorithmic editors, willing to sacrifice a measure of polish to keep an identity-bearing feature of the text intact.

Teacher identity is equally in flux. The presence of AI as a communicative and evaluative partner alters the "social life" of education (Williamson, 2024). Educators may fear displacement as traditional instructional roles are reconfigured (Lee et al., 2024). Yet a holistic integration perspective suggests that AI should mediate rather than replace teacher-student interaction. By offloading administrative tasks and providing preliminary feedback, AI allows teachers to focus on *emotional attunement, ethical guidance,* and *high-level pedagogical orchestration* (Guan et al., 2025). Teachers increasingly function as *cognitive engineers* and *curators of human-machine symbiosis*. However, pre-service teachers often conceptualize AI mechanistically, struggling to integrate it into a relational understanding of pedagogy—highlighting a gap in professional preparation for this existential shift (Guan et al., 2025).

Inclusive identity construction requires deliberate design. AI-mediated environments are not neutral spaces; they shape representation and belonging. Lee et al. (2025) argue that inclusive AI education must intentionally center students' cultural backgrounds and lived experiences. *Digital storytelling*

projects using augmented reality, for example, allow learners to embed personal narratives into curricular tasks, promoting ownership and authenticity. A student who develops a project inspired by a family member's profession, connecting personal identity with technological creation, experiences AI not as an abstract tool but as a medium for self-articulation.

At the same time, the "*black-box"* nature of deep learning systems poses risks. If generative models rely on biased or unrepresentative datasets, they may reproduce gendered, racialized, or cultural stereotypes. Such representations can marginalize learners from underserved populations and subtly constrain identity possibilities. A *Human-in-the-Loop* (HITL) approach is therefore essential, ensuring that educators remain active participants in monitoring, interpreting, and contextualizing AI outputs. Human oversight protects against homogenized identity prescriptions and reinforces learner agency.

To sum up, identity in AI-mediated language learning is dynamic, negotiated, and technologically entangled. AI functions as sandbox, mirror, scaffold, and sometimes gatekeeper. It can expand voice, confidence, and participation while simultaneously reflecting normative standards and algorithmic biases. Whether AI amplifies agency or constrains identity depends on pedagogical framing, AI literacy, transparency, and ethical oversight. Identity construction in the age of AI is therefore more than merely a psychological process but a socio-technical one—shaped by design decisions, data infrastructures, and the critical awareness of learners and educators alike.

9.4 Cultural and Pragmatic Competence in AI Contexts

Cultural and pragmatic competence represent some of the most complex and consequential frontiers of AI-supported language education. While AI systems demonstrate remarkable accuracy in grammatical correction and lexical suggestion, pragmatic competence extends beyond structural correctness. It involves *context-sensitive appropriateness, politeness strategies, relational positioning, sociocultural nuance,* and *the tacit knowledge* that governs how meaning is negotiated within specific communities.

AI-based speaking and writing tools have been shown to *enhance fluency* and *surface-level communicative confidence* (Xu & Ismail, 2025; Wang et

al., 2025). However, contextualized feedback at the discourse level remains limited. AI systems may correct syntactic form while overlooking pragmatic infelicity. For example, a grammatically accurate request may still be socially inappropriate in a high-power-distance context. Without embodied sociocultural embeddedness, AI-generated dialogue can simulate politeness formulas but cannot fully capture lived pragmatic complexity. Learners may therefore internalize decontextualized routines—formulaic expressions detached from cultural grounding—unless teacher mediation provides contextual anchoring.

Pragmatic competence also depends on interpreting subtle em*otional undercurrents, irony, metaphor,* and *non-verbal signals.* Chatbots and generative models often struggle to interpret these dimensions, particularly when cues are culturally specific or interactionally layered (Guan et al., 2025). Equally, the "black-box" architecture of many deep learning models makes it difficult for educators to understand why a response is flagged as inappropriate or reformulated in a particular way (Mawalim et al., 2025). This opacity limits the pedagogical value of AI for teaching complex social nuance.

Direct evidence bears this out. A qualitative case study evaluating ChatGPT's responses to EFL writing, grammar, vocabulary, and conversation tasks found the tool linguistically accurate and broadly useful across those domains, but specifically limited in pragmatics and cultural sensitivity (Islomova, 2026). The asymmetry is instructive: the same system that reliably corrects a subject-verb agreement error or supplies an idiomatic collocation cannot be relied upon to judge whether a request is too direct for the relationship it addresses, or whether an idiom carries connotations its user did not intend. Structural competence and pragmatic competence are evidently not the same capacity, and AI systems currently trained to excel at the first show no guarantee of correspondingly excelling at the second.

Recent developments show partial progress. Advanced systems can annotate pragmatic functions, distinguishing between formal and colloquial expressions and mapping usage across communicative contexts. Large language models are increasingly capable of simulating learners at different CEFR proficiency levels, potentially helping teachers anticipate pragmatic development stages (Benedetto et al., 2025). Multimodal systems can

simulate business etiquette training, intercultural dining scenarios, or culturally specific parting rituals. Yet simulation is not equivalent to sociocultural immersion. The absence of lived community experience constrains the depth of pragmatic modeling.

A direct test of this constraint returned a sobering result. In a pre-post study of 125 EFL undergraduates (87 analyzed), an AI chatbot purpose-built for dynamic assessment of language pragmatics did not outperform either classroom-based dynamic assessment or direct corrective feedback; classroom dynamic assessment, in turn, outperformed direct feedback at both immediate and delayed posttests (Saeedi & Soltani, 2025). If a system built specifically for pragmatics instruction cannot match what a human-mediated classroom procedure already achieves, the case for treating AI as a substitute—rather than a supplement—for pragmatic instruction weakens considerably. A correction notice has since been issued for this article; its specifics were not available for review, and the finding is reported here with that caveat attached.

Cultural representation presents an equally significant challenge. AI training datasets heavily influence the communicative norms reflected in generated outputs. If datasets disproportionately represent Anglophone or Western communicative styles, generated discourse may privilege *WEIRD (Western, Educated, Industrial, Rich, Democratic)* norms. This monocultural bias can result in hallucinated cultural conventions, oversimplified etiquette patterns, or inaccurate modeling of culturally specific acts—such as declining a gift in Chinese culture or navigating hierarchical address forms in collectivist societies.

Research increasingly foregrounds these concerns. Yang et al. (2025) identify cultural bias, data sovereignty, and equitable AI adoption as emerging focal points in AI-language scholarship. Cultural background markedly shapes learner engagement with AI tools. For example, students in collectivist contexts may prioritize relational harmony and instructor guidance in adopting AI technologies, whereas learners in individualistic contexts may exhibit greater autonomous engagement (Ode et al., 2025). AI systems must therefore be examined not only for linguistic performance but for cultural adaptability.

Translation systems illustrate this tension. Deep learning architectures such as BiLSTM can improve lexical and syntactic accuracy, yet cultural nuance often

requires additional evaluative layers. Fuzzy logic algorithms have been proposed to assess "cultural adaptability," ensuring that translation preserves cultural connotation rather than reducing complex concepts to literal approximations (Kong & He, 2025). Machine translation may accurately render denotative meaning while stripping away ceremonial, historical, or metaphorical significance—such as in culturally embedded literary traditions.

Policy frameworks underline these responsibilities. European Union digital education strategies emphasize intercultural dialogue and diversity, while UNESCO highlights cultural preservation and equitable representation in AI deployment. *Intercultural Communicative Competence* (ICC), a core goal of contemporary language education, requires learners to adopt perspective-taking capacities and recognize cultural variability in communicative norms. AI can facilitate exposure to simulated intercultural encounters, but only if its outputs are critically interrogated.

To encourage *intercultural awareness*, educators must design *reflective mediation tasks*. Learners can compare AI-generated politeness strategies across cultural contexts, analyze how different English varieties encode indirectness, or evaluate representation in AI-generated images and dialogues. They can examine whether generated climate descriptions default to four-season Western patterns, disregarding ecological realities in the Global South. They can interrogate how cultural festivals, kinship structures, or indigenous knowledge systems are represented—or omitted.

Innovative approaches attempt to address monocultural bias. Culturally Responsive Lesson Planners (CRLP), grounded in Culturally Responsive Pedagogy (CRP), incorporate interactive semi-automated prompt architectures that collect local cultural parameters before generating instructional material. By integrating regional languages, indigenous knowledge systems, and community assets into content generation, such systems demonstrate that AI grounded in educational theory can produce more culturally appropriate and curriculum-relevant materials. Customized tools with anthropomorphic personas—such as gamified AI writing tutors—can also encourage experimentation in a non-judgmental way while maintaining cultural sensitivity.

Despite these advances, a "*complexity ceiling*" persists. AI struggles with metaphorical density, emotionally charged discourse, and deeply contextual allusions. Pragmatic competence requires understanding when not to speak, how to calibrate deference, and how to interpret silence—dimensions not

easily encoded in statistical models. Cultural and pragmatic knowledge are embedded in social practice and community memory, not merely textual frequency.

This is not solely a limitation of underlying model capability; it is also a limitation of deployment. A content analysis of dialogues generated by three widely used AI language-learning applications—Duolingo, ImmerseMe, and MakesYouFluent—found that all three underutilize the natural-language-processing capabilities available to them for developing learners' emotional and intercultural competencies (Sriganesh & Eamer, 2025). In other words, even where a more pragmatically and culturally attuned exchange is technically achievable, the commercial products through which most learners actually encounter conversational AI have not built toward it. The complexity ceiling described above is therefore partly a property of current products, not only of current models.

Therefore, the human instructor remains indispensable. Educators act as validators, contextualizers, and cultural interpreters. They mediate between AI-generated content and lived sociocultural practice, ensuring that learners develop behavioral culture—the tacit knowledge of effective reaction in specific contexts—rather than relying on surface formulae. Human-in-the-Loop approaches safeguard against bias amplification and ensure that AI amplifies rather than prescribes communicative identity.

AI can support intercultural exposure and pragmatic exploration, but only through critical mediation. Without reflective scaffolding, learners risk internalizing monocultural norms or misinterpreting culturally flattened outputs as universally valid. Cultural and pragmatic competence in AI contexts therefore depends not only on technological sophistication but on pedagogical intentionality, theoretical grounding, and ethical oversight.

9.5 Native-Speaker Norms vs. Pluralistic Englishes

The deployment of AI in language education places the long-standing tension between *native-speaker norms* and *pluralistic Englishes* at the center of contemporary pedagogical debate. Generative AI systems and automated assessment tools are largely trained on dominant English corpora, frequently privileging standardized American and British varieties. As a

result, native-speaker ideology—long critiqued in applied linguistics—can be inadvertently reinforced through algorithmic design.

The critique is not merely inherited from pre-AI applied linguistics; it remains empirically live. Chan (2024) examines the construct of the "native speaker" itself within Hong Kong English, a context where the category has long been contested and unevenly applied despite English's status there as a working language rather than a foreign import—a reminder that the benchmark AI systems are calibrated against is not a neutral technical default but a sociolinguistically unstable one.

Pronunciation scoring systems, for example, often calibrate accuracy against General American or Received Pronunciation benchmarks. Such calibration risks marginalizing World Englishes and legitimate regional varieties. In EFL contexts, learners may interpret lower AI scores not as markers of difference but as evidence of deficiency. This subtle normalization process can perpetuate linguistic hierarchies and internalized deficit narratives.

The problem extends beyond pronunciation. Many NLP-based grammar correction systems flag non-standard dialects or pluralistic Englishes as "errors," reinforcing monolingual native-speaker norms. During alignment and fine-tuning phases, large language models often down-weight texts that exhibit features associated with developing writers or non-standard grammar, privileging polished, formal prose. This bias toward standardized forms risks entrenching linguistic hegemony if not explicitly addressed. OECD (2023) warns that AI systems, if uncritically implemented, can reproduce and amplify structural inequities embedded in training data.

The mechanism by which this happens has been documented directly. Hofmann et al. (2024) show that large language models generate covertly racist decisions about people based on dialect features alone: text bearing surface markers of African American English elicits harsher hypothetical judgments from a model than the same content rendered in standardized American English, even though the same model does not display comparable prejudice when asked directly about race. For language education, the implication is that dialect-based discrimination in these systems is not incidental noise awaiting a tuning fix, but a learned association between a linguistic variety and a social judgment—one reproduced with a fluency that

makes it harder, not easier, to detect than in older, overtly rule-based NLP tools.

Automated assessment raises additional fairness concerns. Studies examining GPT-4 as a grader for short-answer responses have explored whether bias emerges against non-native speakers. While reliability may be high overall, factors such as readability and rhetorical style—often influenced by linguistic background—can inadvertently affect scoring (Rodrigues et al., 2025). Similarly, ASR systems used for oral assessment must be rigorously validated to ensure they do not penalize accented speech, as such penalties would undermine assessment validity for diverse learner populations (da Silva et al., 2025). *Accent bias* in speech recognition remains a documented challenge, with systems often demonstrating higher accuracy for native varieties while struggling with non-native or regional accents.

The same asymmetry surfaces outside grading altogether, in the tools built to police AI use itself. Liang et al. (2023) found that GPT-detection software—increasingly deployed by institutions to flag suspected AI-generated student writing—is biased against non-native English writers, misclassifying their authentically human-written text as AI-generated at markedly higher rates than text produced by native-English-speaking peers. The irony is pointed: a technology adopted to defend academic integrity against AI can itself penalize the very population—multilingual and pluralistic-English writers whose prose departs from a narrow statistical model of "natural" English—that this chapter argues AI systems already underserve.

These risks highlight the persistence of native-speaker ideology in AI-mediated language learning. However, AI also possesses transformative potential for *pluralistic and world Englishes.* Contemporary language education increasingly recognizes *English as a Lingua Franca (ELF),* emphasizing communicative effectiveness over native-like conformity. Zhang et al. (2025) call for pedagogically aligned AI implementation that acknowledges linguistic diversity rather than enforcing monolithic standards. Mohebbi (2025) similarly emphasizes inclusive design and equitable access as foundational to responsible AI integration.

Pluralistic AI design should include *accent-inclusive speech recognition models*, *intelligibility-based assessment frameworks* that distinguish clarity from nativeness, exposure to multiple English varieties across regions, and

transparent explanations of scoring criteria. Feedback should explicitly differentiate between communicative comprehensibility and alignment with a particular phonological norm. When learners understand the basis of scoring, they are less likely to equate standardized output with intrinsic linguistic superiority.

AI can also facilitate *translanguaging* practices. Large language models possess extensive multilingual capacities that can support learners in drawing on their home languages to scaffold meaning-making in English. Rather than treating multilingualism as interference, AI tools can enable bilingual conversational agents, code-switching simulations, and parallel phrasing strategies that reflect authentic global communication patterns. Learners frequently craft prompts in multiple languages to clarify complex concepts, demonstrating agency and viewing multilingual repertoires as resources rather than deficits.

Evidence from *English-medium-instruction (EMI)* writing classrooms supports this reading of multilingual prompting as agentive rather than compensatory. An interpretive qualitative study of 26 Indonesian EMI undergraduates found that generative AI functioned as a genuine writing adjunct supporting multilingual meaning-making and authorial agency—not mere assistance—as students drew on their full linguistic repertoires while composing (Kalisa & Hadi, 2026). The distinction the study draws, between AI as adjunct and AI as assistant, matters pedagogically: an adjunct is incorporated into the learner's own meaning-making process, while assistance implies a process that remains fundamentally the tool's.

In large-scale writing assessment, AI has shown potential to evaluate dimensions such as organizational coherence with consistency. Yet fairness checks are essential to ensure that model-generated response patterns do not systematically disadvantage English Language Learners due to differences in rhetorical tradition or discourse structure. Equity in automated assessment requires continuous bias auditing and dataset diversification.

Beyond mitigating bias, AI can actively support pluralistic Englishes. Metaverse-based chatbot environments have been used to enhance pre-service teachers' confidence in using English as a Lingua Franca, encouraging engagement with diverse global interlocutors rather than adherence to native-speaker norms (Lee et al., 2024). AI-generated contrastive analyzes can highlight sociolinguistic variation, such as differences between

regional varieties, fostering cultural empathy and awareness of language adaptability. Inter-comprehension methodologies can use AI to illuminate shared linguistic patterns across related languages, reducing intimidation and reinforcing interconnected linguistic knowledge.

The broader shift reflects a paradigmatic reorientation in language education—from striving toward native-speaker idealization to embracing communicative effectiveness in multilingual contexts. Reliance on English-medium instruction historically marginalized local languages and knowledge systems. AI, if responsibly designed, can instead amplify linguistic diversity by incorporating region-specific data, supporting translanguaging, and foregrounding global Englishes.

Nonetheless, vigilance remains essential. Without deliberate design choices, AI systems may reproduce Western-centric communicative styles, reinforce gendered or class-based norms, and marginalize non-standard varieties. The "*stochastic parrot*" critique highlights that models reflect patterns in their training data rather than normative linguistic truth. Human-in-the-loop oversight, transparent documentation of datasets, and culturally responsive fine-tuning are necessary safeguards.

To conclude, AI has the potential either to entrench native-speaker hegemony or to democratize English use globally. The outcome depends on whether designers, educators, and policymakers consciously resist monolithic normativity. By prioritizing intelligibility over nativeness, promoting exposure to diverse varieties, integrating multilingual affordances, and embedding fairness auditing into assessment systems, AI can align with pluralistic approaches to language. In doing so, language education can move from monolingual standardization toward inclusive, globally situated communicative competence.

9.6 Conclusion

AI-mediated language education represents not simply a technological development but a sociolinguistic transformation. The shift from CALL and MALL toward AI-integrated environments signals a paradigmatic reconfiguration of interaction, identity, authority, and normativity in language learning (Yang et al., 2025; Zhang et al., 2025). Multimodal affordances—including speech recognition, generative dialogue, image synthesis, and

immersive simulation—expand communicative possibilities and create new spaces for experimentation and participation (Mateos-Blanco et al., 2024; Xu & Ismail, 2025).

However, these affordances are not ideologically neutral. AI systems embed assumptions derived from dominant corpora, design architectures, and alignment processes. Research demonstrates that while AI enhances fluency, engagement, and autonomy (Mohebbi, 2025; Wang et al., 2025), it may simultaneously reinforce standardized linguistic norms and native-speaker ideologies if left critically unexamined (Lyu et al., 2025; OECD, 2023). Pronunciation scoring systems and automated assessment tools risk privileging dominant phonological and rhetorical conventions, subtly shaping learner identity and perceptions of legitimacy.

The intersection of interaction and identity is particularly salient. AI environments often function as low-risk identity sandboxes, enabling learners to experiment with voice and discourse roles (Mohebbi, 2025). Yet algorithmic authority may be perceived as epistemically superior, especially when generative systems provide immediate corrective feedback. UNESCO (2021) emphasizes that AI in education must remain human-centered, requiring transparency, explainability, and sustained pedagogical oversight to prevent the displacement of critical thinking.

Cultural and pragmatic competence further reveal the complexity of AI integration. While generative systems can simulate dialogue and annotate pragmatic functions, they lack lived sociocultural embeddedness (Xu & Ismail, 2025; Wang et al., 2025). Monocultural bias in training datasets may privilege WEIRD communicative norms, requiring educators to actively mediate AI outputs through intercultural reflection and culturally responsive frameworks (Yang et al., 2025).

At the same time, AI holds transformative potential for pluralistic Englishes. Multilingual large language models can support translanguaging, exposure to global English varieties, and intelligibility-based assessment rather than nativeness-based evaluation (Zhang et al., 2025). When designed inclusively, AI can contribute to linguistic democratization rather than homogenization.

Moving forward, the field must cultivate AI literacy that extends beyond operational competence to include ethical awareness, ideological critique, and identity-conscious pedagogy (Ng et al., 2021). A learner-centered,

human-in-the-loop approach ensures that AI amplifies rather than standardizes, supports rather than replaces, and expands rather than constrains linguistic diversity.

As a result, AI-mediated language education should aim *for human-machine symbiosis grounded in inclusivity and pluralism*. When critically integrated, AI can enhance multimodal interaction, support pragmatic development, and affirm diverse linguistic identities. When uncritically adopted, it risks reinforcing structural inequities and monolithic norms. The responsibility, therefore, lies not in the technology itself but in the pedagogical, ethical, and ideological frameworks guiding its use.

Pedagogical Companion for Chapter 9

Key Concepts: multimodal ecologies; text-, voice-, and vision-based interaction; generative media; cultural and pragmatic competence; native-speaker norms; pluralistic and world Englishes; identity; ideology.

Discussion Questions

1. As AI interaction expands from text to voice and vision, which dimensions of communicative competence become newly teachable—and which become newly assessable, with what risks?
2. AI systems trained predominantly on standard varieties model a particular English. Whose English is your AI tutor teaching, and what are the identity consequences for your learners?
3. Can pragmatic competence—irony, politeness, indirectness—be learned from an interlocutor that has no face to lose? Design a task that tests your answer.
4. The chapter connects interaction, identity, and ideology. Give one example of an apparently neutral design choice in a language app that embodies an ideological position.
5. Should language education aim to prepare learners for communication with AI systems as a target context in its own right?

Further Reading

- Hofmann, V., Kalluri, P. R., Jurafsky, D., & King, S. (2024). AI generates covertly racist decisions about people based on their dialect. *Nature, 633*.
- Godwin-Jones, R. (2024). Distributed agency in language learning and teaching through generative AI. *Language Learning & Technology, 28*(2).
- Can, T., Elia, A., & Şimşek, İ. (2017). Immersive game-based vocabulary teaching in Second Life. *RumeliDE*(10).

PART IV

AI IN LANGUAGE ASSESSMENT

Chapter 10. Automated Assessment and Feedback

10.1 Introduction

The rapid integration of Artificial Intelligence (AI) and Large Language Models (LLMs) into language education has fundamentally transformed the ways in which language learning, teaching, and assessment are conceptualized and implemented. Among the various domains influenced by these technological developments, language assessment has experienced one of the most significant shifts, moving from traditional, human-centered evaluation practices toward intelligent systems capable of providing real-time assessment, continuous monitoring, and personalized feedback. Historically, language assessment relied primarily on expert human raters who evaluated learner performance through established rubrics, professional judgment, and periodic testing procedures. While human evaluation remains indispensable for capturing contextual, pragmatic, and sociocultural dimensions of language use, it is often constrained by practical challenges such as time requirements, scoring inconsistency, rater fatigue, and difficulties associated with large-scale implementation. As educational institutions increasingly seek efficient, scalable, and learner-centered assessment solutions, AI has emerged as a powerful force capable of transforming how language performance is evaluated and how feedback is delivered.

The integration of AI into language assessment has been facilitated by remarkable advances in Natural Language Processing (NLP), Machine Learning (ML), Deep Learning (DL), Automatic Speech Recognition (ASR), and, more recently, Large Language Models (LLMs). These technologies enable computers to process, analyze, interpret, and generate human language with a degree of sophistication that was unimaginable only a decade ago. For this reason, AI-powered systems are now capable of evaluating written texts, assessing spoken language, diagnosing grammatical and lexical weaknesses, and providing immediate, individualized feedback to learners. This technological evolution has enabled a transition from static, periodic assessment toward *dynamic, continuous, and adaptive feedback loops* that support learning throughout the educational process (Tan et al., 2025). Beyond this, AI has expanded opportunities for personalized and on-demand feedback, reducing the delays traditionally associated with teacher-generated responses and enabling learners to engage more actively in self-regulated learning processes.

The growing influence of AI coincides with broader shifts in educational assessment paradigms. Contemporary approaches increasingly emphasize *formative assessment*, *learner autonomy, and continuous feedback* rather than relying exclusively on summative examinations. AI technologies are particularly well suited to supporting these goals because they can provide immediate responses, monitor learner progress over time, and adapt feedback according to individual learning needs. Rather than functioning solely as automated grading tools, modern AI systems increasingly serve as intelligent learning companions that support revision, reflection, and self-regulated learning. This transformation aligns closely with contemporary theories of assessment for learning, which position feedback as an integral component of the learning process rather than a post hoc evaluation of performance. According to bibliometric analyses of AI research in language education, assessment-related applications have become one of the fastest-growing areas within the broader field of AI-assisted language learning, reflecting increasing interest among researchers, educators, and policymakers in leveraging AI for educational evaluation (Jaleniauskienė et al., 2023).

Recent developments in Generative AI have accelerated these changes even further. Large Language Models such as ChatGPT and similar transformer-based systems possess capabilities that extend well beyond traditional automated scoring approaches. Unlike earlier rule-based systems that focused primarily on identifying surface-level linguistic errors, contemporary LLMs can analyze complex dimensions of language use, including *coherence, cohesion, discourse organization, argumentation quality, and communicative effectiveness.* Generative AI has introduced sophisticated chatbot technologies that simulate human-like evaluative judgment and interaction, enabling dynamic forms of assessment that extend beyond simple scoring mechanisms. These systems can generate explanatory feedback, provide individualized recommendations, answer learner questions, and engage users in interactive dialogues that support both assessment and learning. As a result, AI-mediated assessment increasingly encompasses not only the evaluation of learning products but also the support of learning processes. Such developments raise important questions regarding the evolving nature of language competence in an AI-rich educational environment, where learners may be expected not only to demonstrate language proficiency but also to interact critically and effectively with intelligent technologies.

Recently, AI technologies offer the capability to rapidly process and evaluate complex language data, opening new avenues for *personalized, adaptive,* and *on-demand feedback*. This growing capability has contributed to the widespread adoption of AI in language assessment and has stimulated innovations in automated writing evaluation, speech scoring and pronunciation assessment, grammar and vocabulary diagnostics, and other forms of diagnostic feedback (Rodrigues et al., 2025). Through these developments, assessment is increasingly being reconceptualized as a continuous component of the learning process rather than a discrete event conducted at the end of instruction. AI systems can identify learner needs in real time, provide targeted support, and facilitate individualized learning pathways that respond dynamically to learners' strengths, weaknesses, and developmental trajectories.

Despite these opportunities, significant concerns remain regarding the implementation of AI in assessment contexts. Researchers have highlighted *issues related to reliability, validity, transparency, algorithmic bias,* and *ethical governance.* While AI systems can provide highly consistent evaluations, consistency alone does not guarantee that assessment outcomes accurately represent language proficiency. Questions persist regarding the extent to which AI systems can capture complex communicative competencies, cultural nuances, creativity, and contextual appropriateness. Further, concerns regarding data privacy, accountability, learner dependence, and the potential overreliance on automated feedback continue to shape discussions of responsible AI implementation in education (Kasneci et al., 2023; Cerratto Pargman et al., 2025). It follows that the future of AI-mediated assessment depends not only on technological advancement but also on the development of frameworks that ensure *fairness, trustworthiness, transparency,* and *pedagogical value*.

Against this backdrop, the present chapter examines the evolving role of AI in automated assessment and feedback within language education. Specifically, it explores four interconnected dimensions of AI-powered language assessment: *automated writing evaluation, speech scoring and pronunciation assessment, grammar and vocabulary diagnostics,* and *the overarching strengths and limitations of AI-generated feedback*. By synthesizing contemporary research and emerging practices, the chapter aims to provide a comprehensive overview of how AI is reshaping language assessment while critically examining the pedagogical, practical, and ethical

considerations associated with its implementation. So, the discussion highlights both the transformative potential of AI-driven assessment systems and the challenges that must be addressed to ensure their effective, equitable, and responsible integration into language teaching, learning, and assessment.

10.2 Automated Writing Evaluation

Automated Writing Evaluation (AWE) represents one of the earliest and most extensively developed applications of AI in language assessment. Writing has traditionally been among the most challenging language skills to assess because effective evaluation requires attention to multiple dimensions simultaneously, including grammatical accuracy, lexical sophistication, coherence, cohesion, content development, organization, and rhetorical effectiveness. Providing detailed feedback on these aspects demands substantial time and expertise from teachers, particularly in contexts involving large numbers of learners. Accordingly, the search for scalable approaches to writing assessment has long motivated technological innovation.

Early automated writing systems were largely rule-based and focused on identifying *surface-level linguistic errors*. These systems relied on predefined grammatical rules and dictionaries to *detect spelling mistakes, punctuation problems, and basic syntactic inaccuracies.* While useful as proofreading tools, they offered only limited pedagogical value because they could not adequately address higher-order dimensions of writing. Their feedback often consisted of simple error identification without meaningful explanations or guidance for revision.

Automated Writing Evaluation (AWE), which encompasses Automated Essay Scoring (AES) and Automated Short Answer Grading (ASAG), utilizes computational techniques to assess written responses and provide instant feedback (Rodrigues et al., 2025). While earlier systems relied heavily on manually engineered linguistic features and statistical classifiers, modern AWE is increasingly powered by pre-trained LLMs such as GPT-3.5 and GPT-4, which are capable of capturing deep semantic meaning and context (Rodrigues et al., 2025).

Advances in NLP and machine learning fundamentally altered this landscape. Modern AWE systems analyze writing at multiple levels, incorporating *lexical, syntactic, semantic, and discourse-related features*. Those AWE systems, employ sophisticated NLP algorithms, deep learning architectures, and LLMs capable of *analyzing discourse-level features, coherence, cohesion, argumentation quality*, and *rhetorical structure.* Rather than relying exclusively on handcrafted rules, contemporary systems learn patterns from large corpora of learner and expert writing. This shift has enabled AI technologies to move beyond error detection and toward more comprehensive evaluations of writing quality. As noted in systematic reviews of automated writing evaluation, modern AWE systems increasingly support both assessment and instruction by combining scoring mechanisms with formative feedback functions (Nunes et al., 2022).

The increasing availability of AI-powered writing evaluation systems reflects the growing need for *scalable assessment solutions* in educational settings. Teachers frequently face challenges in providing timely and detailed feedback to large numbers of students. AI systems address this challenge by delivering immediate responses that can support iterative writing development and self-regulated learning.

Recent research demonstrates that AI-powered writing assessment tools can effectively identify grammatical errors, lexical problems, syntactic complexity issues, and organizational weaknesses while simultaneously offering suggestions for revision and improvement (Wang et al., 2025). Such systems contribute to formative assessment by enabling students to revise their work multiple times before final submission.

These general claims are borne out, with variable strength, by a growing body of quasi-experimental research spanning multiple learner populations. In an 8-week quasi-experiment with 53 intermediate-to-advanced Chinese-as-a-Second-Language undergraduates, Huang and Chen (2026) found greater writing gains and self-efficacy from ChatGPT-4.0 feedback than from traditional teacher support, while Zheng et al. (2026), working with 124 L2 German learners, reported that a domain-specific LLM feedback system (Dr. Write) produced greater gains in writing self-efficacy and clause-level accuracy than either generic-LLM or teacher feedback, though without matching gains in affective engagement. Comparable benefits have been reported outside East Asia and Europe: Divya and Alamelu (2026)

found that DeepL Write and Claude AI significantly improved argumentative-writing scores and metacognitive awareness of argument structure among 40 first-year Indian ESL engineering students. The picture is not uniformly favorable to AI, however. Tang et al. (2026), analyzing 244 texts from 61 Chinese university students over eight weeks, found that teacher feedback produced stronger short-term writing gains while AI-assisted peer feedback produced a steadier, more sustained learning-curve effect—a genuine trade-off rather than a straightforward AI advantage. More strikingly, Thi and Nikolov (2023) found no significant syntactic complexity gains from teacher, automated, or combined feedback across 270 EFL students' texts over a 13-week semester, though effects varied by proficiency level—a null result that cautions against assuming any single feedback source reliably drives measurable linguistic gains.

10.2.1 Automated Essay Scoring

Automated Writing Evaluation (AWE) has evolved considerably from its early rule-based origins into sophisticated systems powered by machine learning, Natural Language Processing (NLP), and Large Language Models (LLMs). Contemporary AWE platforms are capable of analyzing multiple dimensions of writing quality, moving beyond the identification of surface-level linguistic errors to assess more complex features such as coherence, organization, argumentation, and communicative effectiveness. This evolution reflects broader developments in AI-assisted language assessment, where automated systems increasingly function not only as scoring mechanisms but also as tools that support learning through formative feedback and revision guidance (Wang et al., 2025).

One of the most significant developments in this area is the emergence of AI-based essay scoring systems capable of approximating human evaluative judgment. Recent research has explored the psychometric quality of LLMs as essay raters, employing advanced statistical frameworks such as Generalizability Theory (G-Theory) and Many-Facet Rasch Modeling (MFRM) to evaluate reliability, validity, and scoring consistency (Wang et al., 2025). These approaches allow researchers to partition multiple sources of variance within assessment data and identify the relative contributions of scoring dimensions, raters, and tasks. Through such analyzes, AI-based scoring systems can be examined not just for accuracy but also for their

ability to provide stable and meaningful evaluations across different assessment contexts.

Findings from recent studies suggest that LLMs demonstrate strong potential as writing assessors, particularly when analytical scoring rubrics are employed. *Analytical rubrics* require evaluators to score multiple dimensions of writing separately, including content development, organization, vocabulary use, grammatical accuracy, and coherence. Under these conditions, AI systems have demonstrated levels of reliability comparable to those of experienced human raters and, in some cases, have exhibited greater scoring consistency (Wang et al., 2025). Because AI systems are not affected by fatigue, attentional fluctuations, or emotional factors, they can apply evaluation criteria systematically across large numbers of essays, making them particularly attractive for large-scale assessment contexts.

One notable finding concerns *scoring convergence validity*. Studies indicate that AI raters perform most effectively when evaluating dimensions that require explicit textual evidence. For example, dimensions such as content development, organization, and linguistic accuracy are often associated with strong agreement between AI-generated and human-generated scores. Wang et al. (2025) reported that advanced LLMs achieved convergent validity coefficients approaching human-rater agreement levels, demonstrating their ability to identify and evaluate textual features consistently. Such findings suggest that AI systems may provide reliable support for large-scale writing assessment where consistency and scalability are essential.

At the same time, important differences remain between human and AI evaluative processes. *Human raters* often demonstrate greater flexibility when making holistic judgments because they can simultaneously consider multiple contextual and rhetorical dimensions of writing. Their evaluations may incorporate factors such as originality, audience awareness, creativity, and communicative impact. However, this flexibility comes at a cost. Human raters frequently experience what Wang et al. (2025) describe as the "*mental workload*" associated with evaluating complex analytical rubrics. Maintaining consistency across multiple scoring dimensions can be cognitively demanding, particularly when large numbers of essays must be assessed within limited timeframes.

AI systems, by contrast, excel in applying analytical rubrics systematically and consistently. Their evaluations are guided by learned linguistic patterns

and predefined scoring criteria rather than subjective impressions. Consequently, *AI raters* often achieve reliability coefficients comparable to or exceeding those of human evaluators for dimensions such as *ideas and content, organization,* and *language use* (Wang et al., 2025). This consistency represents one of the most significant advantages of automated essay scoring because it reduces variability attributable to rater-related factors and contributes to greater standardization across assessment contexts.

Nevertheless, AI-based scoring systems are *not immune to bias* or *scoring distortions*. One phenomenon observed in both human and AI assessment is the *halo effect*, whereby strong performance in one dimension disproportionately influences evaluations in unrelated areas. For example, a well-developed argument or sophisticated vocabulary may lead to inflated scores for organizational structure or coherence, even when those dimensions are less well developed. Wang et al. (2025) found evidence that both human and AI raters may be susceptible to such effects, highlighting the importance of ongoing validation and rubric refinement. These findings underscore the need to examine not only overall scoring accuracy but also the internal relationships among different scoring dimensions.

Concrete evidence bears out both sides of this picture. In a comparison of 186 introductory-psychology essays graded by eight human instructors and two versions of ChatGPT, Wetzler et al. (2024) found that the AI evaluators were systematically more lenient than human raters at the lower end of the performance range and systematically stricter at the upper end—a divergence pattern invisible in an aggregate correlation coefficient but consequential for individual students near a grading threshold. A comparable disagreement without a corresponding difference in average scores emerged in Guibangguibang and Hesula's (2026) study of ESL student essays rated by three human raters and three AI tools (ChatGPT, CiCi, and Bing): inter-rater consistency between the human and AI groups was low even though overall score levels did not differ significantly between them. Read together, these findings qualify the reliability claims above: AI raters can match human raters in the aggregate while still disagreeing with them, essay by essay, in ways a single summary statistic conceals.

Beyond reliability and validity considerations, AI-based essay scoring raises important questions regarding the nature of writing assessment itself. Writing proficiency is a multidimensional construct encompassing

linguistic, cognitive, rhetorical, and sociocultural competencies. While AI systems have become increasingly sophisticated in identifying textual patterns associated with quality writing, they may still struggle to evaluate nuanced aspects of communication such as creativity, irony, intercultural appropriateness, and disciplinary voice. On this basis, many scholars advocate *hybrid assessment models* in which AI systems complement rather than replace human evaluators. Within such models, AI can efficiently perform routine analytical tasks while human raters contribute contextual interpretation, pedagogical judgment, and qualitative insights that remain difficult to automate.

The growing adoption of AI-based essay scoring therefore reflects a broader transformation in language assessment. Rather than serving solely as automated grading mechanisms, contemporary AI systems increasingly function as components of integrated assessment ecosystems that support *reliability, scalability,* and *formative learning*. Their greatest potential may lie not in replacing human expertise but in augmenting it, creating assessment environments that combine the efficiency of automation with the interpretive strengths of human judgment. As AI technologies continue to evolve, ongoing research will be essential to ensure that automated scoring systems remain psychometrically sound, educationally meaningful, and ethically responsible.

These qualifications are reinforced by broader reviews of the field. Ramesh and Sanampudi's (2021) systematic review of automated essay scoring research concluded that most existing systems still fail to adequately evaluate content relevance and discourse coherence—precisely the dimensions bound up with what an essay is arguing rather than how it is worded. Reliability also varies by learner population rather than holding uniformly across contexts: Lim et al. (2023), working with a Korean learner corpus, found that a neural AWE system built on syntactic-complexity and fluency features achieved high reliability in predicting Korean L2 writing proficiency, evidence that scoring reliability extends, with appropriate retraining, beyond the English-dominant and Chinese-EFL contexts that dominate the literature. Barrot (2026) draws the pedagogical implication explicitly, arguing that generative-AI-based essay scoring risks diminishing teachers' evaluative agency unless institutions build in structured teacher mediation of AI-generated scores—an ecological safeguard against the bias and data-privacy risks that unmediated automated scoring can introduce.

10.2.2 Automated Feedback and Revision Support

Although automated scoring receives considerable attention, many scholars argue that the most valuable contribution of AI lies in its capacity to provide *formative feedback*. Effective feedback plays a crucial role in language development because it helps learners identify weaknesses, understand errors, and improve future performance. However, providing detailed individualized feedback can be extremely demanding for teachers, particularly in large classes.

AI systems address this challenge by generating feedback almost instantaneously. Rather than waiting days or weeks for teacher comments, learners can receive immediate responses that support revision while the writing task remains cognitively salient. Research on automated writing feedback indicates that such immediacy can enhance learner engagement and encourage iterative revision practices (Ranalli, 2021).

Contemporary AI-powered feedback systems operate on multiple levels simultaneously. At the *linguistic level*, they identify grammatical inaccuracies, lexical misuse, and syntactic problems. At the *discourse level*, they evaluate coherence, cohesion, and organizational structure. Some systems also provide recommendations related to genre conventions, audience awareness, and academic style. This *multidimensional feedback* enables learners to engage with writing as a complex communicative process rather than simply correcting isolated errors.

The *pedagogical significance* of automated feedback is particularly evident in its potential to promote learner autonomy. Because feedback is available on demand, learners can independently review, revise, and refine their texts without waiting for teacher intervention. This aligns with self-regulated learning frameworks that emphasize learner responsibility and metacognitive engagement. Rather than functioning solely as recipients of evaluation, learners become active participants in the revision process.

Research has also demonstrated that the effectiveness of automated feedback depends on learner engagement. Simply providing feedback does not guarantee improvement; learners must actively interpret, evaluate, and apply the information they receive. Ranalli (2021) emphasizes that successful integration of automated feedback requires *instructional support* that helps learners develop the skills necessary to use feedback effectively. Seen in

this light, educators continue to play a critical role in mediating interactions between learners and AI systems.

How much of this feedback learners actually take up is itself an empirical question, and the answer is uneven. Ziqi et al. (2024) tracked 45 L2 students revising argumentative essays with ChatGPT-generated feedback and found that they rejected or simply ignored roughly 38% of it overall—about 22% explicitly contested and 16% left unaddressed—with content-level suggestions dismissed significantly more often than form-level corrections, most commonly because students judged them inconvenient to act on or unhelpful. A comparable pattern appeared among 25 B1-level EFL students in Türkiye revising opinion essays: Cengiz et al. (2025) found that students incorporated most ChatGPT feedback on language use but engaged only inconsistently with feedback on essay organization, leaving more than a third of it uncorrected. Together these findings complicate the assumption, implicit in much of the literature on automated feedback, that availability of feedback translates directly into its uptake; as Ranalli (2021) already suggested, that gap is closed by instruction in how to use feedback, not by the feedback system alone.

10.2.3 Formative Assessment, Human–AI Collaboration, and AI-Mediated Feedback

One of the most transformative contributions of AI to language assessment lies in its capacity to support *formative assessment* through adaptive and personalized feedback. Unlike traditional assessment models, which often provide feedback only after task completion, AI-powered systems can deliver immediate responses that guide learners throughout the learning process. This capability is particularly valuable in writing instruction, where timely feedback can promote revision, reflection, and self-regulated learning. Adaptive feedback systems analyze learner performance in real time and generate individualized recommendations based on identified strengths and weaknesses, thereby cultivating continuous improvement rather than merely evaluating final products.

Recent studies have demonstrated that AI-generated adaptive feedback can positively influence learner engagement and writing quality. Learners who receive immediate, personalized feedback are more likely to spend time revising their work, addressing weaknesses, and improving subsequent performance. What is more, adaptive feedback supports the development of

learner autonomy by encouraging students to monitor their own progress and make informed decisions about revision strategies. As AI systems become increasingly sophisticated, feedback is evolving from a corrective mechanism into an interactive pedagogical tool that actively supports the writing process.

10.2.4 Assessing Human–AI Collaboration: The DRIVE Framework

The growing integration of generative AI into writing processes has prompted researchers to reconsider what should be assessed in AI-supported writing environments. Traditional writing assessment focuses primarily on the final written product, evaluating linguistic accuracy, organization, and content quality. However, as learners increasingly collaborate with AI systems during writing tasks, assessment must also account for how effectively students interact with and direct these technologies.

Oliveira et al. (2025) address this challenge through the development of the *DRIVE Framework*, a process-oriented assessment model designed to evaluate human–AI collaboration in writing. The framework introduces two key constructs: *Directive Reasoning Interaction (DRI) and Visible Expertise (VE).*

Directive Reasoning Interaction refers to the learner's ability to guide and manage AI-generated outputs through purposeful prompting. Rather than submitting simple requests for information, highly effective learners actively steer AI interactions by providing detailed instructions, challenging responses, requesting clarifications, and refining outputs through iterative dialogue. According to Oliveira et al. (2025), students who demonstrate strong DRI engage in high-level interactions that resemble collaborative problem-solving rather than passive information retrieval. This capacity to direct AI systems effectively has emerged as an important component of writing competence in AI-enhanced environments.

Visible Expertise, by contrast, refers to the extent to which learners contribute their own knowledge, reasoning, and disciplinary understanding during interactions with AI. Rather than relying entirely on machine-generated content, successful learners integrate external sources, evaluate AI outputs critically, and apply their own expertise to shape the final product. Oliveira et al. (2025) found that students who actively engaged with AI while

contributing substantial personal knowledge produced higher-quality essays than those who depended heavily on AI-generated responses. These findings suggest that future writing assessments may need to evaluate not only textual outcomes but also learner agency, critical thinking, and the quality of human–AI interaction.

The DRIVE framework therefore represents an important shift in assessment philosophy. It recognizes that writing competence in the age of generative AI involves not only linguistic proficiency but also the ability to collaborate effectively with intelligent systems. As a result, future assessment models may increasingly focus on how learners use AI as a cognitive tool rather than treating AI involvement as external to the assessment process.

Construct	Definition	High-performance indicators	Assessment evidence
Directive Reasoning Interaction (DRI)	Learner's ability to guide and manage AI outputs through purposeful prompting	Detailed instructions; challenging and refining outputs; iterative dialogue resembling collaborative problem-solving	Prompt logs; interaction transcripts; revision trajectories
Visible Expertise (VE)	Learner's contribution of own knowledge, reasoning, and disciplinary understanding during the interaction	Substantive content added beyond AI suggestions; corrections of AI errors; justified acceptance/rejection decisions	Annotated drafts; think-alouds; comparison of AI output and final text

Table 5. The DRIVE framework for assessing human–AI collaborative writing (Oliveira et al., 2025)

10.2.5 Reformulation and AI-Mediated Feedback Mechanisms

Another important development in AI-supported writing assessment concerns the use of reformulation-based feedback. Unlike traditional corrective feedback, which typically identifies errors and provides explanations, reformulation presents learners with revised versions of their own texts that preserve original meaning while improving linguistic quality. This approach enables learners to compare their production with target-like language and identify gaps in their interlanguage development.

Chen (2025) conceptualizes this approach as *Written Corrective Feedback (WCF)* supported by generative AI. Within this model, systems such as ChatGPT function as a "*more capable peer*," rewriting learner texts while maintaining the original communicative intent. Through comparison between the original and reformulated versions, learners can observe how vocabulary, syntax, cohesion, and stylistic features may be enhanced. Such comparisons encourage noticing, a process widely regarded as essential for second language development.

The effectiveness of reformulation-based feedback appears to depend substantially on prompt design. Research indicates that generic requests for correction often produce limited pedagogical benefits, whereas prompts specifying particular objectives—such as improving academic style, enhancing lexical sophistication, or increasing syntactic complexity—yield more useful revisions. In these cases, AI-generated reformulations can substantially improve sentence structure, vocabulary diversity, and overall text quality while preserving learner ownership of ideas (Chen, 2025).

10.2.6 Multimodal Assessment and Automated Picture-Cued Writing Evaluation

The development of multimodal large language models has expanded the scope of automated writing evaluation beyond text-only assessment. Emerging systems can now analyze visual and textual information simultaneously, creating opportunities for more authentic and cognitively engaging assessment tasks. One notable example is *Automated Picture-Cued Writing Evaluation (APWE)*, which requires learners to produce written responses based on visual stimuli.

Unlike conventional essay prompts, picture-cued tasks encourage learners to engage in descriptive, interpretive, and inferential writing while drawing upon visual information. Multimodal LLMs can evaluate both the linguistic quality of learner responses and the degree to which written content corresponds accurately to the visual prompt. According to Zhuang et al. (2025), APWE systems support semantic retention and reduce cognitive load by encouraging learners to connect language production with visual comprehension. Furthermore, these systems can *provide highly specific feedback regarding content relevance, descriptive accuracy,* and *organizational effectiveness*.

The integration of multimodal assessment reflects broader trends in contemporary communication, where meaning is increasingly conveyed through combinations of textual, visual, and digital resources. As educational environments continue to embrace multimodal literacy, AI-powered assessment systems are likely to play an increasingly important role in evaluating learners' ability to communicate effectively across multiple modes of representation.

Multimodal learning analytics systems can analyze speech patterns, gesture use, interaction behaviors, engagement levels, collaboration patterns, and other indicators of learning (Mohammadi et al., 2025). Such capabilities may allow educators to develop richer understandings of learner performance.

Concrete deployments illustrate both the appeal and the friction of this approach. A two-year study following 399 students and 17 teachers who used a multimodal learning-analytics tool during collaborative healthcare-education debriefings found participants rated the tool as useful, but also flagged its complexity and raised trust concerns about how its inferences were generated—concerns that transfer directly to language classrooms, where inferring engagement or confusion from speech and gesture data carries similar interpretive stakes (Yan et al., 2024). These reservations are compounded by a limitation in the research base itself: a review of thirty-five studies that explicitly used learning theory to interpret multimodal learning-analytics data found embodied cognition, cognitive load theory, and control-value theory account for most of the theoretical grounding attempted, but that theoretical engagement across the field generally remains limited—meaning many MMLA systems generate rich behavioral signals without a correspondingly rigorous account of what those signals mean pedagogically (Giannakos & Cukurova, 2023). Multimodal assessment in language education will need to close this gap between data richness and interpretive validity before its outputs can count as evidence of communicative competence rather than mere correlates of it.

For example, future speaking assessments may evaluate not only linguistic accuracy but also interactional competence, pragmatic appropriateness, non-verbal communication, and communicative effectiveness. Similarly, collaborative tasks may be assessed through analysis of interaction processes rather than solely final outcomes.

These developments align with broader educational movements toward authentic assessment and competency-based evaluation. Rather than measuring isolated language components, future assessment systems may focus increasingly on learners' ability to communicate effectively across diverse contexts and modalities.

Nevertheless, such developments also raise important ethical concerns regarding privacy, surveillance, and data governance. The collection of multimodal learner data requires robust safeguards to ensure that technological innovation remains aligned with ethical principles and learner rights.

10.2.7 Generative AI and the Transformation Toward Process-Oriented Writing Assessment

Collectively, these developments suggest that automated writing evaluation is moving beyond its original focus on scoring and error detection. Contemporary AI systems increasingly support *formative assessment, process-oriented evaluation, human–AI collaboration,* and *multimodal communication*. Rather than functioning solely as automated raters, AI technologies are becoming integral components of learning environments that *promote revision, reflection,* and *learner agency*.

The emergence of frameworks such as DRIVE, the growing use of reformulation-based feedback, and the development of multimodal assessment systems indicate that future conceptions of writing competence may extend beyond traditional notions of linguistic accuracy and textual quality. In AI-rich environments, successful writers may increasingly be those who can engage critically with intelligent systems, integrate external knowledge effectively, and leverage AI as a tool for communication and learning. In turn, automated writing evaluation is evolving from a technology of assessment into a technology for learning, fundamentally reshaping the relationship between writing, feedback, and educational evaluation.

The emergence of generative AI has introduced new possibilities for writing assessment and feedback. Unlike traditional AWE systems, which primarily evaluate existing texts, generative AI can engage learners in dynamic interactions throughout the writing process. These systems can explain errors, answer questions, provide examples, and suggest alternative formulations in ways that resemble human tutoring.

This capability has led some scholars to describe generative AI as a form of intelligent writing support rather than merely an assessment tool. Within such environments, assessment becomes integrated into learning activities rather than functioning as a separate evaluative event. Learners receive continuous guidance as they write, revise, and reflect on their work.

However, generative AI also introduces new challenges. The ability of AI systems to generate complete essays raises concerns regarding academic integrity and the authenticity of assessment tasks. Educators must therefore reconsider what constitutes evidence of language proficiency in contexts where AI-assisted writing is readily available. Rather than attempting to eliminate AI from educational environments, some researchers advocate redesigning assessment practices to evaluate learners' ability to use AI critically, ethically, and effectively.

This shift suggests a broader transformation in the nature of writing competence. In the AI era, successful writers may increasingly be those who can collaborate productively with intelligent technologies while maintaining ownership of ideas, exercising critical judgment, and demonstrating communicative effectiveness. As language assessment continues to evolve, automated writing evaluation is likely to remain at the forefront of discussions regarding the relationship between human expertise and artificial intelligence in education.

10.3 Speech Scoring and Pronunciation Assessment

The *assessment of speaking* has traditionally been one of the most complex and resource-intensive components of language evaluation. Unlike writing assessment, which allows evaluators to revisit and analyze a permanent textual product, spoken language is *dynamic, multidimensional,* and *highly contextualized*. Effective speaking assessment requires consideration of numerous linguistic and paralinguistic features, including pronunciation, fluency, grammatical accuracy, lexical range, discourse management, interactional competence, and pragmatic appropriateness. For this reason, speaking assessment has historically relied heavily on trained human raters whose judgments are informed by professional expertise and contextual understanding.

The emergence of AI-powered speech assessment has been facilitated by the integration of high-fidelity ASR systems and LLM technologies (Ben-Zion et al., 2025; Xiao et al., 2025). Contemporary frameworks frequently utilize advanced speech recognition models such as OpenAI's Whisper to transcribe spoken language into machine-readable text before performing linguistic analyzes. These systems can evaluate multiple dimensions of oral performance, including speech rate, lexical diversity, grammatical complexity, pronunciation accuracy, discourse coherence, and communicative effectiveness. By automating many aspects of speech evaluation, AI technologies enable more frequent and accessible assessment opportunities while reducing the logistical burdens traditionally associated with oral examinations.

However, human-based speaking assessment presents several challenges. Large-scale implementation is costly and time-consuming, rater training requires significant resources, and scoring consistency can be difficult to maintain across multiple evaluators. These challenges have motivated the development of AI-powered systems capable of automating aspects of speech assessment while maintaining acceptable levels of reliability and validity. Advances in Automatic Speech Recognition (ASR), machine learning, and deep neural networks have enabled substantial progress in this area, making speech scoring and pronunciation assessment one of the fastest-growing domains within AI-assisted language assessment (Jaleniauskienė et al., 2023).

10.3.1 Automatic Speech Recognition and Speaking Assessment

The foundation of most contemporary AI-based speaking assessment systems is *Automatic Speech Recognition technology*. ASR systems convert spoken language into machine-readable text, creating a basis for subsequent linguistic analysis. Once speech is transcribed, AI models analyze linguistic and acoustic features to generate evaluations of learner performance. Recent research suggests that modern ASR systems can achieve high levels of accuracy in decoding spoken language and can support reliable assessment of oral proficiency across a variety of educational contexts. Early speech recognition technologies were often limited by high error rates, particularly when processing non-native speech. Recent advances in deep learning, however, have dramatically improved recognition accuracy, enabling more sophisticated assessment applications. AI models are able to

evaluate lexical diversity, grammatical complexity, speech rate, syntactic accuracy, discourse organization, and fluency indicators. This capability allows assessment systems to move beyond simple pronunciation checks and evaluate broader aspects of communicative competence.

One particularly promising application involves reading fluency assessment. Reading fluency encompasses multiple dimensions, including decoding accuracy, reading speed, prosodic features, and overall oral expressiveness. Research indicates that speech rate remains one of the strongest predictors of perceived fluency, while prosodic features such as intonation, stress, and pausing contribute significantly to overall communicative effectiveness. Studies employing Whisper and related technologies have demonstrated strong performance in adult speech recognition and have enabled automated evaluations of reading fluency with levels of reliability approaching those of human assessors.

Methodologically, this reliability claim rests on moving beyond words-per-minute as the sole fluency metric. Soto (2026) proposes a *multilayer framework* that decomposes oral reading fluency into interpretable indicators—*rate, temporal continuity, pausing, and rhythm*—arguing that scalable, formative literacy assessment requires this fuller decomposition rather than a single rate-based score, since two readers with identical words-per-minute counts can differ substantially in comprehensibility. The methodological implication is that gains in automated speech scoring have come as much from richer construct decomposition as from larger training corpora: a system that reports a single global fluency figure conceals precisely the distinctions a teacher would act on.

The integration of AI into fluency assessment has also facilitated the development of *multimodal evaluation models*. Rather than relying exclusively on textual transcripts, contemporary systems combine acoustic features derived from speech signals with linguistic features extracted from transcribed language. By incorporating information related to speech rate, pause frequency, pronunciation accuracy, lexical diversity, and grammatical complexity, these systems generate more comprehensive representations of speaking proficiency. Such approaches reflect growing recognition that oral communication is a multidimensional construct requiring assessment across multiple linguistic and paralinguistic domains.

One significant advantage of AI-based speech assessment is *scalability*. Whereas traditional oral examinations require substantial human resources, AI systems can evaluate large numbers of speaking samples quickly and consistently. This capability is particularly valuable in online learning environments, standardized testing programs, and institutions serving large student populations. Automated assessment also facilitates more frequent evaluation, allowing learners to receive continuous feedback rather than waiting for formal examination opportunities.

Despite these advantages, ASR-based assessment faces important limitations. *Recognition accuracy* can vary considerably depending on *accent, speech quality, background noise,* and *learner proficiency.* Non-native speech often contains pronunciation patterns that differ substantially from native-speaker norms, creating challenges for systems trained primarily on standard language varieties. As a result, researchers continue to emphasize the importance of validating speech assessment systems across diverse learner populations to ensure fairness and inclusivity.

This caution is well founded empirically, and the relevant population differences begin with age. Attia et al. (2024) document a persistent performance gap in automatic speech recognition between children's and adults' speech, motivating child-adapted models such as Kid-Whisper built specifically to close it—a reminder that ASR systems trained predominantly on adult speech cannot be assumed to transfer to younger learners, whose articulation and disfluency patterns differ systematically from the adult data most models are trained on. Age-related variation reappears further along the developmental range: Mawalim et al. (2025) extended structured, multimodal modeling of speaking-skill indices specifically to young adolescents, treating this population as one that requires dedicated assessment models rather than downward extrapolation from adult- or child-calibrated systems. Geographic and linguistic context matters as well: LLM-based scoring has been adapted to typologically distinct target languages such as Finnish and Finland Swedish (Voskoboinik et al., 2025), underscoring that accuracy figures reported for English-medium systems do not automatically generalize elsewhere.

The bibliometric analysis conducted by Jaleniauskienė et al. (2023) identified automatic speech recognition as one of the most rapidly expanding areas within AI-assisted language education. The growing popularity of ASR

reflects its potential to provide immediate and individualized feedback to language learners while reducing assessment costs and increasing accessibility.

10.3.2 Pronunciation Assessment

Among the various applications of AI in speaking evaluation, *pronunciation assessment* has received particularly extensive attention. Pronunciation plays a crucial role in communicative effectiveness because it directly influences intelligibility and comprehensibility. Traditionally, pronunciation assessment has depended on teacher judgment, making it difficult to provide learners with extensive individualized practice opportunities. AI technologies have transformed this landscape by enabling continuous pronunciation evaluation accompanied by immediate feedback.

Contemporary pronunciation assessment systems analyze both segmental and suprasegmental features of speech. Segmental features refer to individual phonemes and their accurate production, while suprasegmental features include stress, rhythm, intonation, pausing, and speech timing. Research in second language acquisition consistently demonstrates that successful communication depends not only on the accurate production of individual sounds but also on the effective use of prosodic features that contribute to natural and comprehensible speech.

AI-powered pronunciation tools compare learner speech against target-language models and identify areas requiring improvement. Feedback may include visual representations of pronunciation patterns, suggestions for articulatory adjustments, and opportunities for repeated practice. Such systems enable learners to engage in extensive pronunciation training independently, reducing reliance on teacher availability and classroom time.

Recent advances in multimodal AI have further enhanced pronunciation assessment by integrating acoustic and linguistic data within unified evaluation frameworks. Research has demonstrated that multimodal models combining speech features with language-based representations can predict speaking proficiency with considerable accuracy, particularly when evaluating dimensions such as pronunciation, fluency, lexical competence, and coherence. These developments suggest that pronunciation assessment is increasingly moving beyond isolated sound production toward a more *holistic evaluation of spoken communication* (Xiao et. al., 2025)

A notable advantage of AI-mediated pronunciation assessment is its capacity to provide immediate feedback. Learners can practice repeatedly in low-stakes environments while receiving real-time information regarding pronunciation errors, stress patterns, and fluency issues. Such opportunities are particularly valuable because they support extensive oral practice without requiring constant teacher intervention. Moreover, repeated interaction with AI systems may reduce speaking anxiety by providing a non-judgmental environment in which learners can experiment, make mistakes, and improve through practice.

An important pedagogical advantage of AI-mediated pronunciation assessment lies in its capacity to reduce learner anxiety. Many language learners experience apprehension when speaking in front of teachers or peers, particularly when they fear negative evaluation. AI systems provide a non-judgmental environment in which learners can practice repeatedly without social pressure. This increased willingness to engage in speaking activities can contribute positively to oral language development, particularly among learners who are reluctant to participate in traditional classroom interactions.

Nevertheless, pronunciation assessment remains one of the most debated areas of AI-mediated evaluation. Questions persist regarding the appropriateness of native-speaker norms as assessment benchmarks and the extent to which pronunciation assessment should prioritize intelligibility rather than accent reduction. Critics argue that some AI systems may inadvertently reinforce unrealistic native-speaker standards that fail to reflect the multilingual realities of contemporary communication. It follows that many researchers advocate assessment models that emphasize communicative effectiveness and comprehensibility rather than strict conformity to native-speaker pronunciation patterns.

10.3.3 Conversational Agents and Interactive Oral Assessment

The emergence of conversational AI has expanded the possibilities of speaking assessment beyond isolated speech samples. Rather than merely evaluating pronunciation or fluency, conversational agents can engage learners in interactive dialogues that more closely resemble authentic communication. These systems create opportunities for dynamic assessment in which learner responses influence the direction and complexity of interaction.

Recent research demonstrates the potential of LLM-enhanced conversational agents to support both language learning and assessment. Xiao et al. (2025) investigated an LLM-enhanced bilingual conversational agent designed to facilitate dialogic reading among English as a Foreign Language (EFL) children. The system combined bilingual scaffolding, adaptive questioning, and interactive feedback to support oral language development while simultaneously generating valuable assessment data. The findings suggest that conversational agents can create highly engaging learning environments while providing insights into learner comprehension, vocabulary development, and speaking performance.

A notable example is the use of LLM-enhanced conversational agents in dialogic reading activities. Xiao et al. (2025) describe the use of conversational AI systems designed to support oral language development through interactive reading experiences. These systems provide a low-anxiety environment in which learners can engage in meaningful dialogue while simultaneously generating assessment data regarding comprehension, vocabulary use, pronunciation, and communicative competence.

One of the most significant advantages of conversational agents lies in their ability to provide *adaptive feedback*. These systems often employ instructional frameworks such as *CROWD (Completion, Recall, Open-ended, Wh-questions, Distancing)* and *PEER (Prompt, Evaluate, Expand, Repeat)* to scaffold learner responses and support language development. Through these mechanisms, AI systems can adjust interactional complexity according to learner proficiency, ensuring that assessment remains both supportive and appropriately challenging.

A particularly important feature of conversational agents is their ability to provide adaptive support. Traditional oral assessments often present identical tasks to all learners regardless of proficiency level. Conversational AI systems, by contrast, can modify question complexity, adjust interactional support, and provide scaffolding in response to learner performance. This adaptability allows assessment to capture a broader range of abilities while maintaining learner engagement.

The use of bilingual conversational agents further highlights the potential of AI to support inclusive assessment practices. By allowing learners to draw strategically upon their first language while engaging with target-language tasks, such systems may provide a more accurate representation of

communicative competence. This approach aligns with contemporary perspectives on multilingualism that recognize language use as flexible and dynamic rather than strictly compartmentalized.

Traditional language assessments often require learners to operate exclusively within the target language, potentially obscuring underlying conceptual understanding when linguistic limitations interfere with communication. Conversational AI systems increasingly address this challenge by allowing strategic use of learners' home languages during assessment activities.

Xiao et al. (2025) found that bilingual conversational agents supported meaning-making processes by enabling learners to draw upon existing linguistic resources while engaging with target-language content. Such approaches align with contemporary perspectives on multilingualism that recognize language use as flexible and dynamic rather than compartmentalized. By reducing linguistic barriers, *bilingual assessment systems* may provide more accurate representations of learner competence and facilitate more equitable assessment practices.

Obviously, bilingual scaffolding also contributes to learner confidence and engagement. When learners are permitted to access their first language strategically, they may participate more actively in communicative tasks and demonstrate a broader range of knowledge and skills. Accordingly, bilingual AI systems offer promising possibilities for creating more *inclusive assessment environments* that recognize linguistic diversity as an educational resource rather than a deficit.

However, conversational assessment also introduces new challenges. Evaluating interactional competence requires analysis of complex communicative behaviors, including *turn-taking, clarification requests, negotiation of meaning, and pragmatic appropriateness*. While AI systems have become increasingly capable of processing linguistic content, assessing these interactional dimensions remains considerably more difficult. Consequently, ongoing research seeks to improve the ability of conversational agents to evaluate authentic communication while maintaining psychometric rigor.

Direct comparisons between human- and AI-delivered oral tasks temper some of this optimism. Eguchi et al. (2025) undertook exactly such a comparison, eliciting L2 English roleplay performances from 75 Japanese university students under both a human-tutor condition and an autonomous-

turn-taking AI-agent condition and subjecting 24 of the resulting co-constructed interactions to detailed conversation-analytic comparison—the kind of interlocutor-type check that assessment frameworks assuming interactional competence is elicitation-neutral cannot otherwise accommodate. A more direct test of AI-mediated dynamic assessment produced an outright null result: in a pre-post study of 125 EFL undergraduates (87 analyzed), Saeedi and Soltani (2025) found that an AI chatbot providing dynamic assessment of speech acts did not outperform either classroom-based dynamic assessment or simple direct feedback, and that classroom dynamic assessment in turn outperformed direct feedback on both immediate and delayed post-tests—a study for which the journal subsequently issued a formal correction, though the comparative pattern it reports remains among the clearest counter-evidence here to the assumption that conversational AI automatically improves on established, human-mediated assessment approaches.

10.3.4 Multidimensional Speaking Assessment

One of the most significant recent developments in AI-powered speech evaluation is the movement toward multidimensional assessment models. Traditional automated systems often focused on isolated features such as pronunciation accuracy or speech rate. Contemporary approaches increasingly recognize that speaking proficiency is a multifaceted construct requiring evaluation across multiple dimensions simultaneously.

Modern assessment systems integrate *acoustic, linguistic, and discourse-level features* to generate comprehensive proficiency profiles. Acoustic features may include speech rate, pause frequency, pronunciation accuracy, and prosodic variation. Linguistic features encompass grammatical complexity, lexical sophistication, and syntactic accuracy. Discourse-level features examine coherence, cohesion, topic development, and communicative effectiveness. The integration of multiple data sources reflects broader trends toward *multimodal assessment*. Emerging technologies are beginning to incorporate visual information such as facial expressions, eye gaze, and gesture use alongside linguistic analysis. Although these approaches remain in relatively early stages of development, they offer the potential to capture dimensions of communication that are often overlooked in traditional assessment frameworks.

At the same time, the increasing complexity of multidimensional assessment raises important questions regarding transparency and interpretability. As AI systems integrate larger numbers of variables into evaluation processes, understanding how assessment decisions are generated becomes increasingly challenging. This concern aligns with broader discussions surrounding explainable AI in education and highlights the need for systems that provide clear and understandable justifications for their evaluations.

10.3.5 Reliability, Validity, and Fairness in Speech Assessment

The expansion of AI-mediated speech assessment has intensified discussions regarding psychometric quality. *Reliability* remains one of the primary strengths of automated systems because AI models can apply evaluation criteria consistently across large numbers of responses. Unlike human raters, AI systems do not experience fatigue or variability in attention. On this basis, automated assessment often produces highly stable results.

Validity, however, remains more complex. Speaking proficiency encompasses a wide range of cognitive, linguistic, social, and pragmatic competencies. Although AI systems excel at measuring observable linguistic features, they may struggle to capture deeper aspects of communicative competence. Features such as humor, persuasion, cultural appropriateness, and interpersonal sensitivity remain difficult to operationalize computationally. Goh and Aryadoust (2025) reach a consonant conclusion in their review of generative AI's role in foreign language listening and speaking assessment. Surveying the capabilities of spoken-dialogue systems and intelligent personal assistants alongside their current limitations, they conclude that closing the gap between measuring observable speech features and assessing genuinely communicative competence remains one of the field's open problems rather than a solved one, and identify this gap—rather than raw transcription accuracy—as the primary constraint on treating AI speech scores as a full substitute for human judgment.

Fairness presents an equally important concern. Because AI systems learn from training data, they may reproduce biases associated with particular language varieties, accents, or demographic groups. Researchers in responsible AI emphasize that assessment technologies must be evaluated continuously to ensure equitable treatment of diverse learner populations (Cerratto Pargman et al., 2025). Transparency, accountability, and ongoing validation are therefore essential components of ethical AI implementation.

Despite these challenges, speech scoring and pronunciation assessment illustrate the substantial potential of AI to enhance language evaluation. By providing scalable, immediate, and adaptive assessment opportunities, AI systems expand access to meaningful speaking practice and feedback. However, their long-term success will depend on the development of assessment frameworks that balance technological efficiency with pedagogical validity, inclusivity, and human oversight.

10.4 Grammar and Vocabulary Diagnostics

Grammar and vocabulary have traditionally been regarded as foundational components of language proficiency and, consequently, have occupied a prominent position in language assessment. While communicative approaches to language teaching have shifted attention toward meaningful language use, grammatical accuracy and lexical competence continue to serve as important indicators of learners' overall proficiency. The emergence of AI has markedly transformed how these dimensions are assessed, moving diagnostic practices beyond conventional testing formats toward dynamic, continuous, and personalized evaluation systems.

Traditional grammar and vocabulary assessments often rely on discrete-item tests, multiple-choice questions, cloze exercises, and teacher-generated evaluations. Although such instruments can provide useful information regarding learner performance, they frequently offer only limited insight into how learners use language in authentic communicative contexts. AI-powered diagnostic systems address this limitation by analyzing language production as it occurs naturally in written and spoken communication. Through NLP, machine learning, and large-scale language modeling, these systems can identify linguistic patterns, diagnose weaknesses, and provide targeted feedback with a level of immediacy that is difficult to achieve through conventional assessment methods.

The increasing integration of AI into language education has expanded the role of diagnostics from periodic measurement to continuous monitoring. Rather than assessing learners only at predetermined points during instruction, AI systems can track performance over time, identify developmental trajectories, and adapt feedback according to individual needs. This shift aligns closely with contemporary perspectives on formative assessment,

which emphasize the ongoing use of assessment information to support learning and instructional decision-making.

10.4.1 AI-Powered Grammar Diagnostics

Grammar diagnostics represents one of the earliest and most widespread applications of AI in language learning and assessment. Initially, grammar-checking software focused primarily on identifying surface-level errors such as spelling mistakes, punctuation problems, and subject–verb agreement violations. Modern AI systems, however, operate at a far more sophisticated level. They can analyze complex syntactic structures, identify patterns of recurring errors, and provide context-sensitive explanations that support learner understanding. The evolution of grammar diagnostics has been driven largely by advances in NLP and large language modeling. Contemporary systems do not simply compare learner language against predefined grammatical rules; instead, they draw upon extensive language corpora and probabilistic models to evaluate language use within context. This capability enables them to identify not only whether a particular form is grammatically acceptable but also whether it is appropriate for a given communicative situation.

Research reviewed within AI-assisted language learning literature suggests that modern grammar correction systems contribute substantially to learner development by promoting noticing and metalinguistic awareness. Rather than merely correcting errors, effective systems explain grammatical principles, provide examples of appropriate usage, and encourage learners to reflect on their linguistic choices. This pedagogical orientation is particularly important because language acquisition involves more than error elimination; it requires learners to develop an understanding of how grammatical forms function within meaningful communication.

Another notable development involves the integration of generative AI into grammar feedback. Large language models can engage learners in interactive discussions about grammatical issues, answer questions, and provide multiple explanations tailored to different proficiency levels. Such interactions move beyond traditional corrective feedback by creating opportunities for dialogue and exploration. Learners are no longer passive recipients of corrections but active participants in discussions about language use.

At the same time, AI systems can analyze extensive amounts of learner language, allowing for more comprehensive evaluations than traditional classroom-based assessments. Recent studies suggest that LLMs can identify specific grammar and vocabulary patterns associated with different proficiency levels and can classify learner output according to established frameworks such as the Common European Framework of Reference for Languages (CEFR), thereby supporting more precise diagnostic assessment (Benedetto et al., 2025).

One of the most prominent applications of AI in language assessment involves grammar diagnostics. Traditional grammar-checking tools primarily focused on identifying surface-level errors such as spelling mistakes, punctuation problems, and basic syntactic violations. Contemporary AI systems, however, operate at a much deeper level, incorporating contextual and semantic information into their analyzes. As a result, these systems can evaluate grammatical performance more holistically and provide explanations that support learner understanding rather than simply correcting errors.

Recent innovations have introduced *Generative Adversarial Networks (GANs)* into academic English correction systems, enabling more sophisticated analyses of authentic learner writing (Gao, 2025). GAN-based models employ data augmentation techniques, such as back-translation, to generate "pseudo-errors" that expand training datasets and improve the system's ability to recognize diverse grammatical patterns. Through this process, AI systems become increasingly capable of identifying complex grammatical problems while maintaining sensitivity to contextual variation.

Another significant development involves the integration of *Reinforcement Learning (RL)* with dependency parsing techniques. These approaches allow AI systems to evaluate grammatical structures more accurately and generate feedback that improves both textual coherence and semantic consistency (Gao, 2025). Rather than focusing exclusively on isolated linguistic forms, such systems assess how grammatical choices contribute to overall communicative effectiveness. Seen in this light, grammar diagnostics are increasingly aligned with contemporary views of grammar as a functional component of language use rather than a collection of decontextualized rules.

In addition, AI-powered systems can provide educators with detailed diagnostic information regarding recurring learner difficulties. Such analytics enable teachers to identify patterns across individuals and groups, thereby supporting targeted instructional interventions and more effective curriculum planning.

However, grammar diagnostics also reveal some of the limitations of AI systems. Because language is inherently variable and context-dependent, distinguishing between errors and acceptable variation can be challenging. AI systems may occasionally misclassify innovative or contextually appropriate language forms as errors, particularly when operating outside the linguistic patterns represented in their training data. As a result, human oversight remains important, especially in contexts involving advanced language use, creative writing, or multilingual communication.

10.4.2 Vocabulary Diagnostics and Lexical Assessment

Vocabulary knowledge is widely recognized as one of the strongest predictors of language proficiency. Learners with extensive vocabularies generally demonstrate stronger reading comprehension, more effective writing, and greater communicative flexibility. As a result, vocabulary assessment plays a crucial role in language education. AI technologies have significantly enhanced the sophistication of lexical diagnostics by enabling the analysis of vocabulary use in authentic communicative contexts rather than relying exclusively on traditional vocabulary tests.

Contemporary AI systems evaluate vocabulary across multiple dimensions. These include vocabulary size, lexical diversity, lexical sophistication, collocational competence, semantic accuracy, and contextual appropriateness. Through corpus-based analysis and machine learning algorithms, AI can identify patterns in learner language that reveal strengths and weaknesses not easily detectable through conventional assessment methods.

One particularly important contribution of AI is its ability to distinguish between receptive and productive vocabulary knowledge. Traditional vocabulary assessments often focus on recognition, measuring whether learners understand the meaning of particular words. AI-based diagnostics, by contrast, can analyze how learners actually use vocabulary in writing and speaking. This provides a richer picture of lexical competence because

productive vocabulary use reflects learners' ability to select, adapt, and integrate words into meaningful communication.

Recent research has highlighted the potential of *AI-generated storytelling environments* as innovative tools for vocabulary diagnostics (Namaziandost & Çakmak, 2025). Unlike traditional vocabulary tests, which often isolate words from meaningful contexts, AI-driven narrative environments assess lexical knowledge dynamically as learners engage with evolving storylines. Within these environments, learners make decisions that influence narrative progression, and the AI system continuously evaluates their vocabulary use, comprehension, and retention.

One important feature of these environments is the creation of *contextual diagnostic loops*, in which assessment occurs naturally through interaction with the narrative (Namaziandost & Çakmak, 2025). By analyzing learners' choices and responses, the AI system can identify vocabulary strengths and weaknesses in real time. This approach allows assessment to become an integrated component of the learning experience rather than a separate testing event.

Another significant advantage is *adaptive lexical complexity*. As learners demonstrate mastery of particular vocabulary items, AI systems can adjust the complexity of language presented within the narrative, ensuring that learners continue to encounter appropriately challenging input. This adaptive process aligns closely with Vygotsky's concept of the *Zone of Proximal Development*, allowing assessment and instruction to function simultaneously as components of a dynamic learning environment (Namaziandost & Çakmak, 2025).

Bibliometric analyses of AI in language education indicate that vocabulary learning remains one of the most extensively researched areas within AI-assisted language learning (Jaleniauskienė et al., 2023). Many AI-powered learning platforms incorporate adaptive vocabulary diagnostics that continuously monitor learner performance and adjust instructional content accordingly. As learners demonstrate mastery of particular lexical items, systems introduce increasingly challenging vocabulary, thereby supporting incremental language development.

Generative AI has further expanded the possibilities of vocabulary diagnostics. Large language models can evaluate lexical choice within extended discourse, identify inappropriate word usage, suggest more precise alternatives,

and provide contextualized explanations. Such capabilities support vocabulary development not simply by increasing the number of words learners know but by improving their ability to use language effectively and appropriately.

10.4.3 Personalized Diagnostic Assessment

Perhaps the most transformative contribution of AI to grammar and vocabulary assessment lies in its capacity to support personalized diagnostics. Traditional assessments typically apply the same tasks and evaluation criteria to all learners, regardless of individual strengths, weaknesses, or learning trajectories. AI systems, by contrast, can continuously analyze learner data and generate individualized profiles that inform both assessment and instruction.

Personalized diagnostics operate through the collection and interpretation of large quantities of learner performance data. These data may include writing samples, speaking recordings, interaction histories, revision patterns, response times, and assessment results. Machine learning algorithms identify recurring patterns and use predictive models to anticipate future learning needs. As a result, assessment becomes more than merely a retrospective evaluation of performance but a forward-looking process that guides subsequent learning activities. Such systems align closely with adaptive learning paradigms. When learners struggle with particular grammatical structures or vocabulary categories, AI systems can recommend targeted exercises, supplementary explanations, and additional practice opportunities. Conversely, learners who demonstrate mastery can progress to more advanced tasks without unnecessary repetition. This adaptive approach enhances efficiency by ensuring that instructional resources are directed toward areas where they are most needed.

Empirical support for this adaptive logic comes from a pretest-posttest experiment with 484 EFL students assigned to three AI-assisted-learning-strategy conditions—adaptive feedback, speech-recognition exercises, and intelligent tutoring—plus a control group: Feng (2025) found that the AI-assisted conditions reduced learners' cognitive load and improved learning outcomes relative to the control condition, lending empirical weight to the claim that recommending targeted practice in response to diagnosed weaknesses, rather than delivering uniform instruction to every learner,

measurably changes both how hard learners have to work and how much they learn.

The value of personalized diagnostics is particularly evident in diverse learning environments where students possess varying proficiency levels and learning goals. In traditional classrooms, providing individualized diagnostic feedback to every learner can be extremely demanding. AI systems help address this challenge by automating many aspects of data analysis while enabling teachers to focus on interpretation and pedagogical support.

In this respect, personalized diagnostics do not eliminate the role of educators. Rather, they provide teachers with richer information about learner performance, enabling more informed instructional decisions. AI-generated diagnostic reports can highlight patterns that might otherwise remain unnoticed, thereby supporting evidence-based teaching practices.

10.4.4 Diagnostic Assessment as Continuous Learning Support

The integration of AI into grammar and vocabulary diagnostics reflects a broader reconceptualization of assessment. Rather than functioning as isolated measurement events, diagnostics increasingly operate as continuous learning support systems. Feedback is delivered immediately, learning trajectories are monitored dynamically, and instructional interventions are adjusted in response to emerging needs.

This shift aligns with contemporary understandings of formative assessment, which emphasize the importance of timely feedback, learner engagement, and instructional responsiveness. AI systems facilitate these processes by providing continuous streams of information that support both learners and educators. As a result, assessment becomes embedded within everyday learning activities rather than occurring only at designated testing points.

At the same time, researchers caution against viewing AI diagnostics as infallible. The quality of diagnostic feedback depends on the quality of underlying data, the validity of assessment models, and the appropriateness of interpretation. In turn, responsible implementation requires ongoing evaluation, transparency, and human oversight. Diagnostic information generated by AI should inform educational decision-making rather than replace professional judgment.

Thus, AI-powered grammar and vocabulary diagnostics illustrate how intelligent technologies can enhance language assessment by providing detailed, personalized, and actionable information. Their greatest value lies not merely in identifying errors or gaps in knowledge but in supporting continuous language development through targeted and adaptive feedback.

10.4.5 Automated Generation of Diagnostic Items

A related but distinct application of AI to grammar and vocabulary assessment concerns the generation, rather than the scoring, of diagnostic items themselves. Because discrete-item tests remain a common component of diagnostic batteries even as continuous monitoring expands, the labor of writing, piloting, and revising individual grammar and vocabulary items has long constrained how frequently such tests can be refreshed, and generative AI has begun to address this constraint directly.

Lee et al. (2024) developed a ChatGPT-based automatic question generation system for English education using a prompt-engineering protocol and found that the resulting items achieved statistically significant validity ratings overall, although certain question types proved less suited to ChatGPT's generation capabilities than others. A complementary case study by Kic-Drgas and Kılıçkaya (2024) followed a single teacher using AI to generate foreign-language exam questions and identified genuine gains in practicality, customization, and efficiency, set against persistent challenges in item readability, validity, and questions of authorial ownership over AI-drafted test content. Taken together, these findings suggest that automated item generation can meaningfully reduce the workload of diagnostic test construction without yet eliminating the need for teacher review of individual items before they are used to make judgments about learners.

10.5 Strengths and Limitations of AI Feedback

As was put forward before, the rapid expansion of AI-mediated assessment has brought renewed attention to the role of feedback in language learning. While traditional assessment practices often emphasize grading and performance measurement, contemporary educational theories increasingly recognize feedback as one of the most powerful influences on learning. AI systems have transformed feedback processes by making them more

immediate, accessible, scalable, and personalized. At the same time, concerns regarding reliability, bias, learner dependence, and pedagogical appropriateness continue to shape debates surrounding their use.

The following section examines both the strengths and the limitations of AI-generated feedback, highlighting its potential contributions as well as the challenges that must be addressed to ensure responsible implementation in language education.

10.5.1 Strengths of AI Feedback

The increasing integration of AI into language education has fundamentally altered the nature of feedback. Traditionally, feedback has been constrained by the time, workload, and availability of instructors. In many educational settings, particularly those characterized by large class sizes, learners may wait days or even weeks before receiving comments on their performance. AI technologies address this challenge by providing immediate, individualized, and scalable feedback that can support learning throughout the assessment process. As a result, feedback is increasingly viewed not as a retrospective evaluation but as a dynamic mechanism that guides learning in real time.

10.5.1.1 Immediacy and Timeliness

One of the most significant advantages of AI-generated feedback is its immediacy. Research in formative assessment consistently demonstrates that feedback is most effective when delivered while learners remain cognitively engaged with the task. Delayed feedback often loses instructional value because learners may have forgotten the reasoning behind their responses or moved on to new learning activities. AI systems eliminate this delay by generating feedback instantaneously, allowing learners to review, revise, and improve their performance while the learning experience remains fresh.

In writing assessment, for example, automated feedback systems enable learners to revise drafts immediately after submission, promoting iterative writing practices and encouraging reflection on language use. Similarly, speech assessment technologies can provide real-time information regarding pronunciation, fluency, and accuracy, allowing learners to adjust their performance during practice activities. Such immediacy strengthens the connection between assessment and learning by transforming feedback into

an ongoing component of language development rather than an isolated evaluative event.

The capacity for real-time feedback is particularly valuable in writing and speaking assessment. Learners can receive guidance on grammatical accuracy, vocabulary use, organization, pronunciation, and fluency during the learning process rather than only after task completion. This immediacy strengthens self-regulated learning by encouraging reflection, revision, and active engagement with assessment outcomes. Research suggests that students who receive timely AI-mediated feedback often spend more time revising their work and demonstrate higher levels of engagement with learning tasks, leading to improved performance and deeper understanding (Kinder et al., 2025; Tan et al., 2025).

10.5.1.2 Personalization and Adaptivity

Another major strength of AI feedback lies in its capacity for personalization. Unlike traditional feedback systems, which often provide standardized comments to large groups of learners, AI technologies can analyze individual performance data and generate tailored recommendations. Personalized feedback reflects the learner's specific strengths, weaknesses, learning history, and proficiency level, thereby increasing its relevance and effectiveness.

Adaptive feedback systems continuously monitor learner progress and adjust instructional support accordingly. For example, learners struggling with particular grammatical structures may receive additional explanations and practice opportunities, while those demonstrating mastery may be directed toward more advanced tasks. This individualized approach aligns with learner-centered pedagogies that emphasize differentiated instruction and responsiveness to diverse learning needs.

The adaptive capabilities of AI are particularly valuable in language learning because learners often progress at different rates and encounter different challenges. Personalized feedback helps ensure that instructional support remains appropriately challenging while avoiding unnecessary repetition or frustration.

In a quasi-experiment with 124 L2 German learners, a domain-specific large language model feedback system known as Dr. Write produced greater gains in writing self-efficacy and clause-level accuracy than either a generic-LLM

system or teacher feedback, though it did not achieve matching gains in affective engagement (Zheng et al., 2026). The comparison suggests that it is the tailoring of feedback to learner and task—not the mere presence of an LLM—that drives the advantage. A pretest-posttest study assigning 484 EFL students to three AI-assisted-learning strategy groups (adaptive feedback, speech-recognition exercises, and intelligent tutoring) plus a control condition found that the AI-assisted strategies reduced cognitive load and improved learning outcomes relative to the control (Feng, 2025).

10.5.1.3 Scalability and Accessibility

Scalability represents one of the most frequently cited advantages of AI-mediated feedback. Human feedback is inherently constrained by time and available resources. Providing detailed individualized comments for large numbers of learners can place substantial demands on educators and institutions. AI systems overcome these limitations by delivering consistent feedback simultaneously to thousands of users.

This scalability is particularly important in online learning environments, open educational platforms, and large-scale language programs. Learners can access support regardless of geographical location or institutional resources, thereby expanding educational opportunities and promoting greater access to quality language instruction.

Beyond this, AI technologies contribute to educational accessibility by enabling learners to receive feedback outside traditional classroom settings. Mobile applications, conversational agents, and intelligent tutoring systems provide continuous learning support that extends beyond formal instructional hours. Such flexibility is particularly beneficial for adult learners, distance learners, and individuals engaged in self-directed language study.

A six-week controlled experiment with 100 learners found that an HMM-based AI smart-classroom framework raised English translation accuracy by 8.3% and engagement by 12.7% over traditional instruction, with a reported satisfaction rate of 89.2% (Chen & Hu, 2026)—evidence that gains in scale need not come at the cost of learner experience. At the institutional level, a narrative review of 77 studies published between 2018 and 2025 concluded that AI grading and feedback tools in higher education reduce educator workload and speed turnaround, though the same review cautioned that these efficiencies arrive bundled with unresolved concerns about bias, privacy, and

diminished oversight (Deepshikha, 2025)—a qualification the next section takes up in full.

10.5.1.4 Consistency and Reliability

Human evaluation is susceptible to various sources of variability, including fatigue, emotional state, personal biases, and differences in interpretation. AI systems, by contrast, apply evaluation criteria consistently across large numbers of responses. This consistency contributes to reliability and may enhance perceptions of fairness within assessment processes.

Research investigating the use of large language models as raters suggests that AI systems can achieve high levels of scoring consistency when analytical rubrics are employed (Wang et al., 2025). Unlike human evaluators, AI systems do not experience cognitive overload or fluctuations in attention, enabling them to maintain stable performance across extended periods.

Although consistency alone does not guarantee validity, it remains an important characteristic of effective assessment. Reliable feedback provides learners with clear and predictable expectations, thereby supporting transparency and trust within educational environments.

10.5.1.5 Support for Self-Regulated Learning

AI-generated feedback can also contribute considerably to the development of self-regulated learning. Self-regulation involves learners' ability to monitor their progress, identify areas requiring improvement, and take responsibility for their own learning. Because AI systems provide immediate and continuous feedback, learners receive frequent opportunities to evaluate their performance and adjust their strategies accordingly.

The availability of on-demand feedback encourages learners to engage in reflection, revision, and independent problem-solving. Rather than relying exclusively on teacher guidance, learners can actively monitor their development and make informed decisions about subsequent learning activities. Such practices foster learner autonomy and support lifelong learning skills that extend beyond language education.

Among 98 university students working with a knowledge-graph-integrated generative-AI feedback agent, those with high AI self-efficacy exhibited greater self-monitoring, lower cognitive load, and higher-quality self-evaluation on complex revision tasks than their low-AI-self-efficacy peers (Tu

et al., 2026), suggesting that AI feedback's contribution to self-regulation is conditional rather than automatic—it depends on learners already possessing, or being taught, the self-efficacy beliefs needed to exploit feedback productively, a caveat consistent with the self-regulated-learning models introduced in Chapter 3. That engagement with AI feedback is itself multidimensional is confirmed by a validation study spanning focus groups with 12 students and a subsequent survey of 542 university L2 writers, which produced a 22-item scale measuring cognitive, behavioral, emotional, and ethical engagement with ChatGPT feedback (Dong, 2025).

10.5.1.6 Data-Informed Insights for Teachers

Although discussions of AI feedback often focus on learners, educators also benefit from AI-generated data. Diagnostic reports, performance analytics, and learning dashboards provide teachers with detailed information regarding learner progress and common areas of difficulty. These insights support evidence-based instructional decision-making and enable teachers to allocate resources more effectively.

By automating routine assessment tasks, AI systems may also reduce administrative workload, allowing educators to devote greater attention to higher-order pedagogical activities such as mentoring, curriculum design, and individualized support. For this reason, AI feedback has the potential not only to enhance learner outcomes but also to strengthen teaching practices.

10.5.2 Limitations of AI Feedback

Despite its considerable advantages, AI-generated feedback is accompanied by important limitations that must be acknowledged and addressed. Many of these challenges arise from the fundamental characteristics of machine learning systems, which rely on statistical pattern recognition rather than genuine understanding. It follows that the effectiveness of AI feedback depends not only on technological sophistication but also on the pedagogical contexts in which it is used.

10.5.2.1 Limited Contextual Understanding

One of the most frequently discussed limitations of AI systems concerns their lack of genuine contextual understanding. Although contemporary large language models can generate highly sophisticated responses, they do

not comprehend language in the same way humans do. Instead, they identify patterns within vast datasets and predict probable linguistic outputs.

As a result, AI-generated feedback may occasionally overlook contextual nuances, misinterpret learner intentions, or provide recommendations that are technically accurate but pedagogically inappropriate. This limitation is particularly evident in creative writing, persuasive communication, and culturally embedded discourse, where meaning depends heavily on context, audience, and rhetorical purpose.

Language use is inherently situated within social and cultural environments. Human teachers often draw upon contextual knowledge when evaluating learner performance, whereas AI systems may struggle to account for such complexities. Accordingly, feedback generated by AI should be interpreted critically rather than accepted unconditionally.

When seven experienced EFL instructors and three AI chatbots (ChatGPT-3.5, Claude2, and Bing Chat) independently reformulated the same Japanese EFL students' writing, the human reformulations most closely preserved the original text's length and vocabulary (Nakai & Shintani, 2025)—a concrete instance of context loss even in a task that AI otherwise performs well. McKinley and Rose (2026) extend the critique further, arguing that feedback generated by teachers, peers, and AI alike tends to reinforce native-speaker norms as the implicit standard against which multilingual writing is judged, and proposing that feedback practices be reoriented around Global Englishes and translingual principles rather than a single, contextually blind target variety.

10.5.2.2 Hallucinations and Inaccuracies

Generative AI systems are susceptible to hallucinations, a phenomenon in which models generate information that appears plausible but is factually incorrect or unsupported. In educational contexts, hallucinations present a significant concern because learners may not possess sufficient expertise to identify erroneous feedback.

Li et al. (2025) note that hallucinations can result in misleading explanations, incorrect grammatical interpretations, or inappropriate feedback that learners may accept uncritically. Although advances in AI continue to improve reliability, the possibility of inaccurate feedback remains a significant concern. Consequently, AI-generated feedback should be interpreted

critically and, where appropriate, supplemented by human oversight to ensure accuracy and educational value.

Incorrect grammatical explanations, misleading vocabulary suggestions, or inaccurate interpretations of learner performance can undermine educational effectiveness and potentially reinforce misconceptions. Although ongoing improvements in model design continue to reduce hallucination rates, they remain a persistent challenge within AI-assisted education.

For this reason, many scholars emphasize the importance of maintaining human oversight and encouraging learners to verify AI-generated information rather than accepting it unquestioningly.

The difficulty is compounded by the fact that even domain experts struggle to agree on when AI-generated language content is accurate. Five expert linguists asked to rate ChatGPT-4.0's responses to 48 questions on second-language acquisition judged the responses mostly good to very good overall, yet their individual ratings agreed with one another so weak that the favorable aggregate judgment concealed near-total disagreement about which specific responses were trustworthy (Alhazmi & Muftah, 2025). If expert raters cannot converge on which AI outputs are sound, the assumption that learners—who lack that expertise—will reliably detect hallucinated or misleading feedback is difficult to sustain.

10.5.2.3 Algorithmic Bias and Fairness Concerns

AI systems learn from data, and the quality of their outputs depends heavily on the quality of the datasets used during training. If training data contain biases related to language varieties, cultural backgrounds, socioeconomic factors, or demographic groups, these biases may be reflected in assessment outcomes.

Speech assessment systems, for example, may demonstrate lower accuracy when evaluating non-standard accents or underrepresented language varieties. Similarly, writing evaluation models trained primarily on particular genres or cultural contexts may disadvantage learners whose language use differs from those norms.

Researchers concerned with responsible AI in education argue that fairness, transparency, and accountability must remain central considerations in the design and implementation of assessment systems (Cerratto Pargman et al., 2025). Addressing algorithmic bias requires ongoing validation, diverse

training datasets, and careful monitoring of assessment outcomes across different learner populations.

Rodrigues et al. (2025) argue that stylistic and cultural differences may unintentionally disadvantage certain learners if AI systems are trained predominantly on particular language varieties or communicative norms. For example, learners using regionally appropriate expressions or culturally specific rhetorical strategies may receive feedback that incorrectly categorizes such features as errors. These concerns highlight the importance of validating AI systems across diverse learner populations and ensuring that assessment technologies promote equity rather than reinforcing existing inequalities.

AI-text detectors now used in some institutions to flag AI-assisted writing have been shown to misclassify authentic essays by non-native English writers as AI-generated at far higher rates than essays by native speakers, penalizing the very lexical and syntactic patterns that mark developing interlanguage (Liang et al., 2023). Bias of this kind is not confined to detection tools: large language models have also been shown to render harsher or more negative judgments about a person when the same content is expressed in a minoritized dialect rather than a standardized variety, even though the same models show little such bias when asked directly about race or ethnicity (Hofmann et al., 2024). Together these findings indicate that the fairness risks facing AI-mediated assessment extend well beyond speech-recognition accuracy, as discussed above, to how AI systems interpret language variation of any kind—a concern Chapter 11 examines at length.

10.5.2.4 Overreliance and Reduced Learner Agency

Another important concern involves the possibility of learner overreliance on AI-generated feedback. Because AI systems provide immediate and often highly detailed responses, learners may become dependent on automated assistance rather than developing independent evaluative and problem-solving skills.

Excessive reliance on AI can reduce opportunities for critical thinking, reflection, and metacognitive engagement. Learners may focus primarily on correcting errors identified by the system without fully understanding the underlying linguistic principles. Over time, such dependence may weaken

learners' capacity to evaluate their own performance and regulate their learning independently.

Research suggests that overreliance on AI for error correction and idea generation may diminish learners' ability to identify and resolve problems independently (Kohnke et al., 2025). If learners routinely accept AI-generated suggestions without questioning or evaluating them, they may fail to develop the metacognitive skills necessary for autonomous language use. On this basis, educators must encourage learners to engage critically with AI feedback and view it as a resource for reflection rather than as an unquestionable authority. To address this challenge, educators must encourage learners to engage critically with AI-generated feedback, viewing it as a resource for reflection rather than a definitive authority.

The overreliance narrative is complicated, however, by evidence that many learners do not accept AI feedback uncritically. Forty-five L2 students revising argumentative essays with ChatGPT feedback rejected or ignored roughly 38 percent of the feedback they received—22 percent by arguing against it and 16 percent by simply disregarding it—and were significantly more likely to reject feedback on content than on form, citing its inconvenience or perceived unusefulness (Ziqi et al., 2024). Resistance, in other words, can coexist with dependence, and the two should not be conflated—a distinction Chapter 8 explores at greater length in relation to learner autonomy. Where dependence does take hold, its cost is specific: a study of 12 students using ERNIE Bot feedback in L2 writing found technological, educational, and social benefits alongside gaps in AI literacy and a risk of weakened independent writing ability (Mi et al., 2026).

10.5.2.5 Validity Challenges

While AI systems often demonstrate impressive reliability, validity remains a more complex issue. Language proficiency encompasses far more than grammatical accuracy and lexical knowledge. Effective communication involves pragmatic competence, intercultural awareness, creativity, persuasion, empathy, and strategic language use. Many of these dimensions remain difficult to operationalize computationally.

Seen in this light, AI systems may excel at evaluating certain observable linguistic features while failing to capture deeper aspects of communicative competence. This limitation suggests that AI should complement rather

than replace human judgment, particularly in high-stakes assessment contexts where nuanced interpretation remains essential.

Research suggests that AI-generated feedback often lacks the affective support that characterizes effective human teaching (Dai et al., 2025). Although contemporary models can generate increasingly natural and supportive responses, they do not possess genuine empathy or the capacity to understand learners' emotional experiences. As a result, AI systems may struggle to provide the nuanced encouragement and motivational support that many learners require. This limitation highlights the continuing importance of human educators within AI-enhanced learning environments.

Direct tests of AI feedback's effect on language development do not uniformly confirm the benefits described above. Tracking 270 EFL students' writing across a 13-week semester, one study found no significant syntactic-complexity gains from teacher feedback, automated feedback, or a combination of the two—a null result that varied somewhat by proficiency level but held across the sample as a whole (Thi & Nikolov, 2023). Validity concerns also surface in how AI and human raters relate to one another: when three human raters and three AI tools (ChatGPT, CiCi, and Bing) scored the same ESL student essays, overall score levels did not differ significantly between them, yet the inter-rater consistency between individual human and AI judges was low—similar averages concealing considerable disagreement about individual essays (Guibangguibang & Hesula, 2026). Findings of this kind do not negate AI feedback's value so much as they narrow the claims that can responsibly be made for it.

10.5.2.6 Ethical and Privacy Considerations

The increasing use of AI in education raises important ethical questions regarding data collection, privacy, transparency, and accountability. AI assessment systems often require access to substantial amounts of learner data, including written texts, speech recordings, interaction histories, and performance metrics. Ensuring the responsible management of these data is therefore a critical concern.

Learners and educators must understand how data are collected, stored, analyzed, and used. Transparent governance frameworks are essential for building trust and ensuring compliance with ethical standards. Responsible AI initiatives emphasize the importance of human oversight, informed

consent, and accountability mechanisms that protect learners' rights while supporting educational innovation (Cerratto Pargman et al., 2025).

Barrot (2026) sharpens this concern into a claim about agency rather than data handling alone, arguing that generative-AI-driven automated essay scoring risks diminishing teachers' evaluative agency unless an ecological framework requires teachers to critically mediate every AI-generated score—a requirement framed as necessary both for pedagogical soundness and for mitigating bias and protecting learner data, consonant with the augmentation-over-automation principle this book maintains throughout. Johnson et al. (2026) situate the same concern within the wider expansion of AI-based educational assessment—automated scoring, adaptive testing, intelligent tutoring, and generative item creation alike—cautioning that the ethical considerations this expansion raises are scaling alongside the technology's capabilities rather than being resolved by them, a tension Chapter 12 addresses directly.

10.6 Future Directions

The future of AI-mediated language assessment will likely be characterized by increasing integration between human expertise and intelligent technologies. Rather than replacing teachers, emerging assessment systems are expected to function as collaborative partners that augment human judgment while providing scalable and personalized support. Several key developments are likely to shape the next generation of AI-powered assessment.

First, advances in multimodal AI will enable assessment systems to analyze multiple forms of communication simultaneously, including text, speech, gesture, facial expression, and interaction patterns. Such systems may provide more comprehensive representations of communicative competence than traditional assessments focused solely on linguistic output.

Second, explainable AI (XAI) is expected to play an increasingly important role. As assessment systems become more complex, stakeholders will demand greater transparency regarding how scores and feedback are generated. Explainable AI seeks to provide understandable justifications for assessment decisions, thereby enhancing trust, accountability, and fairness.

Third, human–AI collaborative assessment models are likely to become more widespread. In these approaches, AI systems perform routine

analytical tasks while educators focus on interpretation, mentoring, and pedagogical decision-making. Such models recognize that the greatest educational benefits emerge when technological efficiency is combined with human expertise and contextual understanding.

Finally, future research must continue addressing issues of fairness, inclusivity, and responsible implementation. Ensuring equitable outcomes across diverse learner populations will require ongoing validation, ethical governance, and careful attention to the social consequences of assessment technologies.

These forecasts follow from the tensions already documented rather than standing apart from them. Human–AI collaborative assessment is not merely aspirational: Barrot's (2026) ecological framework and the DRIVE framework introduced in Section 10.2.4 both specify what teacher mediation of AI-generated scores should look like, rather than leaving collaboration as an unspecified ideal. The case for explainable AI is similarly evidence-driven: the discrepancies documented above, in which AI and human raters reach similar average scores while disagreeing considerably case by case (Guibangguibang & Hesula, 2026), or in which an AI grader assessing introductory psychology essays proved measurably more lenient at lower performance levels and stricter at higher ones than the human instructors it was compared against (Wetzler et al., 2024), cannot be diagnosed, let alone corrected, without interpretable scoring processes. Reviewing the wider empirical base, Son et al. (2025) likewise conclude that AI's integration into language education will deepen across *natural language processing, automated writing evaluation, dynamic assessment,* and *conversational agents* alike—but only to the extent that research rigor keeps pace, which the null and mixed findings surveyed in this section suggest remains uneven.

10.7 Conclusion

AI has emerged as a transformative force in language assessment, reshaping how learner performance is evaluated and how feedback is delivered. Through automated writing evaluation, speech scoring, pronunciation assessment, and grammar and vocabulary diagnostics, AI systems have expanded opportunities for personalized, scalable, and formative assessment

practices. These technologies enable immediate feedback, support learner autonomy, and provide valuable insights for both learners and educators.

At the same time, important challenges remain. Issues related to validity, bias, transparency, privacy, and learner dependence underline the need for cautious and responsible implementation. AI systems offer powerful tools for supporting language learning, but they cannot fully replace the contextual understanding, professional judgment, and relational support provided by human educators.

Ultimately, the future of language assessment is unlikely to be defined by a competition between humans and machines. Rather, it will be shaped by productive partnerships that combine the strengths of both. When implemented thoughtfully and ethically, AI has the potential to enhance assessment practices, enrich feedback processes, and contribute to more effective, equitable, and learner-centered language education.

Pedagogical Companion for Chapter 10

Key Concepts: automated essay scoring; automated writing evaluation; formative feedback; the DRIVE framework; reformulation; speech scoring; pronunciation assessment; diagnostic assessment; feedback strengths and limitations.

Discussion Questions

1. Trace one piece of writing through the full cycle this chapter describes—scoring, feedback, revision, rescoring. At which step is the most pedagogical value created, and at which is the most validity risk introduced?
2. The DRIVE framework assesses the *process* of human–AI collaboration rather than only its product. What would implementing this shift require of your current assessment practices?
3. Automated speech scoring performs differently across accents. What evidence would you require from a vendor before using such a system for consequential decisions?
4. Which of the six strengths of AI feedback (Section 10.5.1) matters most in your context, and which limitation (10.5.2) most directly threatens it?

5. Is immediate feedback always pedagogically superior to delayed feedback? Identify cases where latency serves learning.

Further Reading

- Xi, X. (2010). Automated scoring and feedback systems. *Language Testing, 27*(3).
- Oliveira, et al. (2025). The DRIVE framework. *Computers and Education: Artificial Intelligence, 9*.
- Suhan, M., & Wolf, M. K. (2025). A comparative study of human, automated scoring model, and GPT-4 ratings. *Language Testing, 43*(1).

Chapter 11. Validity, Fairness, and Academic Integrity in AI-Mediated Language Assessment

11.1 Introduction

The increasing integration of Artificial Intelligence (AI) and Large Language Models (LLMs) into educational assessment has transformed how language proficiency is measured, interpreted, and supported. AI-powered systems now perform tasks ranging from *automated essay scoring and speech assessment to grammar diagnostics and personalized feedback generation* (Wang et al., 2025; Xiao et al., 2025). While these technologies offer significant advantages in terms of *efficiency, scalability, and personalization*, they also raise important concerns regarding *validity, reliability, fairness, transparency, and academic integrity* (Kasneci et al., 2023; Cerratto Pargman et al., 2025). As generative AI becomes increasingly accessible, assessment systems must evolve to ensure that technological innovation enhances rather than undermines educational quality. In turn, researchers have called for new frameworks that address both the opportunities and challenges associated with AI-mediated assessment (Ng et al., 2025).

Historically, language assessment has relied heavily on human judgment, standardized testing procedures, and carefully designed evaluation frameworks intended to measure learners' linguistic competence and communicative abilities. The emergence of AI-powered assessment systems has introduced new approaches capable of processing large volumes of learner data, generating immediate feedback, and evaluating performance across multiple language domains. As a result, assessment is increasingly evolving from a periodic measurement activity into a continuous and data-driven process that supports learning throughout the educational journey.

Despite these promising developments, the integration of AI into language assessment raises important theoretical, pedagogical, and ethical concerns. While AI tools can enhance assessment efficiency and provide individualized support, they also introduce complex challenges regarding the validity, reliability, fairness, transparency, and integrity of the assessment process. Questions remain concerning whether AI systems accurately measure intended learning constructs, whether they can evaluate diverse learner populations equitably, and whether their decision-making processes are sufficiently transparent and accountable. Further, the widespread availability of generative AI tools has

complicated traditional understandings of academic integrity by blurring the boundaries between learner-generated and AI-assisted work.

The growing use of AI in assessment therefore requires educators, researchers, and policymakers to reconsider existing evaluative frameworks and develop new approaches that reflect the realities of AI-rich learning environments. Assessment practices must evolve not only to accommodate technological innovation but also to preserve the core principles of educational measurement. This includes ensuring that assessments remain valid indicators of language proficiency, that learners are treated fairly regardless of linguistic or cultural background, and that assessment outcomes continue to support meaningful learning and academic development.

Against this backdrop, this chapter critically examines four interconnected dimensions of AI-mediated language assessment. First, it explores issues of *validity and reliability*, focusing on the extent to which AI systems accurately and consistently evaluate language performance. Second, it investigates concerns surrounding *fairness, bias, and transparency*, highlighting the ethical implications of algorithmic decision-making. Third, it examines the evolving nature of *academic integrity* in the era of Generative AI and the challenges associated with AI-assisted learning and assessment. Finally, it discusses the urgent need for *assessment redesign*, emphasizing approaches that move beyond traditional testing paradigms toward more authentic, process-oriented, and learner-centered forms of evaluation.

Thus, this chapter argues that the future of language assessment should not be conceptualized as a choice between human and artificial intelligence. Rather, it advocates for a *human–AI hybrid paradigm* that combines the efficiency, scalability, and analytical power of AI technologies with the contextual understanding, ethical judgment, pedagogical expertise, and empathy of human educators. Such an approach offers the greatest potential for developing assessment systems that are not only technologically advanced but also valid, equitable, transparent, and educationally meaningful.

11.2 Validity and Reliability in AI-Based Assessment

Validity and reliability constitute the foundational principles of educational measurement and remain central concerns in the implementation of Artificial Intelligence (AI) within language assessment. The increasing adoption

of AI-powered assessment systems is largely predicated on the assumption that these technologies can evaluate learner performance with levels of consistency and accuracy comparable to those of human assessors. While recent advances in Large Language Models (LLMs) and Automated Essay Scoring (AES) systems have demonstrated considerable promise, ensuring the validity and reliability of AI-generated evaluations remains a complex and evolving challenge.

11.2.1 Construct Validity

Validity refers to the extent to which assessment outcomes accurately represent the knowledge, skills, or competencies they are intended to measure. In language education, validity is particularly important because language proficiency encompasses linguistic, cognitive, pragmatic, and sociocultural dimensions that cannot always be fully captured through computational analysis (Pan, 2024).

Traditional Automated Essay Scoring systems have often been criticized for relying heavily on surface-level linguistic features, such as *word count, sentence length, and lexical diversity*, which may correlate with writing quality without necessarily reflecting deeper dimensions of language proficiency (Pack et al., 2024). Such approaches have led to concerns regarding "*superficial correlation*," where assessment outcomes are influenced by measurable textual characteristics rather than genuine communicative competence.

This concern is corroborated at the level of the field's evidence base, not only in isolated cases: a systematic review of automated essay scoring research and its underlying AI and machine-learning techniques concluded that most existing systems fail to adequately evaluate *content relevance* and *discourse coherence*—precisely the higher-order dimensions that word count, sentence length, and lexical diversity cannot substitute for (Ramesh & Sanampudi, 2021). That this finding recurs across a large and heterogeneous literature suggests superficial correlation is a structural tendency of feature-based scoring approaches rather than a defect specific to any one system.

Recent research suggests that AI systems can achieve high levels of agreement with human raters in writing assessment. Wang et al. (2025) found that large language models demonstrated substantial *reliability* when

evaluating essays according to *analytical rubrics*. However, they also noted that agreement with human raters does not automatically guarantee *construct validity*. While AI systems are highly effective at identifying linguistic features such as grammatical accuracy, lexical diversity, and textual organization, they may struggle to evaluate *creativity, rhetorical sophistication, audience awareness,* and *intercultural appropriateness*. Therefore, validity research increasingly focuses on determining whether AI systems assess meaningful dimensions of communicative competence rather than merely identifying statistical patterns associated with quality writing.

Contemporary Large Language Models attempt to address these limitations by employing non-deterministic algorithms capable of evaluating more complex features of language, including coherence, argumentation, organization, and logical development (Awidi, 2024; Pack et al., 2024). As a result, LLMs can provide evaluations that more closely resemble human judgment than earlier rule-based systems. Nevertheless, researchers continue to question whether these models genuinely understand language constructs or simply generate statistically probable evaluations based on patterns learned during training (Benedetto et al., 2025).

This concern becomes particularly evident when AI systems are required to apply educational frameworks such as the Common European Framework of Reference for Languages (CEFR). Benedetto et al. (2025) found that contemporary LLMs frequently struggle to accurately operationalize CEFR proficiency descriptors when evaluating reading passages and learner writing. Rather than differentiating effectively across proficiency levels, these systems often default to intermediate-level classifications, typically assigning B1 or B2 ratings regardless of actual learner performance. Such tendencies can result in systematic overestimation of learner abilities and undermine the construct validity of assessment outcomes.

Similarly, studies investigating Automatic Short Answer Grading (ASAG) have shown that while models such as GPT-4 often produce evaluations that are consistent across demographic groups, they also demonstrate a tendency toward *leniency* and *grade inflation* when compared with human assessors (Rodrigues et al., 2025). Although consistency is generally desirable, systematically inflated scores may distort representations of learner competence, creating inaccurate impressions of achievement and potentially reducing learners' motivation to improve. For this reason, excessive

leniency poses a direct threat to construct validity because assessment results no longer accurately reflect the abilities they are intended to measure (Rodrigues et al., 2025).

The direction of this divergence is not uniform across the performance range, which matters for how aggregate agreement statistics should be read. In a comparison of 186 introductory psychology essays graded independently by eight human instructors and two versions of ChatGPT, Wetzler et al. (2024) found that the AI graders were systematically more lenient than the human instructors toward lower-scoring essays and systematically stricter toward higher-scoring ones. A single correlation coefficient between AI and human scores can therefore mask opposing biases at opposite ends of the distribution: the same system that appears reliable on average may be simultaneously inflating the weakest performances and suppressing the strongest, which is exactly where a scoring error carries the greatest consequence for the learner.

What is more, researchers caution that even highly advanced AI systems may prioritize linguistic fluency and surface-level coherence over deeper conceptual understanding. Machines do not "read" or interpret texts in the same manner as human educators; rather, they identify statistical relationships between linguistic features and previously learned patterns. This distinction raises important concerns regarding the extent to which AI-generated evaluations capture higher-order cognitive skills such as critical thinking, creativity, and disciplinary reasoning (Pack et al., 2024).

11.2.2 Enhancing Validity Through Multimodal Assessment

To address limitations associated with purely textual analysis, researchers increasingly advocate the development of multimodal assessment frameworks. Rather than relying exclusively on written or spoken language transcripts, multimodal systems integrate multiple sources of information, including acoustic, linguistic, and visual features, to generate more comprehensive evaluations of learner performance.

In spoken language assessment, Mawalim et al. (2025) argue that moving beyond single-modality "black-box" models can markedly improve both interpretability and construct validity. By incorporating acoustic features such as prosody and pronunciation, linguistic features such as vocabulary and grammar, and interactional indicators such as turn-taking behavior,

multimodal systems can evaluate distinct dimensions of communicative competence more effectively. These systems are capable of generating separate scores for fluency, coherence, interactional competence, and linguistic accuracy, thereby aligning automated evaluation more closely with established language assessment rubrics (Mawalim et al., 2025).

Similarly, research on AI-assisted listening assessment suggests that LLMs can support construct validity by manipulating lexical and syntactic complexity in a controlled manner. Aryadoust et al. (2024) found that ChatGPT-4 demonstrated a reliable ability to adjust listening passages according to targeted proficiency levels, thereby facilitating the development of assessment materials that accurately measure specific listening constructs. Such applications highlight the potential of AI not only as an assessment tool but also as a means of enhancing assessment design itself.

11.2.3 Reliability and Consistency

Reliability remains one of the strongest arguments in favor of AI-based assessment. Unlike human evaluators, AI systems do not experience fatigue, emotional influences, or fluctuations in attention. It follows that they can apply scoring criteria consistently across large numbers of responses (Wang et al., 2025).

Reliability refers to the consistency and stability of assessment outcomes across different occasions, raters, and contexts. In AI-mediated assessment, reliability is commonly examined through the consistency of AI-generated scores over time (intrarater reliability) and the degree of agreement between AI systems and human evaluators (interrater reliability) (Pack et al., 2024). Recent studies have reported encouraging results regarding the reliability of contemporary AI systems. Research on English as a Second Language (ESL) writing assessment, for example, found that modern LLMs such as GPT-4 can generate scores that correlate strongly with those assigned by human raters, achieving correlation coefficients as high as 0.73 (Pack et al., 2024). Similarly, investigations of reflective writing assessment indicate that Generative AI tools can achieve moderate to high levels of internal consistency, values comparable to those reported for human assessors (Awidi, 2024).

These findings suggest that AI systems can provide highly consistent evaluations, particularly in contexts where large volumes of learner responses

must be assessed efficiently. Unlike human evaluators, AI systems are not susceptible to fatigue, attentional fluctuations, or emotional influences, allowing them to apply scoring criteria systematically across numerous assessment tasks. Accordingly, AI technologies offer significant potential for enhancing the reliability of large-scale language assessment.

Studies employing Generalizability Theory and Many-Facet Rasch Modeling have demonstrated that AI raters can achieve reliability coefficients comparable to or even exceeding those of experienced human raters (Wang et al., 2025). Similar findings have emerged in speech assessment contexts, where AI systems provide highly consistent evaluations of *pronunciation, fluency, and lexical complexity* (Ben-Zion et al., 2025). Nevertheless, researchers caution that reliability should not be equated with quality. A system may consistently produce scores that are systematically biased or misaligned with assessment objectives. Consequently, reliability must be considered alongside validity and fairness when evaluating AI-based assessment systems.

These aggregate findings should not be read as evidence that AI–human agreement is uniformly strong. Guibangguibang and Hesula (2026) compared three human raters against three widely available AI tools—ChatGPT, CiCi, and Bing—scoring the same set of ESL student essays, and found low inter-rater consistency between the human and AI groups even though the two groups' overall score levels did not differ significantly: a pattern in which mean scores converge while individual judgments diverge, so that agreement computed on group averages can overstate the consistency any one student would actually experience. A comparable inconsistency shows up on the AI side alone. Grading 45 pharmacy students' reflective writing with rubrics generated by three different generative-AI platforms, Dicken et al. (2026) found substantial scoring disagreement not only between the AI platforms and two human graders but among the AI platforms themselves, with one system, Gemini, showing markedly lower agreement than the others. Reliability, in short, is not a fixed property of "AI raters" as a class; it varies by system, task, and the level at which agreement is measured.

However, reliability alone does not guarantee *assessment quality*. A system may consistently produce scores that are systematically biased or fail to capture the intended learning constructs. Therefore, reliability must always

be considered alongside validity when evaluating the educational effectiveness of AI-based assessment systems.

11.2.4 AI as a Complement to Human Assessment

Rather than replacing human evaluators, AI technologies may be most effective when integrated into existing assessment frameworks. One notable example is *Comparative Judgment* (CJ), an assessment approach in which evaluators compare pairs of learner responses rather than assigning scores independently. Although CJ has demonstrated strong reliability, it can be time-consuming and susceptible to inefficiencies during the initial stages of assessment.

To address these challenges, De Vrindt et al. (2025) propose the integration of Natural Language Processing techniques into CJ procedures. By generating preliminary quality estimates based on linguistic features, AI systems can support the allocation of essay comparisons and reduce the burden placed on human assessors. Such hybrid approaches help overcome the "cold-start" problem associated with comparative judgment while preserving the benefits of human evaluation. In this capacity, AI functions not as a replacement for professional judgment but as a tool that enhances assessment efficiency and reliability.

Overall, the evidence suggests that AI possesses substantial potential to support reliable and valid language assessment. However, ensuring assessment quality requires ongoing attention to construct representation, transparency, and alignment with educational objectives. While contemporary AI systems demonstrate impressive capabilities, their effectiveness ultimately depends on how they are designed, validated, and integrated into broader assessment ecosystems. On this basis, the future of AI-based assessment is likely to depend not on fully automated evaluation but on carefully designed *human–AI partnerships* that combine computational efficiency with pedagogical expertise.

11.3 Fairness and Bias in AI Assessment

Fairness is a fundamental principle of educational assessment, requiring that all learners have an equitable opportunity to demonstrate their knowledge, skills, and competencies without being disadvantaged by

factors unrelated to the constructs being measured. In the context of AI-mediated assessment, fairness extends beyond consistent scoring practices to encompass issues of *algorithmic bias, linguistic diversity, cultural representation,* and *equitable treatment* across different learner populations. As Artificial Intelligence increasingly assumes evaluative functions within language education, concerns regarding fairness have become central to discussions of responsible and ethical assessment design.

Risk area	Typical manifestation	Primary mitigation	Chapter discussion
Algorithmic bias in scoring	Systematic under-scoring of particular groups or varieties	Representative training data; group-wise error audits	11.3.1
Detection false positives	Honest work flagged as AI-generated, unevenly across populations	No punitive use without corroboration; redesigned tasks	11.3.2
Accent and variety penalties	Speech systems privileging dominant accents	Variety-stratified validation; human moderation	11.3.3
Opacity of decisions	Unexplainable consequential scores	Explainability requirements; appeal routes	11.3.4
Drift over time	Fairness degrading after deployment	Scheduled re-auditing; monitoring	11.3.5

Table 6. Fairness risks in AI-mediated language assessment and mitigation obligations

11.3.1 Algorithmic Bias

Fairness is one of the most frequently discussed ethical concerns surrounding AI-mediated assessment. Because AI systems learn from training data, they may reproduce biases embedded within those datasets. Such biases can affect learners from *different linguistic, cultural, or socioeconomic backgrounds*, potentially leading to inequitable assessment outcomes (Cerratto Pargman et al., 2025). Seen in this light, AI systems may inadvertently reproduce or amplify social, cultural, linguistic, and educational inequalities present in their training data (Rodrigues et al., 2025).

Research on AI assessment has highlighted concerns regarding *stylistic bias, cultural bias, and linguistic bias*. Writing assessment systems trained

predominantly on specific language varieties may privilege certain rhetorical conventions while penalizing alternative communicative styles (Rodrigues et al., 2025). Similarly, speech recognition technologies may demonstrate reduced accuracy when processing accented speech or underrepresented dialects, thereby disadvantaging some learner populations (Ben-Zion et al., 2025).

Many contemporary AI systems are developed using data that predominantly reflect *WEIRD (Western, Educated, Industrialized, Rich, and Democratic)* contexts, raising concerns regarding their applicability to learners from diverse sociocultural environments (Zheng & Stewart, 2024). When assessment systems are trained primarily on such datasets, they may interpret language use, rhetorical styles, and communicative practices through a narrow cultural lens, potentially leading to inequitable assessment outcomes.

Algorithmic bias can manifest in various ways. If training data u*nderrepresent specific demographic groups, linguistic communities,* or *educational contexts,* AI systems may fail to evaluate those learners fairly and consistently (Kizilcec & Lee, 2022; Rodrigues et al., 2025). Furthermore, assessment algorithms may inadvertently reinforce societal stereotypes or penalize students whose language use diverges from dominant patterns represented in training data (Pan, 2024). Such biases threaten the principle of assessment fairness because evaluation outcomes become influenced by demographic and sociocultural factors rather than by actual language proficiency.

At the same time, recent empirical evidence suggests that bias is not inevitable across all AI assessment applications. For example, studies examining GPT-4's performance in Automatic Short Answer Grading (ASAG) have found relatively stable evaluation standards across gender groups and native-language status categories (Rodrigues et al., 2025). While these findings are encouraging, they do not eliminate broader concerns regarding bias in AI systems. Rather, they highlight the need for continuous validation, auditing, and monitoring to ensure equitable performance across diverse learner populations.

11.3.2 AI Detection Systems and False Positives

Concerns regarding fairness extend beyond scoring systems to AI-detection technologies designed to identify AI-generated content. Educational institutions increasingly employ such tools to monitor academic integrity;

however, research indicates that these systems may themselves introduce substantial inequities.

A particularly significant issue involves *false-positive classifications* among non-native English speakers. Liang et al. (2023, as cited in Pan, 2024) found that AI content detectors disproportionately flag the writing of *multilingual and non-native English writers* as AI-generated. Learners who produce grammatically simple, predictable, or formulaic language may be incorrectly identified as having used AI assistance, even when their work is entirely original. Such errors have serious implications because they can subject students to unwarranted suspicion, disciplinary procedures, and reputational harm.

The scale of this concern extends beyond a single study or tool. A state-of-the-art review spanning 869 peer-reviewed articles on second-language writing assessment published over twenty-five years distills the field into five major themes and explicitly calls for a fairness- and justice-oriented, multilingual turn in how L2 writing is assessed (Plakans & Lee, 2025). Considering that backdrop, the false-positive problem in AI detection is not a peripheral glitch awaiting a technical fix; it is a fresh instance of a much older pattern in which assessment instruments calibrated on one variety of English systematically misread the productions of learners who write in another.

These findings illustrate a broader challenge in AI-mediated assessment: technologies intended to promote fairness and integrity may inadvertently create new forms of bias if their limitations are not carefully considered. As a result, researchers emphasize that AI-detection tools should not be treated as definitive evidence of misconduct but rather as supplementary indicators requiring human interpretation and verification.

11.3.3 Accent and Linguistic Diversity

Speech assessment presents particularly important fairness challenges because pronunciation standards often reflect *native-speaker norms*. Contemporary perspectives on language learning increasingly emphasize intelligibility and communicative effectiveness rather than accent elimination. However, AI systems trained primarily on native-speaker data may inadvertently reinforce discriminatory assumptions regarding acceptable language use (Xiao et al., 2025).

Beyond algorithmic bias, AI assessment systems raise important concerns regarding linguistic diversity and linguistic justice. Language assessment has historically struggled with tensions between standardization and inclusivity, and AI technologies may intensify these challenges if they privilege dominant language varieties at the expense of alternative forms of expression.

In speech assessment, Automatic Speech Recognition (ASR) systems frequently demonstrate reduced accuracy when processing *non-standard accents, regional dialects,* and speech patterns that differ from those represented in training datasets (Pan, 2024). Research indicates that models trained primarily on standardized adult speech often struggle to recognize the acoustic characteristics of *children's speech and diverse pronunciation patterns* accurately (da Silva et al., 2025). As a result, genuine language abilities may be underestimated, while learners with non-standard speech varieties may be penalized despite demonstrating effective communicative competence.

This age-related gap is documented outside educational assessment too, which suggests it reflects a property of the underlying speech technology rather than a limitation specific to any one testing system. Attia et al. (2024) name the phenomenon directly in their effort to build "Kid-Whisper," a child-adapted version of an existing automatic speech recognition model, undertaken specifically to address what they describe as a persistent performance gap between children's and adults' speech recognition accuracy. A learner assessed through an off-the-shelf ASR pipeline is therefore not simply being scored on pronunciation or fluency; the learner's age interacts with the recognizer's own error profile before any scoring rubric is applied, and younger test-takers absorb the larger share of that error.

The limitations of AI in oral assessment highlight what Pan (2024) describes as the continuing "unmatched ability of the human ear" to perceive subtle phonetic nuances, contextual variation, and communicative intent. While AI technologies have achieved remarkable progress in speech recognition, they continue to face challenges in accurately interpreting the diversity of human speech found in multilingual educational settings.

Similar concerns emerge in writing assessment. Generative AI systems often promote standardized forms of language and may inadvertently discourage linguistic variation. Research suggests that AI writing tools frequently

"*sanitize"* learner texts by replacing *dialectal, regional, or culturally specific language* features with homogenized forms aligned with dominant professional or academic norms (Lyu et al., 2025). Such practices risk *marginalizing speakers* of *World Englishes* and other legitimate language varieties by implicitly presenting a narrow conception of acceptable language use.

This homogenizing tendency has a documented analogue in how large language models treat dialectal features even outside assessment tasks specifically. Hofmann et al. (2024) found that large language models can generate covertly racist decisions about people based on features of their dialect—bias that operates beneath the level of overtly discriminatory content. The relevance for assessment is direct: a scoring or feedback system that treats dialectal or first-language-influenced structures as errors to be corrected is not making a narrowly linguistic choice but is operating in a space where covert, dialect-linked judgment has already been documented in general-purpose language models.

In addition, LLMs may favor native-speaker conventions when evaluating vocabulary choice, grammatical structures, and discourse organization, potentially disadvantaging learners whose language reflects multilingual identities or first-language influences (Pan, 2024). In global language education contexts, where linguistic diversity is increasingly recognized as a resource rather than a deficit, such tendencies raise important ethical and pedagogical concerns.

To address these concerns, researchers advocate assessment frameworks that prioritize communication success over conformity to idealized linguistic standards. Such approaches recognize the multilingual realities of contemporary communication and support more inclusive assessment practices (Cerratto Pargman et al., 2025).

11.3.4 Responsible AI

Responsible AI frameworks emphasize transparency, accountability, fairness, and human oversight as essential components of educational assessment (Cerratto Pargman et al., 2025). Institutions implementing AI technologies must therefore ensure that assessment systems undergo continuous validation and monitoring. Transparency regarding data collection,

scoring procedures, and model limitations is essential for maintaining stakeholder trust and promoting equitable educational opportunities.

11.3.5 Ensuring Fairness Through Continuous Auditing

Addressing fairness and bias in AI assessment requires ongoing evaluation and oversight. Researchers increasingly emphasize that fairness cannot be assumed simply because an AI system performs accurately on average. Instead, assessment systems must be continuously audited to identify disparities across demographic groups, linguistic communities, and educational contexts (Song et al., 2024).

Responsible AI frameworks advocate transparent *reporting practices, diverse training datasets, bias testing protocols, and human oversight mechanisms* to mitigate inequitable outcomes (Pan, 2024). Human evaluators play a particularly important role in identifying cases where AI-generated judgments may be influenced by cultural, linguistic, or contextual factors not adequately captured by automated systems.

To conclude, fairness in AI-mediated language assessment requires more than technical accuracy. It demands a commitment to linguistic inclusivity, cultural sensitivity, and educational equity. As AI becomes increasingly integrated into language education, ensuring that assessment systems respect and accommodate the diversity of learner experiences will remain one of the most important challenges facing researchers, educators, and policymakers. The future success of AI-based assessment depends not only on its efficiency and reliability but also on its capacity to provide equitable opportunities for all learners regardless of their linguistic, cultural, or demographic backgrounds.

11.4 Academic Integrity in the Era of Generative AI

The rapid proliferation of Generative Artificial Intelligence (GenAI) tools has fundamentally altered traditional *understandings of authorship, originality, learning, and academic honesty.* Since the public release of systems such as ChatGPT, educators, researchers, and institutions worldwide have engaged in extensive debates concerning the implications of AI-assisted content generation for *academic integrity*. While AI technologies offer substantial educational benefits, including personalized learning support,

feedback generation, and idea development, they also introduce significant challenges regarding the authenticity of student work and the validity of assessment outcomes (Lai et al., 2024; Lobalsamo et al., 2024).

Historically, academic integrity frameworks have been built upon the assumption that submitted work primarily reflects a learner's own knowledge, effort, and intellectual engagement. Generative AI complicates this assumption by enabling students to produce sophisticated texts, summaries, essays, and responses with minimal effort. As a result, educational institutions increasingly face the challenge of distinguishing between legitimate AI-assisted learning and inappropriate dependence on AI-generated content. These developments have prompted the emergence of new concepts, policies, and assessment practices designed to address the evolving relationship between AI and academic honesty.

11.4.1 New Challenges for Assessment

The emergence of generative AI has fundamentally altered traditional understandings of academic integrity. Tools such as ChatGPT can generate essays, summarize texts, answer questions, and provide extensive linguistic support. While these capabilities offer substantial educational benefits, they also complicate the interpretation of assessment outcomes because it is often difficult to determine the extent of learner contribution (Kasneci et al., 2023).

Traditional assessment practices were developed under the assumption that submitted work represented an individual learner's independent performance. In AI-rich environments, however, students may collaborate with intelligent systems during task completion, making it increasingly difficult to distinguish between human and machine contributions (Ng et al., 2025).

One of the most significant consequences of Generative AI is the emergence of what scholars have termed "*AIgiarism*"—the use of AI-generated content without appropriate disclosure, attribution, or critical engagement (Chan, 2024; Huang et al., 2025). Unlike traditional plagiarism, which involves copying existing human-authored material, AIgiarism involves presenting machine-generated content as original student work. Because LLM outputs are often unique, context-specific, and highly human-like in style, they are particularly difficult to identify through conventional plagiarism detection systems.

This challenge has profound implications for language assessment. Tasks that once served as reliable indicators of learner proficiency may no longer provide valid evidence of student competence if AI systems can complete them with minimal human input. In turn, educators increasingly question whether traditional writing assignments, take-home essays, and other product-oriented assessments remain appropriate measures of learning in AI-rich environments (Pan, 2024).

Research suggests that concerns extend beyond simple misconduct. The widespread availability of AI-generated content raises broader questions regarding the educational purpose of assessment itself. If students rely heavily on AI to complete assessment tasks, there is a risk that they may bypass the cognitive processes that assessment is intended to develop, including critical thinking, creativity, problem-solving, and independent reasoning (Lyu et al., 2025). In this sense, the challenge posed by Generative AI is not simply one of detecting misconduct but also one of preserving meaningful learning opportunities.

11.4.2 Detection and Attribution

Educational institutions have responded by developing AI-detection tools designed to identify machine-generated content. However, research indicates that such systems often produce false positives and false negatives, limiting their reliability as standalone solutions (Kasneci et al., 2023). For this reason, many scholars argue that assessment design should focus less on detection and more on creating tasks that encourage authentic learner engagement and critical thinking.

In response to growing concerns about AI-assisted cheating, many institutions have implemented AI-detection software designed to identify machine-generated text. However, emerging evidence suggests that such technologies remain highly unreliable and should be approached with caution.

Experimental studies indicate that even experienced educators struggle to distinguish between human-written and AI-generated texts. Fleckenstein et al. (2024), for example, found that teachers correctly identified only approximately 37.8% of AI-generated essays in some assessment contexts. This finding highlights the increasingly sophisticated nature of AI-generated language and makes clear the limitations of human judgment alone in identifying AI-assisted work.

Similarly, AI-detection systems have demonstrated significant limitations. Because Generative AI produces *non-static* and *highly variable outputs*, traditional plagiarism detection tools are largely ineffective in identifying AI-generated content (Lobalsamo et al., 2024). More concerningly, empirical investigations have revealed that many AI detectors exhibit unacceptable error rates, frequently producing both false positives and false negatives (Desaire et al., 2023; Popkov & Barrett, 2024). *False-positive classifications* are particularly problematic because they may incorrectly accuse students of academic misconduct despite the originality of their work.

These findings suggest that AI-detection technologies cannot currently serve as definitive evidence of misconduct. Instead, they should be viewed as supplementary tools whose outputs require careful human interpretation and verification. Overreliance on automated detection systems may create additional ethical concerns, including unjust accusations and increased distrust between students and educators.

It is worth distinguishing what this literature actually establishes from what is often asserted about it. What is established is narrower than headline accuracy figures suggest: applying off-the-shelf machine-learning tools to a narrow, specific writing domain has achieved very high accuracy under controlled research conditions—Desaire et al. (2023), for instance, distinguished human from ChatGPT-authored academic science writing with over 99% accuracy this way. What has not been established is that this accuracy generalizes to the conditions under which institutions actually deploy detection: heterogeneous assignments, unfamiliar topics, and a student population that includes multilingual writers whose baseline style already resembles what detectors flag. Compounding the problem, Perkins et al. (2024b) demonstrated that simple techniques are sufficient to bypass widely used GenAI text detectors altogether, meaning that even a favorable laboratory accuracy figure describes performance against unmotivated text, not against a student who has encountered the same bypass techniques. The asserted claim—that AI detectors can be trusted to identify undisclosed AI use—is accordingly weaker than the demonstrated one: that they can be trained, under narrow conditions, to tell two specific corpora apart.

11.4.3 Ethical AI Use

The rise of Generative AI has also created complex ethical ambiguities regarding acceptable academic practices. Unlike traditional plagiarism,

where ethical boundaries are relatively well defined, AI-assisted learning often exists within a continuum of acceptable and unacceptable uses.

Rather than prohibiting AI entirely, many researchers advocate the development of AI literacy and ethical usage guidelines. Learners should be encouraged to use AI as a learning support tool while remaining responsible for evaluating, revising, and contextualizing AI-generated outputs (Ng et al., 2025). Such approaches position AI as a collaborative educational resource rather than a substitute for learner effort.

Research indicates that many students distinguish between different forms of AI assistance. While outright submission of AI-generated work is generally perceived as dishonest, learners often view activities such as brainstorming, outlining, proofreading, language editing, and idea generation as legitimate forms of academic support (Huang et al., 2025). This distinction reflects broader debates concerning whether AI should be treated as a prohibited external resource or as an educational tool comparable to spell checkers, grammar assistants, or online reference materials.

The existence of these ethical gray areas complicates institutional policy development. Blanket prohibitions on AI use are increasingly viewed as impractical because students can access AI technologies easily outside formal educational environments. Beyond this, prohibitive approaches may discourage the development of AI literacy skills that are becoming increasingly important in academic and professional contexts. As a result, many scholars argue that educational institutions should focus less on eliminating AI use and more on nurturing responsible, transparent, and ethical engagement with AI technologies (Huang et al., 2025).

11.4.4 Educational Consequences of Overreliance on AI

Beyond concerns regarding misconduct, researchers have expressed growing apprehension about the potential educational consequences of excessive reliance on Generative AI. While AI systems can provide valuable support, there is concern that extensive dependence on automated assistance may diminish learners' opportunities to develop essential cognitive and metacognitive skills.

Educators have reported concerns that routine reliance on AI chatbots for writing, problem-solving, and idea generation may weaken students' capacity for independent *critical thinking, creativity, and intellectual*

persistence (Lyu et al., 2025). Similarly, Pan (2024) argues that overreliance on AI-generated responses risks devaluing educational credentials if assessment tasks no longer accurately reflect students' own abilities and achievements.

These concerns are particularly relevant in language education, where the *process of composing, revising, and reflecting on language use* is itself an important component of learning. If AI assumes responsibility for these processes, learners may receive polished outputs without engaging in the cognitive effort necessary for language development. It follows that the challenge for educators is not simply determining whether AI should be used, but rather identifying how it can be integrated in ways that support rather than replace learning.

11.4.5 Institutional Responses and Emerging Policies

Recognizing the limitations of detection-based approaches, many educational institutions are shifting toward policies that emphasize transparency and responsible AI use. Rather than prohibiting AI outright, these institutions encourage students to disclose how AI tools were used during task completion and to document their interactions with AI systems.

For example, some universities now require students to submit AI usage statements, include acknowledgments of AI assistance, or provide chat logs demonstrating how AI contributed to the development of their work (Lobalsamo et al., 2024). Such approaches align with broader movements toward academic transparency and allow educators to evaluate not only final products but also the processes through which those products were created.

These policies reflect a growing recognition that AI literacy should become an integral component of contemporary education. Instead of framing AI solely as a threat to academic integrity, institutions increasingly view responsible AI use as a skill that learners must develop in order to participate effectively in future academic and professional environments.

11.5 Assessment Redesign for AI-Rich Learning Environments

The widespread availability of generative AI necessitates a reconsideration of assessment design. Traditional assessments emphasizing memorization and routine language production may no longer provide valid indicators of

learner competence because such tasks can be completed easily with AI assistance (Kasneci et al., 2023).

Researchers increasingly advocate *process-oriented assessment* approaches that evaluate learner *decision-making, revision practices, and critical engagement with AI-generated content* (Oliveira et al., 2025). Frameworks such as DRIVE emphasize learners' ability to direct AI interactions effectively and integrate their own expertise into the writing process. These approaches recognize that future language competence may involve not only linguistic proficiency but also the capacity to collaborate critically and ethically with intelligent technologies.

Similarly, *multimodal assessments, project-based tasks, reflective portfolios,* and *authentic communicative activities* are gaining prominence because they provide richer evidence of learner understanding and reduce dependence on product-focused evaluation (Pan, 2024). Such assessment designs align more closely with contemporary educational goals emphasizing creativity, critical thinking, and lifelong learning.

To maintain the validity of language learning evaluations in an environment where GenAI is ubiquitous, higher education institutions and teachers must undergo a paradigm shift in how assessments are designed and administered.

Rather than engaging in an unwinnable "arms race" with AI text detectors, educators must transition from product-focused to process-oriented assessments. By assessing the process through which students produce their work, rather than just the final output, instructors can target forms of expression that are difficult for AI to replicate, such as personal reflection, critical evaluation, and contextual creativity (Jin et al., 2025).

Assessment redesign should leverage established pedagogical frameworks, such as Bloom's Taxonomy, to elevate the cognitive demands placed on students. Educators should design higher-order assessment items—requiring deep analysis, synthesis, and evaluation—to ensure that AI-driven responses alone are insufficient to pass (Perkins et al., 2024a). For example, instructors can design assignments where students generate a draft using an LLM and are then graded on their ability to critically critique, edit, and reference the AI-generated artifact (Perkins et al., 2024a). This approach transforms GenAI from a cheating mechanism into a tool for active

learning, forcing students to exercise their linguistic knowledge and metacognitive skills to evaluate the machine's output.

The inability to rely on traditional take-home assignments necessitates a redesign of the assessment landscape. Educators are moving away from "reproduction-based" tasks toward *authentic assessment* and *Higher-Order Thinking Skills (HOTS)* (Lobalsamo et al., 2024; Pan, 2024).

- *Process-Oriented Tasks:* Assessments are shifting to evaluate the revision process, including how students critically engage with and edit AI-generated feedback (Meyer et al., 2024; Wiboolyasarin et al., 2024).
- *Multimodal Portfolios:* Instructors are incorporating video-based portfolios, live discussions, and recorded responses that are less susceptible to AI subversion (Lobalsamo et al., 2024).
- *Updated Rubrics:* Standardized rubrics are being modified to include criteria for originality, innovation, and the extent to which learners elevate AI-produced language into clearer, more concise responses (Lobalsamo et al., 2024).
- *Prompt Engineering as Assessment*: Students can be assessed on their ability to craft sophisticated, well-structured prompts to elicit high-quality linguistic output, treating AI as a "*thinking partner*" rather than a substitute (Knoth et al., 2024; Pan, 2024).

11.6 Human–AI Hybrid Assessment

As AI technologies become increasingly integrated into educational assessment, a growing body of research suggests that the future of language assessment is unlikely to be characterized by the complete replacement of human evaluators. Instead, scholars advocate *human–AI hybrid assessment models*, which combine the efficiency, consistency, and scalability of AI systems with the *contextual understanding, pedagogical expertise,* and *ethical judgment of human assessors* (Wang et al., 2025; Cerratto Pargman et al., 2025). Such approaches recognize that while AI can effectively evaluate many observable aspects of language performance, *human evaluators* remain essential for interpreting nuanced communicative behaviors, contextual factors, and learner-specific needs. Thus, the most promising path forward involves the development of Human–AI hybrid assessment models, often referred to as *Human-in-the-Loop (HITL) systems* or *Human-Centred AI (HCAI)* frameworks. These approaches emphasize collaboration between human educators and intelligent technologies, combining the

efficiency, scalability, and analytical capabilities of AI with the contextual understanding, ethical judgment, and pedagogical expertise of human assessors (Alfredo et al., 2024).

Human–AI hybrid assessment represents a collaborative framework in which AI systems and educators perform complementary roles throughout the assessment process. While AI systems excel at processing large volumes of data, identifying linguistic patterns, and generating immediate feedback, human educators remain uniquely capable of interpreting nuanced communicative behaviors, evaluating creativity and intercultural competence, and providing the emotional and relational support that is central to effective teaching and learning (Meyer et al., 2024; Pan, 2024). Consequently, hybrid assessment models seek not to replace human judgment but to augment it through carefully designed forms of technological assistance.

In a hybrid assessment environment, AI is typically employed to automate routine and lower-level evaluative tasks. These may include grammar checking, vocabulary analysis, syntactic evaluation, pronunciation diagnostics, and preliminary scoring procedures. By reducing the administrative burden associated with such tasks, AI enables educators to devote greater attention to higher-order dimensions of language performance, including argument quality, coherence, critical thinking, intercultural awareness, creativity, and communicative effectiveness (Gayed et al., 2022; Pan, 2024). This redistribution of assessment responsibilities allows teachers to focus on aspects of language learning that remain difficult to evaluate through automated systems alone.

One particularly valuable application of hybrid assessment involves the use of AI as a "*second rater*" rather than as an autonomous evaluator. In such models, AI systems provide immediate formative feedback, preliminary scores, or diagnostic analyzes, while human assessors retain responsibility for interpreting results and making final evaluative decisions (Meyer et al., 2024). This arrangement offers several advantages. First, it enhances efficiency by accelerating feedback delivery and reducing grading workload. Second, it improves transparency because educators remain actively involved in the assessment process. Third, it provides an important safeguard against algorithmic errors, biases, and misclassifications that may arise within automated systems.

Whether this transparency is experienced as reassuring is itself an empirical question, and the answer complicates a simple assumption that more disclosure is always better. In an experiment in which undergraduates evaluated essay scores and feedback attributed to either an AI system or a human rater, most participants could not correctly identify which source had actually produced a given evaluation—yet explicitly disclosing that an evaluation was AI-generated significantly reduced students' confidence in it (Thomas et al., 2026). Hybrid models that route routine scoring to AI and reserve borderline or high-stakes decisions for human review must therefore reckon with a corollary: the visible presence of AI in the loop, even where it demonstrably improves accuracy or consistency, can itself erode the perceived legitimacy of the outcome unless the rationale for that division of labor is made legible to the students being assessed.

Human-in-the-loop frameworks are particularly important when assessment decisions involve uncertainty. Contemporary hybrid systems increasingly incorporate confidence thresholds that automatically route "borderline" responses, unusual learner outputs, or low-confidence AI predictions to human experts for manual review (Alfredo et al., 2024). Such mechanisms ensure that complex or ambiguous cases receive appropriate professional attention while allowing AI systems to handle routine assessments efficiently. This collaborative model preserves assessment quality while leveraging the speed and consistency of automation.

Hybrid assessment approaches have also demonstrated considerable potential within Comparative Judgment (CJ) frameworks. Comparative Judgment relies on evaluators making relative comparisons between learner responses rather than assigning scores directly. Although this approach has been shown to improve reliability, it can be resource-intensive when implemented at scale. De Vrindt et al. (2025) propose integrating Natural Language Processing techniques into CJ procedures to address these limitations. By generating preliminary quality estimates and identifying optimal response pairings, AI systems can reduce the burden on human judges and alleviate the "cold-start" problem frequently encountered in comparative judgment systems. Importantly, human assessors retain full control over final quality judgments, ensuring that automated processes support rather than replace professional expertise.

The success of Human–AI hybrid assessment depends not only on technological design but also on educator preparedness. Researchers emphasize the importance of developing AI literacy among teachers to ensure that intelligent technologies are used critically and responsibly (Lyu et al., 2025). Effective AI literacy extends beyond technical competence and includes understanding the capabilities and limitations of AI systems, recognizing potential sources of bias, interpreting automated outputs appropriately, and making ethically informed decisions regarding AI integration. Without such knowledge, educators may either place excessive trust in AI-generated evaluations or reject potentially beneficial technologies altogether.

This concept of *calibrated trust* is increasingly recognized as a cornerstone of successful AI integration. Teachers must develop the capacity to engage with AI neither uncritically nor skeptically, but rather as informed professionals capable of evaluating AI-generated outputs within broader pedagogical contexts (Lyu et al., 2025). Calibrated trust enables educators to leverage the advantages of AI while maintaining appropriate levels of human oversight and accountability.

From a broader educational perspective, Human–AI hybrid assessment aligns closely with contemporary visions of responsible and ethical AI implementation. By *preserving human agency* and *decision-making authority*, hybrid models address many of the concerns associated with fully automated assessment systems, including *algorithmic bias, transparency, fairness,* and *accountability* (Pan, 2024). At the same time, they retain the efficiency, scalability, and personalization benefits that make AI technologies attractive within modern educational environments.

Then, the future of language assessment is likely to be characterized not by competition between humans and machines but by productive collaboration between them. Human–AI hybrid assessment offers a framework in which technological innovation enhances rather than diminishes the role of educators. AI contributes speed, consistency, and analytical power, while human assessors provide contextual interpretation, ethical judgment, empathy, and pedagogical insight. This synergy enables assessment systems that are simultaneously data-driven and human-centered, combining technological sophistication with educational responsibility (Pan, 2024).

One of the primary advantages of hybrid assessment models lies in their ability to *balance reliability* and *validity*. AI systems frequently demonstrate high

levels of scoring consistency because they apply evaluation criteria systematically across large numbers of responses (Wang et al., 2025). Human raters, however, contribute contextual interpretation and professional expertise that help ensure assessment outcomes accurately reflect intended learning constructs. Through collaborative assessment processes, AI-generated evaluations can be reviewed, interpreted, and, when necessary, adjusted by human experts. Such practices reduce the risk of algorithmic errors while maintaining efficiency.

Hybrid assessment also provides an effective mechanism for addressing concerns regarding *fairness* and *bias*. Although AI systems can inadvertently reproduce biases present within training datasets, human evaluators can identify potentially problematic scoring patterns and intervene when necessary (Cerratto Pargman et al., 2025). Continuous monitoring by educators therefore serves as an important safeguard against unintended discrimination based on linguistic, cultural, or demographic factors. This oversight contributes to more equitable assessment practices and strengthens stakeholder trust in AI-supported evaluation systems.

The emergence of generative AI has further expanded the relevance of hybrid assessment. In AI-rich educational environments, assessment increasingly focuses not only on learners' final products but also on their ability to engage critically and responsibly with intelligent technologies. Oliveira et al. (2025) argue that future assessment frameworks should evaluate learners' capacity to direct AI interactions effectively, assess the quality of prompts they generate, and examine how they integrate AI-generated outputs into their own learning processes. Such perspectives suggest that human evaluators play a crucial role in assessing dimensions of *learner agency, critical thinking,* and *ethical decision-making* that remain difficult to capture through automated systems alone.

Hybrid assessment models are particularly valuable in writing assessment. AI systems can provide immediate feedback on grammatical accuracy, lexical sophistication, and organizational structure, while human instructors evaluate argument quality, originality, audience awareness, and rhetorical effectiveness. Similarly, in speaking assessment, AI technologies can analyze pronunciation, fluency, and speech rate, whereas human evaluators assess interactional competence, pragmatic appropriateness, and

communicative effectiveness. This complementary relationship enables more comprehensive and balanced evaluations of language proficiency.

The case for reserving interactional competence to human judgment is sharpened by research on what changes when the interlocutor itself is automated, not merely the scorer. Eguchi et al. (2025) used conversation analysis to examine 24 co-constructed interactions drawn from 75 Japanese university students who each completed parallel English roleplay tasks with a human tutor and with an autonomous, turn-taking AI agent, treating the two interlocutor types as directly comparable data for interactional-competence research. Because interactional competence is constituted jointly by both parties to a conversation—through turn-taking, repair, and the negotiation of meaning—a hybrid design that lets AI take the interlocutor role in a speaking assessment is not automating a fixed task in the way that AI grammar-checking automates a fixed task; it is substituting one half of the interaction being measured, which is a different and considerably less settled proposition.

Beyond improving assessment quality, human–AI collaboration also supports *professional development for educators*. AI-generated analytics can provide teachers with detailed insights into learner performance patterns, allowing them to make more informed instructional decisions. Rather than replacing educators, AI functions as a decision-support tool that enhances teachers' capacity to provide targeted intervention and personalized support. On this basis, hybrid assessment aligns with contemporary educational perspectives that view AI as a partner in teaching and learning rather than a substitute for human expertise.

Looking ahead, human–AI hybrid assessment is likely to become a defining feature of future language assessment systems. As AI technologies continue to evolve, the most effective assessment models will be those that leverage the strengths of both humans and machines. AI contributes speed, consistency, and scalability, while human evaluators provide contextual understanding, ethical judgment, empathy, and pedagogical insight. Together, these complementary capabilities can support assessment practices that are not only more efficient but also more valid, equitable, and educationally meaningful.

11.7 Process-Based and Formative Assessment

The emergence of Generative Artificial Intelligence (GenAI) has accelerated a significant shift in educational assessment from traditional product-oriented evaluation toward process-based and formative assessment approaches. Historically, language assessment has focused primarily on final products such as essays, presentations, examinations, and completed assignments. However, as GenAI systems become increasingly capable of producing sophisticated written and spoken outputs, assessing only the final product may no longer provide valid evidence of learners' knowledge, skills, and competencies (Al-khresheh, 2024; van Niekerk et al., 2025). Seen in this light, educators and researchers increasingly advocate assessment frameworks that prioritize the processes through which learners develop, revise, and refine their work rather than evaluating final outcomes alone.

Process-based assessment focuses on documenting and evaluating learners' developmental trajectories, decision-making processes, revision practices, and engagement with learning tasks. Rather than assessing only the final outcome, educators examine the stages through which learners generate, refine, and improve their work. This approach aligns particularly well with contemporary understandings of language learning as a dynamic and iterative process. By evaluating planning, drafting, revising, reflecting, and self-correcting activities, process-based assessment provides richer evidence of linguistic development and learner agency than product-focused approaches alone (Pan, 2024).

Language learning is widely recognized as a dynamic and iterative process involving continuous cycles of planning, drafting, revising, reflecting, and restructuring knowledge. Within this perspective, writing is not viewed as a static product but as a complex cognitive and social activity that develops through repeated interaction, feedback, and refinement (Wiboolyasarin et al., 2024). GenAI technologies can support this process by functioning as virtual tutors, collaborative partners, or "*thinking companions*" that provide immediate and individualized assistance throughout different stages of learning. Rather than replacing learners' cognitive efforts, AI systems can scaffold performance by offering targeted corrective feedback, generating alternative formulations, and prompting deeper reflection on language use (Pack et al., 2024; Wiboolyasarin et al., 2024).

The emergence of GenAI has further strengthened the rationale for process-oriented assessment. As AI tools become increasingly capable of generating fluent written and spoken language, the ability to evaluate learners' interactions with these technologies becomes an important component of assessment. Rather than prohibiting AI use, many scholars argue that assessment should focus on how learners employ AI critically, ethically, and strategically throughout the learning process (Kasneci et al., 2023). In this view, assessment extends beyond the evaluation of linguistic products to include learners' abilities to evaluate AI-generated outputs, revise machine-generated suggestions, and integrate external support into their own communicative practices.

One of the most significant contributions of AI to process-oriented assessment is its capacity to facilitate continuous formative feedback. Unlike traditional summative assessments, which primarily evaluate achievement at the conclusion of a learning period, formative assessment is designed to support ongoing development by providing learners with information that can guide improvement. A substantial body of research suggests that integrating feedback into formative assessment practices enhances learning outcomes more effectively than relying solely on summative evaluation (Cobbold & Wright, 2021; Morris et al., 2021). AI technologies strengthen this process by delivering real-time, adaptive feedback that responds dynamically to learners' evolving needs.

Formative assessment plays a particularly important role within this evolving landscape. Unlike summative assessment, which primarily measures achievement at the conclusion of a learning period, formative assessment is designed to support ongoing learning and improvement. AI technologies offer substantial opportunities to strengthen formative assessment through immediate feedback, continuous progress monitoring, and personalized learning support. Generative AI systems can provide learners with real-time guidance regarding grammar, vocabulary, organization, and language use, enabling them to revise their work iteratively and develop greater awareness of their strengths and weaknesses (Meyer et al., 2024).

Empirical evidence supports the educational value of AI-mediated formative feedback. Studies indicate that feedback generated by Large Language Models (LLMs) can improve revision quality, increase task engagement, and enhance learner motivation and positive emotional experiences (Meyer

et al., 2024). Similarly, Fleckenstein et al. (2023) found that adaptive AI-generated feedback contributes positively to text revision processes and subsequent performance outcomes. More recently, Kinder et al. (2025) demonstrated that adaptive feedback produced by an LLM substantially improved the quality of justifications in pre-service teachers' diagnostic reasoning tasks when compared with static expert-written feedback. These findings suggest that AI systems can support not only linguistic development but also higher-order cognitive processes associated with reflection, reasoning, and self-regulated learning.

Measuring this engagement rigorously requires more than task-performance data. Drawing on focus groups with 12 students and a subsequent survey of 542 university-level second-language writers, Dong (2025) developed and validated a 22-item scale that decomposes learners' engagement with ChatGPT-generated feedback into four distinct dimensions—cognitive, behavioral, emotional, and ethical—treating the ethical dimension, whether and how learners judge it appropriate to act on machine-generated suggestions, as no less measurable than whether they act on it at all. Instruments of this kind matter for process-based assessment specifically, because a revision history can document that a learner engaged with AI feedback without revealing which of these dimensions was actually operating: a purely behavioral record of whether a suggestion was accepted or rejected collapses distinctions between compliant, critical, and indifferent engagement that formative assessment is meant to tell apart.

A practical illustration of process-based assessment can be found in collaborative writing interventions that integrate AI support into multiple stages of the writing process (Wiboolyasarin et al., 2024). Such interventions typically involve three interconnected phases. First, learners engage in collaborative outlining activities, working in small groups to establish the structure, purpose, and key arguments of a writing task. Second, students produce individual drafts while utilizing AI-powered writing assistants to receive written corrective feedback and revision suggestions. Third, group members collaboratively integrate their revised sections into a final document, focusing on coherence, cohesion, organization, and overall communicative effectiveness. This multi-stage approach encourages learners to actively engage with feedback, reflect on their language choices, and participate in collaborative knowledge construction.

The increasing availability of *learning analytics* (LA) further strengthens the potential of process-oriented assessment. Contemporary AI-supported learning environments can collect and analyze large quantities of learner interaction data, including revision histories, feedback engagement patterns, writing behaviors, and participation metrics. These activity traces enable instructors to monitor learner progress in real time and identify emerging linguistic or conceptual difficulties as they occur (Acuna & Bansal, 2022; Alfredo et al., 2024; Armas-Cervantes et al., 2024; Maniyan et al., 2024). By providing detailed insights into learner performance, learning analytics support timely pedagogical interventions and facilitate more personalized forms of instruction.

In AI-enhanced environments, the role of the educator consequently evolves from that of a primary evaluator to that of *a facilitator, mentor, and learning designer*. Rather than focusing exclusively on grading completed work, instructors can guide learners through developmental processes, support critical reflection, and help students interpret and apply AI-generated feedback effectively (Pan, 2024). This shift aligns with contemporary learner-centered pedagogies that emphasize autonomy, self-regulation, and active engagement in knowledge construction.

Further, process-based assessment offers a promising response to concerns regarding *academic integrity* in the age of GenAI. Because AI systems can easily generate final products that resemble authentic student work, assessment designs that evaluate drafting histories, revision processes, reflective commentaries, peer interactions, and learner decision-making provide more robust evidence of genuine learning. Van Niekerk et al. (2025) argue that educators can strengthen academic rigor by requiring students not only to generate content with AI tools but also to critique, evaluate, and improve AI-produced outputs according to established disciplinary and academic standards. Such approaches encourage learners to engage critically with AI rather than passively accepting machine-generated responses.

Research indicates that AI-generated feedback can promote *learner motivation, engagement, and self-regulated learning* when integrated appropriately into educational contexts. Meyer et al. (2024), for example, found that AI-supported feedback mechanisms contributed positively to students' revision practices and emotional engagement with writing tasks. Such findings suggest that

AI can serve not just as an assessment tool but also as a learning facilitator that supports formative development throughout the assessment cycle.

Beyond this, process-based assessment aligns closely with broader educational goals emphasizing *critical thinking, problem-solving, collaboration, and learner autonomy*. By documenting learner progress over time through portfolios, reflective journals, revision histories, and AI interaction logs, educators can obtain more comprehensive evidence of learning than traditional one-time assessments permit. These approaches also reduce the likelihood that assessment outcomes are unduly influenced by AI-generated products because they focus on the learner's developmental journey rather than solely on the final performance.

The evidence for this optimism is not yet uniformly reassuring, particularly regarding the guidance mechanisms that process-based and hybrid models depend on to keep AI use critical rather than passive. In a survey of 1,444 U.S. college-preparatory high school students, more frequent AI use predicted lower self-confidence and a widening gap toward reliance on AI, while students' perceptions of the guidance their schools provided did not predict either outcome (Xu et al., 2026). If institutional guidance does not reliably shift how students calibrate their reliance on AI, then process-based designs that assume disclosure requirements or AI-use policies will, on their own, cultivate the critical engagement they are meant to document may be resting on a mechanism not yet shown to work at scale.

As a conclusion, process-based and formative assessment frameworks represent a necessary evolution in language assessment for AI-rich educational environments. By emphasizing learning trajectories, feedback engagement, reflection, and revision rather than final products alone, these approaches preserve assessment validity while supporting deeper and more meaningful learning. As AI technologies continue to develop, assessment systems that foreground process, learner agency, and continuous improvement are likely to play an increasingly important role in ensuring both educational quality and academic integrity.

11.8 AI-aware Assessment Literacy

The increasing integration of Artificial Intelligence (AI) into language education and assessment has created a pressing need for both educators and learners to develop AI-aware assessment literacy. As assessment practices

become increasingly mediated *by Large Language Models (LLMs)*, *learning analytics, automated feedback systems, and generative AI application*s, stakeholders require competencies that extend beyond traditional assessment literacy. Successfully navigating AI-enhanced learning environments demands not only technical familiarity with AI tools but also a critical understanding of how these technologies function, their limitations, ethical implications, and influence on teaching, learning, and assessment practices.

AI literacy is generally defined as *a collection of competencies that enable individuals to critically evaluate AI technologies, communicate and collaborate effectively with AI systems, and utilize these technologies responsibly in educational, professional, and everyday contexts* (Long & Magerko, 2020). With the rapid emergence of generative AI technologies, this concept has expanded into what researchers describe as Generative AI literacy, which includes competencies such as prompt formulation, output evaluation, AI-content detection, and the critical assessment of machine-generated information (Annapureddy et al., 2025). Within educational settings, AI-aware assessment literacy represents the application of these competencies specifically to assessment design, implementation, interpretation, and feedback practices.

For learners, AI-aware assessment literacy involves developing the capacity to engage critically with AI-generated feedback and assessment outputs rather than accepting them unquestioningly. A central component of this competence is feedback literacy, which refers to learners' ability to interpret, evaluate, and implement feedback effectively to improve their performance (Alfredo et al., 2024; Wiboolyasarin et al., 2024). In AI-mediated environments, students increasingly interact with automated systems that provide recommendations regarding grammar, vocabulary, organization, coherence, and content development. While such feedback can facilitate learning, students must also learn to identify inaccuracies, limitations, and potential biases within AI-generated recommendations. This critical stance is particularly important because contemporary AI systems remain susceptible to producing inaccurate information, fabricated references, and so-called "*hallucination*s," in which outputs appear plausible despite lacking factual accuracy (Awidi, 2024; Lai et al., 2024; Lobalsamo et al., 2024). As a result, learners must be trained more than merely as recipients of AI-generated feedback but as evaluators and validators capable of distinguishing useful guidance from unreliable outputs. Such competencies strengthen

learner autonomy and ensure that AI functions as a learning aid rather than a substitute for critical thinking and independent judgment.

Another essential dimension of AI-aware assessment literacy is *prompt engineering*, which has emerged as a crucial digital literacy skill in AI-rich educational environments. The effectiveness of generative AI systems is highly dependent on the quality and specificity of user prompts. Students who can formulate clear, context-sensitive, and purposeful prompts are more likely to obtain meaningful and pedagogically useful outputs (Knoth et al., 2024; Pan, 2024). As a result, prompt construction is increasingly viewed not only as a technical skill but also as an academic competency associated with problem-solving, communication, and inquiry. Assessment practices may therefore need to evaluate learners' ability to interact productively with AI systems rather than focusing exclusively on final AI-assisted products.

For educators, AI-aware assessment literacy constitutes an equally important professional competency. Teachers and assessment designers must possess *sufficient knowledge to integrate AI technologies responsibly while maintaining educational quality, validity, reliability, and fairness.* This competency extends beyond technical proficiency and requires an understanding of how AI systems process information, generate evaluations, and produce feedback (Knoth et al., 2024; Markus et al., 2024; Song et al., 2024). Educators must be capable of critically interpreting AI-generated outputs, identifying potential inaccuracies, and determining when human judgment should override automated recommendations.

The concept of Technological Pedagogical Content Knowledge (TPACK) provides a useful framework for understanding these competencies.

Effective AI integration requires educators to combine technological knowledge, pedagogical expertise, and disciplinary content knowledge when designing learning activities and assessments (Mishra et al., 2023). Within AI-supported assessment environments, teachers must understand not only how AI technologies function but also how they can be employed to support learning outcomes while preserving assessment validity and academic integrity. Responsible AI integration therefore depends on educators' ability to align technological tools with sound pedagogical principles and assessment objectives.

Equally, AI-aware assessment literacy requires educators to engage critically with the ethical dimensions of AI use. Researchers emphasize that teachers must understand issues related to *algorithmic bias, data privacy, transparency, accountability,* and *the broader socio-technical implications of AI systems* (Akgun & Greenhow, 2022; Ma et al., 2025; Ng et al., 2021). Awareness of these issues is particularly important because many AI systems operate as "black boxes," making it difficult to understand how specific outputs or assessment decisions are generated (Alfredo et al., 2024; Latif & Zhai, 2024). Without sufficient AI literacy, educators may either place excessive trust in automated systems or reject potentially valuable technologies due to uncertainty and misunderstanding.

The issue of *algorithmic bias* further illustrates the importance of AI-aware assessment literacy. Research has demonstrated that some AI-based assessment tools and AI-content detectors may disproportionately disadvantage multilingual learners and non-native English speakers by incorrectly identifying authentic writing as AI-generated or by favoring dominant linguistic norms (Liang et al., 2023; Pan, 2024). Educators who possess strong AI literacy are better positioned to recognize such limitations, critically evaluate automated outputs, and ensure that assessment practices remain equitable and inclusive. Similarly, students equipped with AI literacy are more capable of understanding how automated systems influence assessment outcomes and how such technologies can both support and constrain learning opportunities.

Recent scholarship also emphasizes the role of AI literacy in fostering *responsible* and *ethical AI adoption* across educational contexts. As institutions increasingly incorporate AI-supported assessment practices, stakeholders must understand not only how to use AI tools but also when, why, and under what conditions they should be used (Belkina et al., 2025; Kong et al., 2025; Pei et al., 2025; Valeri et al., 2025). AI-aware assessment literacy therefore encompasses critical awareness, ethical responsibility, and informed decision-making. These competencies help ensure that AI serves educational goals rather than undermining them and that human agency remains central to assessment processes.

As is evident, AI-aware assessment literacy serves as a crucial safeguard for maintaining human control and pedagogical responsibility within increasingly automated educational environments. Rather than allowing AI

systems to dictate assessment decisions, AI literacy empowers educators and learners to engage critically with technology, evaluate its outputs thoughtfully, and employ it in ways that support meaningful learning. As AI becomes an increasingly permanent component of language education, cultivating AI-aware assessment literacy will be essential for ensuring that assessment remains ethical, transparent, equitable, and pedagogically sound. By cultivating these competencies, educational institutions can prepare both teachers and students to participate effectively in AI-enhanced learning environments while preserving the central role of human judgment, critical reflection, and educational responsibility. In this sense, AI-aware assessment literacy represents not merely a technological skill but a *fundamental educational competency* for the future of language teaching, learning, and assessment.

11.9 Conclusion

AI has introduced unprecedented opportunities for enhancing language assessment through improved efficiency, consistency, personalization, and accessibility. At the same time, it has generated significant challenges related to validity, fairness, transparency, and academic integrity. Research demonstrates that AI systems can support reliable assessment and meaningful feedback when implemented responsibly; however, technological sophistication alone cannot guarantee educational quality (Wang et al., 2025; Cerratto Pargman et al., 2025).

The future of language assessment will likely be characterized by collaborative human–AI partnerships in which intelligent technologies provide analytical support while human educators contribute contextual understanding, ethical judgment, and pedagogical expertise. Such partnerships offer the greatest potential for developing assessment systems that are not only technologically advanced but also educationally meaningful, equitable, and trustworthy.

Pedagogical Companion for Chapter 11

Key Concepts: construct validity; reliability; multimodal assessment; algorithmic bias; detection false positives; accent diversity; academic integrity; assessment redesign; hybrid human–AI assessment; process-based assessment; assessment literacy.

Discussion Questions

1. Construct a validity argument for one AI-assessed task in your context: what is the construct, what is the evidence, and where is the weakest inference?
2. AI-text detectors produce false positives that fall unevenly across learner populations. What institutional policy follows from taking this fact seriously?
3. The chapter proposes redesigning assessment for AI-rich environments rather than policing AI out of them. Redesign one existing assessment accordingly and identify what is lost in translation.
4. In hybrid human–AI assessment, which judgment must remain human, and on what principled basis—expertise, accountability, or something else?
5. What does "assessment literacy" now include that it did not include five years ago?

Further Reading

- Perkins, M., Roe, J., et al. (2024). Simple techniques to bypass GenAI text detectors. *IJETHE, 21*.
- Jung, J. Y., Tyack, L., & von Davier, M. (2025). Automated scoring in international large-scale assessments. *CAEAI, 8*.
- Chapelle, C. A., & Douglas, D. (2006). *Assessing language through computer technology*. Cambridge University Press.

PART V

Ethics, Equity, Policy, & Future Directions

Chapter 12. Ethics, Equity, and Policy in AI-Mediated Language Education

12.1 Introduction

The integration of Artificial Intelligence (AI), Generative Artificial Intelligence (GenAI), and Large Language Models (LLMs) into language teaching, learning, and assessment has rapidly evolved from a futuristic possibility into an immediate educational reality (Belkina et al., 2025; Jin et al., 2025). Recent advances in Natural Language Processing (NLP), Machine Learning (ML), multimodal AI, learning analytics, and transformer-based language models have significantly expanded the capabilities of educational technologies, enabling increasingly personalized, adaptive, and data-driven learning experiences. AI-powered systems can now *generate instructional materials, facilitate language practice, provide automated feedback, evaluate learner performance, support assessment processes, and deliver individualized learning pathways* at an unprecedented scale (Pan, 2024). In turn, AI has emerged as one of the most transformative technological developments shaping contemporary language education.

The growing adoption of AI across educational contexts has generated considerable enthusiasm among educators, researchers, policymakers, and institutional leaders. Numerous studies suggest that AI technologies can enhance learner engagement, support self-regulated learning, personalize instruction, increase assessment efficiency, and reduce administrative workloads for teachers (Kasneci et al., 2023; Pan, 2024). Generative AI applications such as ChatGPT, Gemini, Claude, and other LLM-based systems have further accelerated this transformation by enabling interactive communication, automated content generation, intelligent tutoring, and real-time feedback. As a result, AI is increasingly viewed not simply as a supplementary educational tool but as a transformative force capable of reshaping language learning and teaching practices in fundamental ways.

Despite these opportunities, the rapid expansion of AI in education has also generated substantial *ethical, social, cultural, and regulatory concerns*. As educational systems increasingly adopt AI technologies, they encounter a range of complex socio-technical and sociocultural challenges that extend beyond questions of technological effectiveness (Wiese et al., 2025). Researchers argue that discussions surrounding AI must move beyond its

operational capabilities and address broader issues of responsibility, fairness, governance, accountability, and societal impact (Kasneci et al., 2023; Ng et al., 2025). While AI systems may improve efficiency, scalability, and personalization, they simultaneously introduce concerns related to privacy, surveillance, algorithmic bias, transparency, learner autonomy, data ownership, and unequal access to technological resources. For this reason, the implementation of AI in language education cannot be understood solely through technological or pedagogical perspectives; it must also be examined through ethical, legal, sociocultural, and political lenses.

One of the most pressing ethical concerns involves the extensive collection and analysis of *learner data*. AI-supported educational systems frequently rely on large volumes of student-generated information, including written texts, spoken interactions, behavioral logs, assessment records, learning analytics, and engagement metrics. These data enable AI systems to provide adaptive instruction, personalized feedback, and predictive insights regarding learner performance. However, they also create significant risks associated with privacy violations, unauthorized data sharing, algorithmic profiling, and educational surveillance (Alfredo et al., 2024; Jin et al., 2025). As educational institutions increasingly adopt AI-powered platforms, questions regarding informed consent, data governance, security, and ownership of learner-generated content have become central to discussions of responsible AI implementation.

Closely related to these concerns is the issue of *algorithmic bias*. Because AI systems are trained on large-scale datasets that often reflect existing social, cultural, and linguistic inequalities, their outputs may unintentionally reproduce or amplify forms of discrimination (Song et al., 2024). In language education, such biases can manifest through unequal treatment of multilingual learners, non-native speakers, speakers of regional dialects, and individuals from underrepresented cultural or linguistic backgrounds. It follows that scholars emphasize the importance of transparency, fairness auditing, explainability, and human oversight to ensure that AI systems support equitable educational outcomes rather than reinforcing existing disparities (Ma et al., 2025).

Questions of *equity* and *access* further complicate the global implementation of AI technologies. Although advanced AI tools are becoming increasingly available in educational settings, access remains uneven across

regions, institutions, and socioeconomic groups. Significant disparities persist between the *Global North* and the *Global South* in terms of technological infrastructure, internet connectivity, financial resources, institutional capacity, and digital literacy (Belkina et al., 2025; Pan, 2024). These inequalities raise concerns that AI may exacerbate existing educational divides if access to technological innovation remains concentrated among already privileged populations. Moreover, the predominance of English-language datasets and *English-centric AI* development has generated growing concerns regarding *linguistic imperialism, language homogenization,* and *the marginalization* of less-resourced and minoritized languages.

At the same time, AI technologies offer considerable potential for promoting *inclusion* and *accessibility.* AI-powered tools can support learners with disabilities through speech recognition, text-to-speech systems, adaptive learning pathways, personalized scaffolding, multimodal content delivery, and assistive communication technologies. Such innovations have the potential to create more inclusive educational environments that accommodate diverse learner needs and learning preferences (Ng et al., 2021). Nevertheless, ensuring that these benefits are distributed equitably requires careful attention to accessibility, inclusive design, educational policy, and ethical implementation.

The growing influence of AI has also prompted educational institutions, governments, and international organizations to develop policies and regulatory frameworks governing its use. Universities, schools, accreditation agencies, and ministries of education increasingly recognize the need for clear guidelines addressing academic integrity, assessment practices, ethical AI use, data governance, institutional accountability, and teacher responsibilities (Jin et al., 2025; Pan, 2024). Simultaneously, international organizations such as UNESCO and the OECD advocate human-centered approaches to AI governance that prioritize transparency, equity, accountability, human rights, and educational well-being. These developments underscore the importance of policy as a mechanism for balancing technological innovation with ethical responsibility.

Underlying these debates is the recognition that AI is not just a technological tool but a *sociotechnical phenomenon* that shapes educational values, practices, identities, and power relations. Decisions regarding AI adoption influence how knowledge is produced, how learning is assessed, whose

languages and cultures are represented, and who benefits from technological advancement. Therefore, educators, researchers, and policymakers must critically examine not only what AI can do but also what it should do within educational contexts.

Against this backdrop, this chapter examines three interconnected dimensions of AI implementation in language education: *ethics, equity,* and *policy*. First, it explores ethical issues surrounding data privacy, surveillance, algorithmic bias, learner data ownership, and responsible AI use in classrooms. Second, it investigates questions of equity and access by examining digital divides, Global North–Global South disparities, linguistic imperialism, multilingualism, and inclusion for diverse learner populations. Third, it analyzes emerging policy frameworks, institutional guidelines, and regulatory approaches designed to govern AI use in educational settings. Through this discussion, the chapter seeks to provide a comprehensive understanding of how language education can harness the transformative potential of AI while safeguarding ethical principles, promoting social equity, and ensuring responsible governance in an increasingly AI-mediated world.

12.2 Ethics of AI in Language Education

The rapid expansion of Artificial Intelligence (AI) in language education has generated unprecedented opportunities for personalization, efficiency, and learner support. At the same time, the growing reliance on AI-powered systems has raised significant ethical concerns regarding p*rivacy, surveillance, bias, accountability, transparency,* and *the responsible use* of educational technologies. As AI increasingly mediates interactions between learners, teachers, institutions, and educational content, ethical considerations have moved from the periphery to the center of discussions surrounding AI adoption in education (Kasneci et al., 2023; Pan, 2024). Ensuring that AI contributes positively to language learning requires not only technological innovation but also careful attention to ethical principles that safeguard learners' rights, well-being, and educational opportunities.

Educational technologies have never been value-neutral. AI systems are designed, trained, and deployed within social, cultural, and political contexts that influence how they operate and whom they benefit. Accordingly, decisions regarding data collection, algorithm design, assessment practices,

and feedback mechanisms can have profound implications for educational equity and learner autonomy. Researchers increasingly emphasize that responsible AI implementation requires balancing innovation with ethical safeguards that promote transparency, fairness, inclusivity, and human oversight (Akgun & Greenhow, 2022; Ma et al., 2025).

The ethical challenges associated with AI in language education are particularly complex because language itself is deeply *connected to identity, culture, communication,* and *social participation*. Unlike many other educational domains, language learning involves personal expression, intercultural interaction, and the development of communicative competence. AI systems that analyze learner language, generate feedback, or evaluate performance therefore influence not only academic outcomes but also how learners perceive themselves and their linguistic abilities. These considerations highlight the importance of examining AI through an ethical lens.

The following sections explore four central ethical issues in AI-enhanced language education: *data privacy and surveillance, algorithmic bias, ownership of learner data,* and *responsible AI use* in classrooms.

12.2.1 Data Privacy and Surveillance

One of the most significant ethical concerns surrounding the integration of Artificial Intelligence (AI) into language education involves the collection, storage, analysis, and governance of *learner data*. Contemporary AI-powered educational systems rely extensively on data to provide *personalized instruction, adaptive feedback, predictive analytics, learning recommendations, and automated assessment*. Every interaction learners have with AI-supported platforms—including written responses, spoken language samples, feedback requests, revision histories, behavioral logs, clickstream activities, and learning analytics—may be recorded and analyzed to optimize educational outcomes (Alfredo et al., 2024; Jin et al., 2025). While these capabilities offer substantial pedagogical benefits, they simultaneously introduce complex ethical and legal challenges concerning privacy, surveillance, data ownership, and learner autonomy.

From an educational perspective, data-driven AI systems offer significant advantages. *Learning analytics* can identify struggling learners, monitor progress over time, personalize learning pathways, and provide instructors with detailed insights into student engagement and performance (Alfredo

et al., 2024). AI-driven platforms can adapt instruction to individual needs and deliver targeted interventions that support more effective language development. Such capabilities have contributed considerably to the growing adoption of AI technologies across language education contexts.

However, the extensive collection of learner data raises substantial concerns regarding privacy and educational surveillance. Students frequently generate large volumes of personal information while interacting with AI systems, often without fully understanding how their data are collected, processed, stored, shared, or reused. Educational institutions increasingly rely on commercial AI platforms whose data governance practices may not always be transparent to users. Consequently, critical questions emerge regarding *informed consent, data ownership, third-party access, data retention*, and the long-term use of educational information (Kasneci et al., 2023).

Empirical evidence on how learners themselves experience these risks complicates any assumption that awareness alone produces caution. A qualitative study using linguistic textual analysis of questionnaire responses from ICT-savvy users of AI-driven foreign-language-learning applications found that these users were well aware of the privacy threats such tools pose, yet, precisely because of their own technical knowledge, felt comparatively unconcerned about their personal vulnerability to misuse (Pikhart & Klímová, 2025). Technical literacy, in other words, can produce a false sense of security as readily as it produces genuine caution.

These concerns have become even more pronounced with the widespread use of generative AI (GAI) tools such as ChatGPT and other large language models. Language learning frequently requires students to submit open-ended essays, reflective journals, discussion posts, voice recordings, and conversational interactions that may contain highly personal information. When such artifacts are processed by external AI systems, significant ethical and legal questions arise regarding how third-party providers manage, store, and potentially reuse sensitive learner data (Lyu et al., 2025; Markus et al., 2025). Educational institutions increasingly emphasize that interactions with commercial AI systems must occur within clearly defined boundaries that protect personal, confidential, and proprietary educational information from unauthorized use or downstream commercial exploitation (Jin et al., 2025).

The issue of educational *surveillance* has become particularly salient within AI-enhanced learning environments. AI technologies can continuously monitor learner behavior, engagement patterns, participation levels, revision activities, and academic performance. While such monitoring may support early intervention and personalized learning, excessive surveillance risks undermining learner autonomy and transforming educational environments into spaces of constant observation. Researchers warn that the normalization of continuous data collection may alter learner behavior, reduce feelings of trust, and contribute to anxiety or diminished agency among students (Akgun & Greenhow, 2022).

The expansion of AI-powered monitoring technologies further complicates these concerns. Educational institutions increasingly employ AI-assisted proctoring systems, plagiarism detection tools, and behavioral monitoring software to identify potential instances of academic misconduct. Some of these technologies track student activities beyond traditional classroom settings, collecting data about online behaviors, screen activity, location information, or patterns of digital engagement. Such practices have generated concerns regarding whether learners are adequately informed about the extent of monitoring and whether meaningful consent is truly possible within compulsory educational environments (Perkins et al., 2024b). Critics argue that excessive monitoring may fundamentally alter the educational relationship by shifting the classroom from a space of trust and intellectual exploration toward one characterized by surveillance and control (Iqbal et al., 2025).

Survey evidence suggests this concern is not confined to a small circle of critics. A survey of 2,000 Japanese respondents found that *privacy* and *dystopian-surveillance concerns* significantly predicted negative attitudes toward emotional AI in educational contexts, accounting for 44% of the variance in such attitudes, and toward AI-enabled children's toys, accounting for 38% (Ho et al., 2024). Nearly half the variance in attitudes toward affectively sensitive classroom AI explained by surveillance concern alone indicates a mainstream wariness that institutions cannot simply assume away.

Privacy concerns are especially pronounced in language education because language-learning activities often involve highly personal forms of communication. Learners frequently discuss personal experiences, cultural perspectives, beliefs, emotions, and opinions through written assignments, discussion forums, and conversational AI systems. The collection and analysis

of such interactions create substantial ethical responsibilities regarding confidentiality, security, and appropriate data usage. Institutions must therefore establish transparent policies governing what data are collected, who can access them, how long they are retained, and under what circumstances they may be shared or repurposed.

The growing use of Large Language Models (LLMs), learning analytics, and Multimodal Learning Analytics (MMLA) systems has intensified these concerns. Many AI systems require vast quantities of data to function effectively, yet the origins, collection methods, and limitations of training datasets often remain opaque to educators and learners alike. This lack of transparency raises important questions concerning data ownership, informed participation, and the ethical boundaries of educational data collection (Long & Magerko, 2020; Ng et al., 2021). What is more, AI-supported educational environments may obscure when data are being collected and what forms of information are being analyzed, increasing the risks of continuous *surveillance, algorithmic profiling,* and *unauthorized data use*.

Emerging AI technologies are also capable of collecting increasingly sophisticated forms of learner data. *Multimodal learning analytics* systems may *track eye movements, facial expressions, vocal characteristics, physiological responses,* and *other behavioral indicators* in an effort to personalize learning experiences and predict educational outcomes (Giannakos & Sharma, 2020; Feng et al., 2025). While these capabilities may improve adaptive instruction, they also create profound ethical questions regarding how such data should be interpreted, stored, and protected. Learners may not fully understand the scope of behavioral information being collected, nor the implications of allowing educational systems to infer cognitive, emotional, or motivational states from these data.

For these reasons, researchers increasingly emphasize the importance of robust data governance frameworks capable of regulating *data quality, security, privacy, transparency, and accountability* (Hooshyar et al., 2025). International regulatory frameworks similarly stress principles such as informed consent, transparency, purpose limitation, data minimization, and learner control over personal information. Educational institutions are therefore expected to ensure that AI systems comply with relevant legal and ethical standards while maintaining trust among learners, educators, and other stakeholders.

Some institutional responses translate these governance principles into classroom practice rather than policy language alone. In Finland, 213 fourth- and seventh-graders across twelve classrooms co-designed AI applications within a distributed-scaffolding framework built to foster children's data agency, understanding, creativity, and ethical awareness (Kahila et al., 2024). A parallel initiative in the United States, the Institute for Student-AI Teaming, pursues a similar logic by embedding 'AI Partner' technologies within K-12 collaborative learning teams through equity-focused design processes conducted jointly with educators and a diverse range of students (D'Mello et al., 2024). Such projects treat data governance as something learners practice rather than merely consent to.

To sum up, privacy protection must be regarded as a foundational requirement rather than a secondary consideration in AI-enhanced language education. Although learner data can support more effective teaching, assessment, and personalization, their collection and use must be governed by ethical principles that prioritize human rights, dignity, autonomy, and educational well-being. Responsible AI implementation requires educational institutions to balance the pedagogical benefits of data-driven innovation with the fundamental obligation to protect learner privacy, establish transparent ownership boundaries, and prevent excessive surveillance within educational environments.

12.2.2 Algorithmic Bias

Algorithmic bias represents one of the most significant ethical challenges associated with the use of Artificial Intelligence (AI) in language education. Although AI systems are often perceived as objective, neutral, and data-driven, their outputs are fundamentally shaped by the datasets, design choices, and computational assumptions embedded within their development. On this basis, AI technologies may inadvertently reproduce, reinforce, or even amplify existing social, cultural, linguistic, and educational inequalities. In language teaching, learning, and assessment, such biases can influence feedback quality, instructional recommendations, automated evaluations, and learning opportunities, potentially disadvantaging particular groups of learners (Akgun & Greenhow, 2022; Song et al., 2024).

The origins of algorithmic bias can largely be traced to the data used to train AI systems. Contemporary Large Language Models (LLMs) are trained on vast collections of digital texts that reflect existing social

structures, historical inequalities, cultural assumptions, and power relations. Because these datasets frequently overrepresent dominant languages, cultures, and perspectives, AI systems may internalize patterns that privilege certain groups while marginalizing others (Ma et al., 2025; Rodrigues et al., 2025). As a result, *algorithmic outputs* may appear technically sophisticated while simultaneously reproducing underlying societal biases.

In foreign language education, algorithmic bias is particularly concerning because linguistic diversity constitutes a fundamental aspect of the learning process. Learners bring *diverse linguistic repertoires, cultural backgrounds, communication styles, and educational experiences* to the classroom. However, AI systems often operate according to dominant linguistic norms that may not adequately reflect this diversity. Seen in this light, language learners whose language use differs from the patterns most strongly represented within AI training data may receive less accurate evaluations, lower-quality feedback, or inappropriate instructional recommendations.

One area where algorithmic bias becomes especially visible is automated language assessment. AI-powered scoring systems are increasingly used to evaluate writing, speaking, pronunciation, grammar, and vocabulary performance. While these systems frequently demonstrate high levels of scoring consistency, concerns remain regarding whether they evaluate all learners equitably. Learners from diverse linguistic, cultural, and educational backgrounds may produce language that differs from the patterns most commonly represented in training datasets. As a result, automated systems may misinterpret legitimate linguistic variation as error, thereby producing evaluations that do not accurately reflect learners' actual communicative competence (Pack et al., 2024; Benedetto et al., 2025).

Recent studies illustrate both the promise and limitations of AI-based assessment. Research examining GPT-4 and other advanced language models has shown that these systems can evaluate open-ended responses with considerable consistency and may avoid explicit demographic discrimination in controlled testing environments (Rodrigues et al., 2025). Nevertheless, the risk of writing-style, dialectal, and cultural bias remains substantial. If an AI evaluation system implicitly favors particular discourse structures, rhetorical styles, or culturally dominant linguistic conventions, it may systematically disadvantage learners whose language use reflects different sociocultural backgrounds (Benedetto et al., 2025). Such outcomes

undermine the validity and fairness of assessment by measuring conformity to dominant norms rather than actual language proficiency.

Accent bias presents another significant concern, particularly in AI-mediated speaking assessment. Automatic Speech Recognition (ASR) systems typically perform most accurately when processing speech varieties that resemble those represented in their training datasets. As a result, speakers with regional accents, non-native pronunciations, or less commonly represented varieties of English may experience lower recognition accuracy. In assessment contexts, this can negatively affect automated evaluations of pronunciation, fluency, and oral proficiency, raising important concerns regarding fairness and equity. Learners may be penalized not because of inadequate communicative ability but because their speech diverges from the linguistic norms embedded within the system.

Researchers have similarly identified instances of dialect prejudice within AI systems. Because many models are trained primarily on standardized language varieties, they may systematically undervalue non-standard dialects, regional language practices, or multilingual forms of communication (Baker & Hawn, 2022; Hofmann et al., 2024). Such tendencies are particularly problematic in language education, where learners should be encouraged to appreciate linguistic diversity rather than perceive non-standard varieties as inherently deficient. AI systems that privilege a narrow conception of language risk reinforcing linguistic hierarchies and marginalizing legitimate forms of linguistic expression.

Algorithmic bias also extends beyond language assessment to generative AI applications. Generative systems frequently produce outputs that reflect dominant cultural assumptions, social stereotypes, and demographic patterns present in their training data. For example, studies have shown that *image-generation systems* often reproduce narrow Western beauty standards and default to white, Westernized representations when presented with culturally neutral prompts (Iqbal et al., 2025). Although such examples originate outside language assessment, they illustrate broader concerns regarding the ways AI systems can normalize particular cultural perspectives while rendering others less visible. Similar patterns may emerge in language education when AI-generated examples, narratives, or instructional materials privilege certain cultural viewpoints over others.

The predominance of English-language and Western-centric training data further contributes to algorithmic bias. Most commercially available AI systems have been developed using datasets dominated by standardized English and content produced in economically powerful regions of the world (Nyaaba & Zhai, 2025; Wang et al., 2025). In turn, AI systems may generate culturally insensitive outputs that fail to account for local realities, multilingual contexts, or diverse educational traditions. Researchers have documented instances in which AI systems default to assumptions based on Western seasonal patterns, educational experiences, or sociocultural norms while overlooking the lived realities of learners from the Global South (Nyaaba & Zhai, 2025). Such tendencies reinforce concerns regarding cultural representation and educational equity.

A particularly troubling manifestation of algorithmic bias involves AI-content detection systems. These technologies are increasingly used to identify potential misuse of generative AI in academic settings. However, empirical studies have demonstrated that some AI detectors are more likely to classify writing produced by multilingual and non-native English speakers as AI-generated, even when the text is entirely human-authored (Liang et al., 2023). Such false positives may lead to unwarranted accusations of academic misconduct, disproportionately affecting learners whose linguistic patterns differ from dominant native-speaker norms. These findings underline the potential consequences of deploying imperfect AI systems in high-stakes educational contexts.

Additional ethical concerns arise regarding intellectual property, authorship, and learner identity. Researchers have noted that LLMs may inadvertently reproduce copyrighted material or encourage forms of language production that diminish learners' individual voices (Boulhrir & Hamash, 2025; Ho & Lee, 2025). If AI systems consistently encourage particular stylistic conventions or linguistic forms, they may contribute to the homogenization of language use and reduce opportunities for authentic self-expression. In language education, where the development of individual voice and communicative identity is often a key objective, such outcomes warrant careful consideration.

Addressing algorithmic bias requires coordinated technological, pedagogical, and institutional responses. Researchers increasingly advocate for the development of more inclusive AI systems trained on linguistically,

culturally, and demographically diverse datasets (Song et al., 2024). Regular auditing procedures, transparency requirements, fairness evaluations, and explainable AI mechanisms can help identify problematic patterns before they negatively affect learners. Educational institutions should also ensure that AI-supported decisions remain subject to human review, particularly in high-stakes contexts such as assessment, placement, and academic integrity investigations.

Whether such responses take hold appears to depend less on available technical capability than on the values embedded within the organizations building these systems. A survey of 194 AI professionals, analyzed through a Value-Belief-Norm model, found that personal diversity, equity, and inclusion values—themselves shaped by the perceived diversity climate of the professional's own organization—predicted the sense of moral responsibility that in turn drove adoption of inclusive, responsible AI development practices (Falcon & Carillo, 2026). Calls for more representative datasets and routine fairness auditing are therefore unlikely to succeed as purely technical mandates detached from that organizational culture.

Human oversight remains essential because educators can recognize contextual factors, cultural nuances, and learner characteristics that may not be adequately captured by automated systems. Teachers are uniquely positioned to identify potentially biased outcomes, interpret learner performance holistically, and ensure that assessment practices remain equitable and educationally meaningful. This perspective aligns with emerging human-centered approaches to AI governance that emphasize collaboration between human expertise and technological systems rather than complete automation (Alfredo et al., 2024).

This concern has been formalized into an ecological framework for automated essay scoring, which argues that generative AI-driven scoring risks diminishing teachers' evaluative agency unless teachers are positioned to critically mediate AI-generated scores rather than defer to them, treating each score as a provisional judgment to be weighed against contextual knowledge of the learner—a stance the framework presents as necessary for mitigating bias and protecting data privacy alike (Barrot, 2026).

All in all, algorithmic bias is more than merely a technical limitation but an ethical, social, and educational challenge. Ensuring fairness in AI-enhanced language education requires ongoing attention to representation, diversity,

transparency, accountability, and inclusion. As AI technologies become increasingly embedded within language teaching, learning, and assessment, institutions must actively work to prevent automated systems from reproducing existing inequalities. Ethical AI implementation depends not only on technological sophistication but also on a sustained commitment to educational justice, linguistic diversity, and equitable opportunities for all learners.

12.2.3 Ownership of Learner Data

AI-powered educational systems rely heavily on learner-generated information to provide personalized instruction, adaptive feedback, automated assessment, predictive analytics, and learning recommendations. Every interaction learners have with AI technologies—including written assignments, prompts submitted to generative AI systems, spoken recordings, assessment responses, revision histories, conversational exchanges, feedback interactions, and behavioral engagement data—may be collected, processed, stored, and analyzed within digital platforms (Alfredo et al., 2024; Jin et al., 2025). While these data support increasingly sophisticated educational services, they also raise important questions regarding who owns educational data, who controls access to them, and who ultimately benefits from their collection and use.

Through history, educational data were generated and maintained primarily within institutional environments under the supervision of schools, universities, and instructors. However, the rapid expansion of commercial AI technologies has fundamentally altered this landscape. Many contemporary AI systems are developed and operated by private technology companies whose business models depend heavily on large-scale data collection, algorithm refinement, and continuous model improvement. As students increasingly use generative AI tools for language learning, writing assistance, translation, feedback, and assessment preparation, substantial amounts of learner-generated content may become incorporated into proprietary technological ecosystems (Kasneci et al., 2023). This development raises important concerns regarding whether learners retain ownership of the content they create and whether educational institutions can adequately protect student interests when third-party providers are involved.

The issue is particularly significant in language education because learner-generated content often contains deeply personal, creative, and intellectual

contributions. Essays, reflective journals, discussion posts, oral presentations, language portfolios, and conversational interactions represent not only evidence of linguistic development but also expressions of learners' identities, experiences, opinions, and cultural perspectives. When such materials are uploaded to AI-powered systems, questions emerge concerning intellectual property rights, authorship, access, and control. Some scholars argue that *students should retain ownership* of all educational artifacts they create, regardless of the technological platforms through which those artifacts are generated, submitted, or analyzed (Akgun & Greenhow, 2022). Others emphasize the need for clearer *legal and institutional frameworks that define ownership rights* and responsibilities among learners, educational institutions, technology developers, and platform providers.

The emergence of Generative AI has further complicated discussions surrounding ownership and authorship. Increasingly, learners engage in collaborative interactions with AI systems when producing written and spoken outputs. Students may receive AI-generated suggestions, revisions, translations, content recommendations, or even substantial portions of draft text that contribute to final assignments. In such contexts, determining authorship becomes considerably more complex. Questions arise regarding the extent to which AI-generated contributions should be attributed, whether learners can claim full ownership of AI-assisted work, and how educational institutions should evaluate intellectual contributions in AI-mediated learning environments (Pan, 2024). These concerns extend beyond traditional discussions of academic integrity and intersect with broader debates surrounding creativity, intellectual labor, and human–machine collaboration.

Recent scholarship highlights growing tensions among multiple stakeholders regarding ownership of both AI-generated content (AIGC) and the learner data used to train AI systems (Ode et al., 2025). Technology companies often rely on user interactions to improve proprietary models, creating situations in which student submissions may contribute indirectly to future system development. Educational stakeholders increasingly question whether learners should have greater control over how their content is used and whether students should be explicitly informed when their educational data contribute to commercial model refinement (Jin et al., 2025). These concerns become particularly acute when student-authored content is incorporated into proprietary systems without explicit consent, compensation, or meaningful opportunities for opting out.

The collection of large-scale language datasets also raises broader questions concerning intellectual property rights. When students submit essays to automated grading systems, plagiarism detection software, or AI-powered learning platforms, ownership of those texts may become ambiguous. In some cases, learners may unknowingly grant extensive rights to technology providers through complex terms-of-service agreements that are rarely read or fully understood (Perkins et al., 2024). For this reason, scholars have raised concerns that the intellectual property rights of learners may be compromised when educational content is used for purposes extending beyond its original instructional context.

Data ownership concerns are closely linked to issues of transparency and informed consent. Many learners remain unaware of how their interactions with AI systems are stored, analyzed, shared, or potentially reused. Terms of service agreements are frequently lengthy, highly technical, and inaccessible to non-specialist users. As a result, students may unknowingly authorize extensive uses of their educational data without fully understanding the implications. Ethical AI implementation therefore requires clear communication regarding what data are collected, how they are used, who has access to them, how long they are retained, and whether they may contribute to future system training or commercial activities (Ma et al., 2025).

Educational institutions play a central role in addressing these challenges. Universities, schools, and language programs bear significant responsibility for ensuring that learner data are managed ethically, transparently, and responsibly. This includes establishing policies governing data collection, storage, retention, sharing, deletion, and ownership; evaluating the privacy and governance practices of third-party AI providers; and ensuring compliance with relevant legal and regulatory requirements. Institutions must also provide learners with meaningful opportunities to understand, manage, and exercise their rights regarding personal information and educational records. Such measures contribute to greater trust, transparency, and accountability within AI-supported learning environments.

The increasing use of learning analytics further amplifies concerns regarding ownership and control. Contemporary educational platforms can generate extensive profiles of learner behavior by tracking participation patterns, engagement metrics, interaction histories, performance indicators, and learning trajectories (Alfredo et al., 2024). While these data may

support personalized learning and early intervention strategies, they also create persistent digital representations of learners that may remain within institutional or commercial systems long after educational activities have concluded. Questions therefore arise regarding how long such records should be retained, whether learners can access, transfer, or delete their data, and how institutions can prevent misuse, unauthorized disclosure, or commercial exploitation.

Emerging frameworks for responsible AI governance increasingly emphasize the principle of learner agency. From this perspective, students should not be viewed merely as sources of data but as active stakeholders who possess rights concerning the information they generate and the ways in which it is used. Ethical AI systems should therefore provide mechanisms that allow learners to understand, manage, access, and control their personal information. Such approaches align closely with broader human-centered models of AI governance that prioritize transparency, autonomy, accountability, and respect for individual rights (Ng et al., 2021; Ma et al., 2025).

As is obvious, questions of learner data ownership extend far beyond technical or legal considerations and touch upon broader issues of power, autonomy, authorship, and educational justice. As AI systems become increasingly integrated into language teaching, learning, and assessment, educational institutions must ensure that learners remain informed participants rather than passive providers of data. Responsible AI governance requires transparent ownership policies, meaningful consent mechanisms, ethical data management practices, and a sustained commitment to protecting the rights and interests of learners within increasingly data-driven educational environments. The challenge is not simply determining who possesses learner data but ensuring that the collection and use of those data serve educational purposes while respecting learner dignity, autonomy, and intellectual ownership.

12.2.4 Responsible AI Use in Classrooms

While AI technologies offer significant opportunities for personalization, accessibility, efficiency, and learner support, their educational value ultimately depends not on technological sophistication alone but on the principles, values, and pedagogical practices that govern their implementation. Responsible AI in education refers to the development and deployment of

AI systems that are human-centered, transparent, fair, accountable, and aligned with educational goals and societal values (Kasneci et al., 2023; Goellner et al., 2024; Ma et al., 2025). Within language education, responsible AI use requires balancing technological innovation with pedagogical integrity, learner well-being, educational equity, and ethical responsibility.

A fundamental principle of responsible AI implementation is maintaining *meaningful human oversight.* Although AI systems can automate numerous educational functions—including feedback generation, content creation, language assessment, learning analytics, and instructional support—educational decision-making should not be delegated entirely to algorithms. Human educators possess *contextual understanding, pedagogical expertise, ethical judgment, cultural awareness,* and *emotional intelligence* that remain beyond the capabilities of current AI systems. It follows that researchers increasingly advocate human-in-the-loop approaches in which AI functions as a supportive tool while teachers retain ultimate responsibility for instructional and assessment decisions (Alfredo et al., 2024; Pan, 2024). Such approaches allow institutions to benefit from the efficiency and scalability of AI while preserving the professional judgment and relational dimensions that are essential to effective language education.

Responsible AI use also requires *transparency* regarding when, how, and why AI systems are being employed. Learners and educators should understand the functions AI performs, the types of data it uses, and the limitations of its outputs. However, many contemporary AI systems operate as complex "black boxes," making it difficult for users to understand how particular recommendations, scores, feedback comments, or instructional suggestions are generated (Latif & Zhai, 2024). This lack of transparency can undermine trust, limit accountability, and prevent users from critically evaluating AI-generated outputs. Responsible implementation therefore requires clear communication regarding system capabilities, limitations, data usage practices, and the role AI plays within educational processes.

Closely connected to transparency is the principle of e*xplainability*. Educational stakeholders should be able to understand the reasoning underlying AI-generated recommendations and decisions, particularly in high-stakes contexts such as assessment, placement, and academic evaluation. Explainable AI seeks to provide interpretable outputs that allow educators and learners to understand why specific feedback, scores, or recommendations

have been generated (Song et al., 2024). Such transparency not only enhances trust but also enables users to verify, challenge, or contextualize AI-generated conclusions when necessary. In language education, where assessment outcomes can markedly influence learner confidence, motivation, and academic trajectories, explainability becomes especially important.

Academic integrity represents another central dimension of responsible AI use. Generative AI tools can support learners in brainstorming ideas, revising texts, translating content, generating examples, and refining language production. While these capabilities may enhance learning, they also create opportunities for misuse, dependency, and the circumvention of genuine intellectual effort. Increasingly, educators recognize that prohibiting AI use altogether is neither realistic nor pedagogically productive. Instead, responsible implementation requires clear guidelines regarding acceptable uses of AI, transparency requirements, attribution practices, and expectations for learner engagement (Pan, 2024). Such approaches encourage students to use AI as a learning partner and reflective tool rather than as a substitute for critical thinking and active participation.

Researchers further emphasize that responsible AI literacy should be viewed not merely as a regulatory requirement but as a core educational competency (Wiese et al., 2025). Rather than approaching AI ethics solely through restrictive policies, language educators are increasingly encouraged to embed ethical decision-making directly into classroom practice and curriculum design (Ng et al., 2025). This perspective recognizes that learners require opportunities to engage critically with AI technologies, understand their societal implications, and develop the capacity to use them responsibly in academic and professional contexts.

Several practical pedagogical approaches have been proposed to support the development of responsible AI literacy. One strategy involves analogy-based and experiential pedagogies in which learners participate in role-playing activities, simulations, or embodied exercises that demonstrate how AI systems classify information and how bias can emerge within algorithmic processes (Iqbal et al., 2025). Such activities help learners understand that AI outputs are not inherently neutral but are shaped by human decisions regarding data collection, categorization, and model design.

Another important approach involves encouraging students to critically evaluate misinformation, deepfakes, AI hallucinations, and biased outputs.

By analyzing the credibility, accuracy, and social consequences of AI-generated content, learners develop the critical literacy skills necessary to navigate increasingly automated information environments (Wiese et al., 2025). Such activities support not only language development but also broader competencies related to media literacy, digital citizenship, and ethical reasoning.

Collaborative ethical co-design offers a further avenue for responsible AI implementation. Researchers advocate involving students directly in discussions about classroom AI policies, privacy frameworks, and the design of AI-supported learning activities (D'Mello et al., 2025). By participating in these decisions, learners gain firsthand understanding of the trade-offs associated with data collection, personalization, and educational technology use. Such participatory approaches promote learner agency and reinforce the principle that educational technologies should serve learners rather than merely extract data from them.

The ethical foundations of responsible AI have also been articulated through broader normative frameworks. Floridi and Cowls (2022) identify five foundational principles that should guide AI implementation: *beneficence, non-maleficence, autonomy, justice, and explicability.* Applied to language education, these principles emphasize the importance of ensuring that AI technologies contribute positively to learning, avoid causing harm, respect learner autonomy, promote fairness, and provide transparent explanations for their actions. These principles offer a valuable framework for evaluating whether AI systems align with educational values and ethical responsibilities.

AI literacy constitutes another indispensable component of responsible AI use. Both educators and learners require the knowledge and skills necessary to understand AI technologies, critically evaluate their outputs, recognize their limitations, and identify potential ethical concerns. For educators, AI literacy includes understanding the pedagogical, technical, and ethical dimensions of AI integration. For learners, it involves developing the ability to assess AI-generated information critically, recognize bias and misinformation, and use AI tools responsibly and effectively (Long & Magerko, 2020; Ng et al., 2021). Without adequate AI literacy, users may place excessive trust in automated outputs or fail to recognize inaccuracies, biases, and ethical risks associated with AI systems.

The issue of overreliance on AI has received increasing attention within recent scholarship. Researchers warn that excessive dependence on AI-generated suggestions may contribute to what has been described as "*metacognitive laziness*," whereby learners rely on automated systems rather than engaging in independent reflection, problem-solving, and critical thinking (Nyaaba & Zhai, 2025; Yan et al., 2025). Responsible AI use therefore requires educational practices that position students as active evaluators and decision-makers rather than passive recipients of machine-generated outputs. Human-centered approaches encourage learners to function as the "cultural authorities" of their own learning experiences, maintaining agency and ownership over educational processes.

Responsible AI implementation further requires sustained attention to *fairness, diversity, and inclusion*. Educational AI systems should support learners from diverse linguistic, cultural, and socioeconomic backgrounds rather than privileging dominant groups. Achieving this goal requires continuous monitoring for algorithmic bias, regular auditing of system performance, and the inclusion of diverse perspectives during AI development and implementation (Akgun & Greenhow, 2022; Ma et al., 2025). Educational institutions must therefore establish governance mechanisms that ensure AI contributes to equitable educational opportunities rather than reproducing existing inequalities.

Institutional policies play a critical role in supporting these objectives. Schools and universities increasingly develop guidelines that define acceptable uses of AI, establish responsibilities for educators and learners, regulate data governance practices, and address concerns related to academic integrity and assessment (Jin et al., 2025). Effective policies provide frameworks for ethical decision-making while remaining sufficiently flexible to accommodate the rapid pace of technological change.

Thus, responsible AI use in language classrooms requires viewing AI as a tool that enhances rather than replaces human teaching and learning. Educational goals extend far beyond efficiency and measurable performance to include critical thinking, creativity, collaboration, intercultural competence, ethical reasoning, and personal growth. AI should therefore be deployed in ways that support these broader educational objectives rather than narrowing learning experiences to what can be easily automated or

quantified. Human-centered approaches emphasize that technology should serve learners and educators—not the reverse (Alfredo et al., 2024).

In summary, responsible AI use in language education depends upon a combination of human oversight, transparency, explainability, ethical governance, academic integrity, AI literacy, and learner agency. As AI becomes an increasingly influential component of educational practice, institutions must ensure that technological innovation remains aligned with pedagogical values, ethical principles, and human well-being. Responsible AI is not simply a technical challenge but an ongoing ethical commitment to nurturing meaningful, equitable, and learner-centered education in an increasingly AI-mediated world.

12.2.5 Accountability When AI Systems Harm a Learner

The governance principles surveyed above—*transparency, fairness auditing, human oversight*—specify what institutions should do in general, but they say comparatively little about who answers for a specific learner's specific harm: a misgraded essay that alters a placement decision, a false accusation of AI-generated writing, or an adaptive system that quietly fails a learner it was never designed to serve. A qualitative systematic review of 71 peer-reviewed studies published between 2020 and 2024 used thematic coding and SWOT synthesis to map the governance tensions surrounding agentic AI in global education, finding these tensions clustered around unresolved questions of institutional alignment, responsibility, and the distributed autonomy of increasingly autonomous systems (Mariyono, 2026).

The legal terrain is no more settled. A legal analysis of autonomous, agentic AI systems used in academic assessment, student evaluation, recruitment, and workforce decisions examines precisely these regulatory and liability questions, treating the allocation of responsibility among vendor, institution, and supervising educator as an open problem rather than one existing law already resolves (Atre, 2026). At the level of everyday institutional practice, a survey of 75 university educators has begun mapping how such governance mechanisms are perceived and how ready institutions are judged to be for agentic AI adoption in higher education (Brohi & Jhanjhi, 2026)—a stocktaking exercise whose very existence suggests the field is still describing this problem more than resolving it.

Two further strands complicate any easy reliance on a human in the loop as the default answer. An empirical study of whether, why, and which employees comply with directives issued by AI systems rather than human supervisors has informed the development of competency frameworks for human-oversight roles, on the premise that oversight functions as a safeguard only if the human involved is positioned and willing to exercise independent judgment rather than defer automatically to the system (Jamshaid et al., 2025). A conceptual framework for school AI systems takes a complementary design-level approach, proposing low-level modelling of actor properties that keeps a human presence identifiable and traceable within a system's architecture and thereby reduces the risk of misuse or misinterpretation (Isop, 2025). Whether either approach has been adopted at scale remains unclear: a systematic review that screened 22,711 articles down to 553 full-text studies on responsible-AI principles and practice found extensive published material on principles such as accountability and transparency, synthesized into six cross-cutting findings spanning both foundations and implementation (Gunasekara et al., 2025)—a synthesis whose necessity suggests how immature the passage from stated principle to institutional practice remains. Until institutions can specify, in advance, who answers when a given system fails a given learner, human-in-the-loop oversight risks diffusing responsibility rather than assigning it, precisely the gap the Governance-bound pillar of the PAGE framework, introduced in Chapter 13, is designed to close.

12.3 Digital Divides and Linguistic Imperialism

While Artificial Intelligence (AI) offers significant opportunities to transform language teaching, learning, and assessment, its benefits are not distributed equally across educational contexts. The rapid expansion of AI-powered learning tools has intensified longstanding concerns regarding educational inequality, digital access, linguistic justice, and the unequal distribution of technological resources. As AI increasingly shapes educational practices worldwide, questions concerning who has access to these technologies, whose languages are represented within AI systems, and whose educational needs are prioritized have become central to discussions of equity in language education (Pan, 2024; Belkina et al., 2025).

One of the most persistent challenges involves the existence of *digital divides* that affect learners' ability to access and benefit from AI-enhanced educational opportunities. Digital divides extend beyond disparities in access to computers and internet connectivity; they also encompass differences in technological infrastructure, institutional resources, digital literacy, AI literacy, and opportunities for meaningful technology use. Although AI technologies have become increasingly accessible in many parts of the world, substantial inequalities remain among countries, educational institutions, and socioeconomic groups. Learners in well-resourced environments often benefit from advanced AI platforms, reliable internet access, digital learning ecosystems, and AI-literate educators, whereas students in under-resourced settings may lack even basic technological infrastructure necessary to engage with AI-supported learning tools (Kasneci et al., 2023; Pan, 2024).

Recent scholarship suggests that the deployment of AI in education may inadvertently *deepen existing inequalities* if access to technological resources remains uneven. Wealthier regions, institutions, and educational systems possess the infrastructure and financial capacity necessary to utilize premium AI services, secure digital platforms, cloud-based learning environments, and advanced analytics systems. In contrast, marginalized communities and underfunded schools often struggle to access foundational technological resources, creating barriers to participation in AI-enhanced educational opportunities (Agyare et al., 2025; López-Costa et al., 2025). As a result, the educational advantages associated with AI risk becoming concentrated among already privileged populations.

These disparities are particularly visible when comparing educational contexts across the Global North and the Global South. Many of the most influential AI technologies are developed within economically powerful countries that invest heavily in technological research, innovation, and educational technology infrastructure. Accordingly, learners and institutions in these regions often gain earlier access to emerging technologies, specialized training opportunities, and supportive policy frameworks. By contrast, many educational institutions in developing regions face challenges related to limited funding, inadequate infrastructure, restricted internet access, and insufficient professional development opportunities (Belkina et al., 2025). Such inequalities raise concerns that AI may reinforce rather than reduce global educational disparities.

Significant is that access to technology alone does not guarantee equitable participation. Effective engagement with AI-enhanced learning environments requires a range of competencies, including *digital literacy, AI literacy, critical thinking skills, and the ability to evaluate and use technological outputs responsibly* (Ng et al., 2021; Ma et al., 2025). Learners who lack these competencies may be unable to benefit fully from AI tools even when technological access is available. Similarly, educators who have not received adequate training may struggle to integrate AI into pedagogically meaningful ways. Consequently, digital divides involve not only material resources but also the human capacities required to participate effectively in technologically mediated educational environments.

Systematic reviews of the empirical record confirm that this *dual burden*—material access and human capacity—persists even where AI-based instruction demonstrably works. A PRISMA-guided review of 22 empirical studies, drawn from 189 screened, found that AI tools produced significant gains across reading, writing, listening, speaking, and vocabulary in school EFL settings, yet digital-divide and teacher-readiness challenges persisted regardless of the gains recorded (Kundu & Bej, 2025). A broader systematic review of 63 higher-education studies similarly found that human-centered approaches—combining AI with blended learning, augmented reality, and gamification—outperformed technology-centric implementations, while digital-divide and faculty-readiness gaps continued to limit which institutions could adopt them (Mahyiyuddin & Erlita, 2026). Pedagogical effectiveness and equitable access, across both reviews, are separate achievements: demonstrating the first does not secure the second.

Alongside concerns regarding access, scholars increasingly highlight the relationship between AI and linguistic imperialism. *Linguistic imperialism* refers to the dominance of particular languages—most notably English—in global systems of communication, education, and knowledge production. Contemporary AI technologies often reinforce these hierarchies because they are predominantly trained on English-language datasets and developed within English-dominant technological ecosystems. Hence, AI systems typically demonstrate substantially higher levels of performance, accuracy, and functionality when operating in English than when supporting many other languages (Pan, 2024; Benedetto et al., 2025).

This imbalance has profound implications for language education. Learners studying English benefit from extensive AI-supported resources, including conversational agents, automated writing evaluation systems, intelligent tutoring systems, translation tools, and adaptive learning platforms. In contrast, learners and educators working with less-resourced languages frequently encounter lower-quality technological support, reduced functionality, and less accurate outputs. On this basis, AI technologies risk reinforcing the global dominance of English while simultaneously marginalizing local, indigenous, and minority languages.

Researchers have increasingly warned that AI may contribute to new forms of digital linguistic inequality if language representation remains concentrated among dominant global languages (Song et al., 2024). Because most AI systems are optimized for standardized English and other widely represented languages, they may inadequately support multilingual learners, regional language varieties, and less commonly taught languages. Such disparities influence not only technological performance but also educational opportunities, potentially affecting which languages are learned, taught, funded, and valued within educational systems.

The challenge extends beyond language representation to broader issues of *cultural* and *epistemological dominance*. Some scholars describe this phenomenon as *technological colonialism*, arguing that the global dissemination of AI systems developed primarily within Western contexts may reproduce historical power imbalances by exporting linguistic norms, cultural assumptions, and educational values to diverse learning environments (Akgun & Greenhow, 2022). AI systems frequently embody implicit assumptions regarding communication styles, assessment practices, knowledge structures, and language use that may not align with local educational traditions or cultural realities. Seen in this light, the adoption of AI technologies without critical adaptation may contribute to forms of cultural homogenization that diminish linguistic and cultural diversity.

These concerns are especially relevant for learners in the Global South. Researchers have noted that the growing dominance of English-language AI systems may marginalize indigenous knowledge systems, local languages, and culturally specific forms of expression (Nyaaba & Zhai, 2025). Furthermore, many AI models establish evaluation baselines based on standardized and highly polished forms of English, potentially disadvantaging

English Language Learners (ELLs) and speakers of *World Englishes* whose linguistic practices differ from dominant norms (Wang et al., 2025). Such tendencies risk reinforcing existing linguistic hierarchies while limiting recognition of diverse communicative practices.

The unequal distribution of AI benefits has therefore become a major concern within discussions of educational equity. While AI promises to democratize access to educational resources, its benefits currently remain more accessible to institutions and regions possessing greater economic and technological capacity (Boulhrir & Hamash, 2025; Prinsloo, 2020). Without deliberate intervention, AI may contribute to a widening gap between those who can leverage advanced educational technologies and those who cannot.

Addressing these challenges requires a commitment to equitable and inclusive AI development. Researchers increasingly advocate for multilingual AI systems, culturally responsive technologies, and greater representation of underrepresented languages within *AI training datasets* (Ma et al., 2025). Such initiatives can help ensure that technological innovation benefits a broader range of learners and communities rather than reinforcing existing inequalities. Equally important are investments in infrastructure, digital literacy, AI literacy, and teacher professional development, which are essential for reducing disparities in access and participation.

Educational policymakers also play a crucial role in promoting equitable AI adoption. Policies that expand technological access, support open educational resources, encourage local AI innovation, and invest in underserved educational communities can help mitigate global disparities. Importantly, equity should not be viewed solely as a matter of technological distribution but as a broader commitment to ensuring that all learners can participate meaningfully in AI-enhanced educational environments regardless of their linguistic, cultural, geographic, or socioeconomic backgrounds.

The transformative potential of AI in language education cannot be fully realized without addressing the inequalities that shape both access to technology and representation within technological systems. Digital divides and linguistic imperialism remind educators and policymakers that technological innovation alone does not guarantee educational justice. Rather, equitable outcomes require deliberate efforts to ensure that AI technologies

support linguistic diversity, cultural inclusion, educational accessibility, and equal opportunities for learners across diverse global contexts.

12.3.1 Global South and Global North Perspectives

Discussions of equity in AI-enhanced language education frequently highlight significant differences between the Global North and the Global South. Although Artificial Intelligence is often presented as a universally applicable educational innovation, its adoption, implementation, governance, and educational impact vary considerably across geographical, economic, political, and sociocultural contexts. These differences shape not only access to AI technologies but also how educational institutions perceive their value, formulate policies, allocate resources, and integrate AI into teaching, learning, and assessment practices.

The Global North generally encompasses *economically developed countries* characterized by *advanced technological infrastructure, substantial investment in research and development, widespread internet connectivity, and mature digital learning ecosystems*. Educational institutions in these regions often possess the financial, technological, and human resources necessary to experiment with emerging AI technologies, provide extensive professional development opportunities, and develop institutional frameworks governing AI use (Jin et al., 2025). As a result, many schools and universities across North America, Western Europe, and parts of East Asia have become early adopters of generative AI applications, intelligent tutoring systems, learning analytics platforms, and AI-supported assessment technologies.

Within these contexts, AI is frequently framed as a catalyst for educational innovation. Institutional policies and academic discussions often emphasize the potential of AI to cultivate individual creativity, support personalized learning, improve efficiency, and promote critical engagement with emerging technologies (Belkina et al., 2025). In turn, policy debates in the Global North frequently focus on issues such as *academic integrity, authorship, innovation, intellectual property, and the ethical use of AI* within educational environments. The central concern is often how to integrate AI while preserving *originality, critical thinking, and learner autonomy*.

By contrast, many countries within *the Global South* continue to face significant structural challenges that influence AI adoption and implementation.

Educational institutions across parts of *Africa, Asia, Latin America, and other developing regions* often encounter *limitations related to technological infrastructure, internet connectivity, financial resources, and access to educational technologies* (Agyare et al., 2025; Belkina et al., 2025; Ode et al., 2025). While educators and policymakers in these regions frequently recognize the transformative potential of AI, practical barriers often restrict the extent to which such technologies can be incorporated into everyday educational practice. For this reason, global discussions of AI in education may sometimes reflect assumptions and experiences that inadequately represent the realities of under-resourced educational contexts.

Comparative research across country pairs supports this reading and, importantly, locates the gap in infrastructure and economics rather than in attitudes toward the technology itself. Semi-structured interviews with 40 Iranian and Italian EFL teachers found that Italian teachers felt ready to integrate AI tools into their teaching, whereas Iranian teachers reported readiness barriers rooted specifically in economic and infrastructural factors (Ghiasvand & Kogani, 2024). A comparative interview study of six language teachers from Eastern (Vietnam, Iran, Indonesia) and Western (Spain, Germany, UK) countries reached a consonant conclusion: all six were positive about AI-assisted language learning, yet voiced differing equity concerns tied to their national contexts rather than differing enthusiasm for the technology itself (Huertas-Abril & Palacios-Hidalgo, 2023). The same pattern recurs at a larger scale: a PRISMA-guided systematic review of 22 empirical studies, screened down from 189, found that AI tools produced significant gains in reading, writing, listening, speaking, and vocabulary in school EFL settings, yet concluded that *digital-divide and teacher-readiness challenges* persisted across those settings (Kundu & Bej, 2025). A mixed-methods study spanning multiple institutions likewise found that AI-driven personalized English instruction improved student motivation, participation, and language skills, while identifying unequal technology access, alongside teacher-training needs, as the binding constraint on wider adoption (Bernal Parraga et al., 2025). Taken together, this evidence reframes the Global North–Global South divide discussed throughout this section: the obstacle is rarely a deficit of will.

These disparities are not just technological but also educational, social, and political. Educational priorities often differ substantially between regions. In many Global North contexts, AI is commonly viewed as a mechanism

for enhancing personalization, efficiency, and educational innovation. Discussions frequently focus on optimizing learning outcomes, improving assessment systems, and developing sophisticated digital learning ecosystems. In contrast, institutions in the Global South may prioritize expanding educational access, addressing teacher shortages, supporting multilingual populations, overcoming infrastructural limitations, and maximizing resource accessibility (Jin et al., 2025). It follows that AI solutions designed within one context may not necessarily address the most pressing educational challenges faced by another.

Researchers increasingly caution against adopting a one-size-fits-all approach to AI implementation in education (Pan, 2024). Educational technologies developed within particular sociocultural, linguistic, and economic environments often embed assumptions regarding pedagogy, assessment practices, communication norms, learner behavior, and educational values. When these technologies are transferred across contexts without adaptation, they may fail to account for local educational realities, cultural traditions, or linguistic diversity. This concern has contributed to growing calls for context-sensitive and locally responsive approaches to AI development and deployment.

Language education provides a particularly compelling illustration of these challenges. Many AI systems have been trained predominantly on linguistic data derived from English and other globally dominant languages. As a result, educational applications frequently demonstrate higher levels of accuracy, functionality, and pedagogical effectiveness for learners studying English than for those learning or using less-resourced languages. For institutions operating within multilingual societies—particularly across Africa, Latin America, South Asia, and Southeast Asia—such limitations can reduce the educational relevance of AI technologies while simultaneously reinforcing existing linguistic inequalities (Song et al., 2024).

The Global South perspective also highlights broader concerns regarding technological dependency and educational sovereignty. Much of the AI infrastructure currently used worldwide is controlled by a relatively small number of technology companies located within economically dominant countries. This concentration of technological power raises important questions concerning data governance, intellectual ownership, educational autonomy, and the ability of local institutions to shape AI systems according

to their own cultural, linguistic, and educational priorities. Scholars argue that sustainable and equitable AI adoption requires greater investment in local research capacity, multilingual AI development, and regionally relevant educational innovation (Ma et al., 2025).

These concerns are closely linked to discussions of linguistic imperialism. The predominance of *English-language datasets* and *Western-centric technological development* can marginalize local languages, indigenous knowledge systems, and culturally specific forms of expression (Nyaaba & Zhai, 2025). Researchers have warned that overreliance on English-dominated AI ecosystems risks reinforcing historical patterns of linguistic and cultural dominance while limiting opportunities for educational technologies to support local identities and knowledge traditions. In assessment contexts, AI systems often establish evaluation baselines based on *standardized forms of English*, potentially disadvantaging English Language Learners (ELLs), multilingual speakers, and users of pluralistic Englishes whose language practices differ from dominant norms (Wang et al., 2025).

The emergence of *culturally responsive AI design* offers one promising response to these concerns. Researchers have developed frameworks such as the Culturally Responsive Lesson Planner (CRLP), which utilizes Interactive Semi-Automated (ISA) prompt architectures to gather culturally specific parameters before generating educational content (Nyaaba & Zhai, 2025). For example, teachers may incorporate local linguistic and cultural resources—such as the Asante Twi language in Ghana—into AI-generated instructional materials. Such approaches help ensure that AI systems draw upon locally relevant knowledge rather than relying exclusively on generic Western-centric datasets (Kitsara, 2022; Nyaaba & Zhai, 2025). These innovations demonstrate how AI can be adapted to support cultural diversity rather than contributing to cultural homogenization.

Notably, the Global South should not be viewed solely through a deficit perspective. Across developing regions, researchers, educators, and institutions are actively contributing innovative AI solutions designed to address local educational challenges. In Ghana, for example, design science research has produced GenAITEd Ghana, a curriculum-aligned conversational AI agent developed and evaluated specifically to support responsible, culturally responsive teacher education (Nyaaba et al., 2026). In many cases, resource constraints have encouraged creative approaches to

technology integration, including mobile learning initiatives, low-cost AI applications, open educational resources, and community-based digital learning projects. These developments demonstrate that meaningful educational innovation can emerge from diverse contexts rather than exclusively from technologically dominant regions.

International collaboration offers an important pathway toward more equitable AI development and implementation. Partnerships among universities, governments, educational organizations, and technology providers can facilitate knowledge exchange, capacity building, multilingual AI development, and the creation of culturally responsive educational technologies. However, such collaborations must be grounded in principles of mutual respect, reciprocity, and equitable participation rather than reproducing existing hierarchies of knowledge production and technological authority.

As is evident, examining AI through Global South and Global North perspectives underlines the importance of contextualizing educational innovation. The effectiveness, ethics, and equity of AI cannot be evaluated independently of the social, economic, cultural, and political environments in which technologies are developed and deployed. Responsible AI adoption therefore requires moving beyond universal solutions toward approaches that recognize local needs, support linguistic and cultural diversity, strengthen *educational sovereignty*, and promote equitable participation in the benefits of technological advancement. Only through such context-sensitive approaches can AI contribute meaningfully to more inclusive and globally equitable language education.

12.3.2 English Dominance and Minoritized Languages

The rapid growth of Artificial Intelligence (AI) in language education has intensified longstanding concerns regarding the global dominance of English and the marginalization of less-resourced, indigenous, and minoritized languages. While AI technologies have significantly expanded opportunities for language learning, communication, assessment, and knowledge access, they have also exposed and, in some cases, reinforced structural inequalities embedded within global digital language ecosystems. Because most contemporary AI systems are trained on *large-scale datasets dominated by English-language content*, they generally perform substantially better in English than in many other languages. This imbalance carries

important implications for linguistic diversity, educational equity, cultural representation, and language preservation.

English currently occupies a privileged position within AI development and digital knowledge production. Much of the textual data used to train Large Language Models (LLMs) originates from English-language websites, academic publications, news media, social networks, and online communication platforms. Accordingly, AI systems typically possess more extensive linguistic knowledge, greater accuracy, and richer functional capabilities when processing English than when operating in many other languages (Pan, 2024). Learners studying English therefore benefit from access to highly sophisticated AI-supported educational tools, including intelligent tutoring systems, automated writing evaluation platforms, conversational agents, translation services, adaptive feedback technologies, and multimodal learning applications.

In contrast, many indigenous, regional, and less-resourced languages remain considerably underrepresented within AI development. These languages often lack the extensive digital corpora required for training advanced AI systems, resulting in lower levels of system performance, reduced functionality, and limited educational support. As a consequence, speakers and learners of these languages frequently have fewer opportunities to benefit from AI-enhanced educational environments. Researchers have therefore warned that technological advancement may unintentionally widen existing linguistic inequalities if language representation remains concentrated among dominant global languages (Benedetto et al., 2025).

Recent benchmark testing gives this warning a concrete, alarming shape. Testing of Gemini 2.5 Flash found refusal rates for unsafe or sensitive prompts near 90% when the model operated in English, compared with only 35% to 55% across four West African languages—Yoruba, Hausa, Igbo, and Igala—indicating severe degradation in the model's safety behavior once it operates outside its dominant training language (Faruna, 2026). A gap of this magnitude means the same system can function as a reasonably well-governed tool in English while functioning as a considerably less safe one for learners who happen to be studying, or being taught, in a minoritized language—a disparity no amount of pedagogical goodwill on the part of individual teachers can correct.

The issue extends beyond technical functionality to broader questions of linguistic justice. Language is more than merely a tool for communication; it is also a carrier of cultural knowledge, collective memory, identity, history, and community belonging. When AI systems prioritize dominant languages, they may inadvertently reinforce global linguistic hierarchies that privilege certain language communities while marginalizing others. Consequently, scholars have expressed concern that AI-driven educational systems may contribute to processes of language shift and linguistic homogenization if educational institutions increasingly favor languages that receive stronger technological support (Song et al., 2024).

These concerns are closely connected to debates surrounding linguistic imperialism. Historically, the global spread of English has been linked to economic, political, and cultural power. The predominance of English within digital environments has created new forms of linguistic inequality that now extend into AI-supported education. Because most LLMs are trained primarily on English-dominant and Western-centric internet corpora, they often reinforce standardized English norms and dominant linguistic conventions (Benedetto et al., 2025). As a result, AI-generated feedback, assessment systems, and instructional recommendations may privilege standard English varieties while marginalizing pluralistic Englishes, regional dialects, multilingual communication practices, and minoritized languages.

Design-based research on generative AI tools bears this out directly. A design-and-development study that tested three separate ChatGPT configurations for English language teaching found consistent pedagogical limitations in the models' capacity to represent pluralistic Global Englishes rather than standard forms (Lee et al., 2025). The finding matters pedagogically, not simply technically: a language classroom that draws its models of English from AI systems tuned to a single standardized variety teaches, however unintentionally, that only that variety counts as correct English.

This dynamic has important educational consequences. Learners may increasingly perceive English as more valuable, useful, or technologically advantageous because of the extensive AI resources available for English-language learning. While proficiency in English undoubtedly offers significant academic and professional opportunities, educational policies that prioritize English at the expense of local languages risk undermining linguistic diversity and cultural heritage. In multilingual societies, educational

institutions frequently face the challenge of balancing participation in global knowledge economies with the preservation of local linguistic identities and indigenous knowledge systems.

The challenges are particularly significant in regions where local, regional, and indigenous languages play essential social and cultural roles. In many educational contexts across Africa, Latin America, Asia, and Oceania, local languages remain central to cultural continuity, community engagement, and identity formation. However, the limited representation of these languages within AI systems can reduce their visibility and perceived relevance within increasingly digital educational environments. Researchers have therefore emphasized that language technologies should support multilingualism rather than accelerate *language marginalization* (Nyaaba & Zhai, 2025).

At the same time, AI technologies also offer significant opportunities for *language preservation* and *revitalization*. When multilingual capabilities are intentionally prioritized, AI can facilitate language documentation, translation, content generation, language learning, and the preservation of endangered linguistic resources. Recent initiatives demonstrate how AI can be leveraged to support rather than undermine linguistic diversity. For example, universities and research institutions across the Global South have increasingly advocated the inclusion of indigenous and regional languages within AI ecosystems. North-West University in South Africa has actively promoted the integration of indigenous African languages into AI platforms, recognizing that participation in AI development is essential for ensuring that these languages remain visible and functional within future digital environments (Jin et al., 2025).

Efforts to expand linguistic representation within AI systems are gaining momentum globally. Researchers increasingly advocate for the development of *inclusive multilingual AI models* capable of supporting a broader range of languages, dialects, and cultural contexts (Ma et al., 2025). Such initiatives acknowledge that technological innovation should strengthen rather than diminish linguistic diversity. Expanding multilingual datasets, incorporating underrepresented languages into training corpora, and developing culturally responsive AI systems are essential steps toward reducing linguistic inequality and promoting more equitable access to AI-supported educational opportunities.

Promising developments have already begun to emerge. Open-source multilingual AI initiatives have sought to create models capable of translating and processing hundreds of underrepresented languages, thereby challenging the dominance of English-centric systems (META, 2022). Similarly, advances in multilingual transformer architectures and cross-lingual content scoring have improved the capacity of AI systems to evaluate language production across diverse linguistic backgrounds, creating opportunities for more equitable automated assessment practices (Horbach et al., 2024). Such developments suggest that AI can evolve beyond monolingual paradigms toward more inclusive and linguistically diverse educational applications.

Yet the relationship between *AI-supported translation* and *language preservation* is not straightforwardly positive, and the record here is more mixed than the preceding examples suggest. Technically, low-resource language translation has advanced: systems for endangered languages that lack enough authentic parallel text for a purely data-driven approach have increasingly combined large-language-model-generated synthetic data with more traditional rule-based translation methods to close that gap. But technical capability and beneficial adoption are not the same thing. A survey study of Basque speakers' machine-translation adoption habits and perceptions found that growing MT use raises concerns for a carefully planned process of minority-language revitalization (Aranberri & Iñurrieta, 2024), suggesting that even technically competent AI translation can complicate, rather than simply support, a minoritized language community's efforts to expand its own use rather than be translated around it. The two observations together caution against treating multilingual AI capability as self-evidently a gain for linguistic diversity: capability is necessary for language preservation but plainly not sufficient, and its effects depend on how communities choose to use it.

Educational institutions also play a critical role in promoting linguistic inclusion within AI-enhanced learning environments. Language programs can encourage critical discussions regarding language representation in AI systems, raise awareness of linguistic bias, and support the adoption of technologies that value multilingualism. Educators can help learners understand the sociopolitical dimensions of language and technology while promoting appreciation for linguistic diversity and cultural plurality. These

educational practices contribute to a more balanced and equitable integration of AI into language learning environments.

Essentially, the relationship between AI and language diversity represents one of the most significant ethical and educational challenges facing contemporary language education. Ensuring that AI supports rather than undermines linguistic diversity requires coordinated efforts among researchers, developers, educators, policymakers, and language communities themselves. The success of AI-enhanced language education should not be measured solely by technological sophistication or efficiency but also by its capacity to support multilingualism, protect linguistic heritage, promote cultural inclusion, and provide equitable educational opportunities for speakers of all languages. In an increasingly AI-mediated world, linguistic diversity must be viewed not as a technological obstacle but as a fundamental educational and *cultural resource* deserving active protection and support.

12.4 Inclusion, Accessibility, and Special Needs

One of the most promising aspects of Artificial Intelligence (AI) in language education is its potential to promote *inclusion* and *accessibility for diverse learner* populations. While much of the discourse surrounding AI focuses on ethical risks, bias, and inequality, AI technologies also offer significant opportunities to support learners who have traditionally encountered barriers to participation in educational environments. Through *adaptive learning systems, assistive technologies, speech recognition applications, personalized feedback mechanisms, and multilingual support tools,* AI has the capacity to make language education more *accessible, flexible,* and *responsive* to individual learner needs (Ng et al., 2021; Pan, 2024).

Inclusion in education is grounded in the principle that all learners, regardless of their physical, cognitive, linguistic, cultural, or socioeconomic backgrounds, should have equitable opportunities to participate in meaningful learning experiences. AI-supported educational systems can contribute markedly to this goal by providing individualized support that accommodates diverse learning profiles. Unlike traditional one-size-fits-all instructional models, AI technologies can analyze learner performance patterns and adapt instructional content, pacing, feedback, and learning pathways according to individual needs. Such personalization is particularly valuable in language education, where learners frequently demonstrate substantial

variation in proficiency levels, learning preferences, communication styles, and educational backgrounds.

For *learners with disabilities*, AI technologies offer numerous possibilities for enhancing participation and educational outcomes. Speech recognition systems can support learners with motor impairments by enabling voice-based interaction with educational platforms. Text-to-speech and speech-to-text technologies can assist learners with visual impairments, dyslexia, and reading difficulties by providing alternative pathways for accessing educational content. Similarly, AI-powered captioning and transcription systems can improve accessibility for deaf and hard-of-hearing learners by facilitating access to spoken language materials and classroom interactions (Ng et al., 2021). These technologies help reduce participation barriers and create more inclusive learning environments where diverse learners can engage more fully with language-learning activities.

A systematic review of AI applications for children with *special education needs*, conducted in the context of pandemic-era school closures, corroborates the pedagogical promise of these tools while sounding a caution consistent with the equity concerns raised elsewhere in this chapter: the review found that AI aided written language, reading, listening, memory, and arithmetic skills among learners with special needs, but noted that access barriers in developing-nation contexts constrained which learners could actually benefit (Bah & Artaria, 2020). Inclusion and the digital divide, in other words, are not separate problems to be solved in sequence; for a learner with a disability in an under-resourced setting, they are the same problem viewed from two sides.

The potential of AI extends beyond physical accessibility to support learners with diverse cognitive and learning needs. Adaptive learning systems can provide personalized scaffolding, differentiated instruction, and targeted feedback based on individual learner performance. Such systems may be particularly beneficial for learners who require additional support, extended practice opportunities, alternative instructional approaches, or individualized pacing. By continuously monitoring learner progress and identifying areas of difficulty, AI technologies can deliver timely interventions tailored to specific educational needs, thereby enhancing engagement, confidence, and learning success.

Recent research has highlighted the potential of AI to support neurodivergent learners through highly *individualized educational interventions*. AI-powered conversational agents, adaptive learning platforms, and intelligent tutoring systems can be customized to accommodate different cognitive profiles and communication preferences. For example, AI has been successfully employed to identify and recommend personalized communication strategies in autism education, enabling more effective learner engagement and interaction (Lampos et al., 2021). Similarly, generative AI systems can provide individualized scaffolding, structured guidance, and low-pressure conversational environments that reduce language performance anxiety and support learners who may struggle within traditional classroom contexts (Ronksley-Pavia et al., 2025; Xiao et al., 2025).

Teacher acceptance of these applications appears to hinge on a narrower factor than general enthusiasm for AI. A survey and interview study of Moroccan preservice teachers, using an extended technology-acceptance model to examine attitudes toward AI for teaching students with ADHD, found that *perceived usefulness* was the strongest predictor of acceptability (Amouri et al., 2025), suggesting that teacher buy-in for special-needs applications will track demonstrated classroom value rather than novelty or institutional mandate. Learners' own assessments introduce a further qualification. 69 Hong Kong secondary students in an AI-literacy programme viewed generative AI therapy chatbots as accessible and valued them as a source of supplementary support, but explicitly identified them as lacking genuine empathy and warned against over-reliance on them (Tse, 2026). Accessibility gains of this kind, in short, are real but bounded: useful supplements that learners themselves do not mistake for substitutes.

AI also offers significant opportunities to support educators working in inclusive classrooms. Generative AI tools are increasingly being used to assist teachers in designing *differentiated instructional materials, adapting learning resources, and and drafting individualized education plans* (IEPs) for learners with special educational needs—an application area that remains under-researched and requires careful human oversight. Such applications may enhance teachers' ability to provide individualized support while reducing administrative workload. On this basis, AI has the potential not only to improve learner outcomes but also to strengthen the capacity of educators to create more inclusive educational environments.

The benefits of AI-supported *accessibility* extend beyond disability-focused applications. Learners from diverse linguistic and cultural backgrounds may also benefit from *AI-powered translation tools, multilingual conversational agents, adaptive language-learning platforms, and culturally responsive educational resources*. These technologies can facilitate participation among multilingual learners by helping them navigate unfamiliar linguistic environments and access educational content more effectively. As educational institutions become increasingly diverse, such tools may contribute to more inclusive and equitable learning experiences for students from a wide range of backgrounds.

Despite these opportunities, researchers caution that accessibility benefits are not automatically guaranteed. AI systems themselves may generate new forms of exclusion if they are not designed with diverse users in mind. Educational technologies developed primarily for dominant populations may fail to account for the needs of learners with disabilities, speakers of underrepresented languages, or individuals from marginalized communities (Song et al., 2024). Seen in this light, accessibility and inclusion must be integrated into every stage of AI development and implementation rather than treated as secondary considerations.

The framework of *Universal Design for Learning (UDL)* provides a valuable foundation for guiding inclusive AI implementation. UDL emphasizes the provision of multiple means of engagement, representation, and expression to accommodate learner diversity. AI technologies can support these principles by offering flexible pathways for accessing content, demonstrating understanding, and participating in learning activities. Researchers increasingly argue that responsible AI integration should be explicitly aligned with UDL principles to ensure accessibility for all learners, particularly those with disabilities and special educational needs (Jafarian & Kramer, 2025). *AI-powered text simplification tools, customizable text-to-speech systems, adaptive feedback mechanisms, and multimodal content delivery* represent practical examples of how AI can support UDL-oriented educational environments (Kohnke, Zou, Su, et al., 2025).

Personalization frameworks developed specifically for language instruction reinforce this UDL-oriented direction. A conceptual synthesis proposes integrating Gardner's theory of *multiple intelligences* with AI tools such as chatbots in EFL instruction, arguing that doing so allows feedback

to be personalized in ways that enhance *engagement, motivation, and inclusivity for learners with different profiles* (Pitychoutis & Al Rawahi, 2024). Implementation, however, remains uneven even where the intent is explicitly inclusive: an exploratory study of 15 ESP teachers in Moroccan higher education found that integrating a 6Cs framework with inclusive AI learning design was hindered in practice by low student motivation, material limitations, and time constraints (Belkbir, 2024)—a reminder that inclusive design principles compete with the same resource constraints already discussed as a driver of digital divides more broadly.

Another important consideration concerns the accessibility of AI technologies themselves. While AI-powered tools may enhance inclusion for some learners, unequal access to devices, internet connectivity, assistive technologies, and digital literacy training may prevent others from benefiting from these innovations. Accessibility must therefore be understood not only in terms of individual learner characteristics but also in relation to broader issues of technological access, affordability, and educational equity. Educational institutions must address both pedagogical and infrastructural barriers if they are to realize the inclusive potential of AI.

Institutional responsibility is particularly important in this regard. Increasingly, educational policies emphasize that if a specific AI tool is inaccessible to a student because of disability-related barriers, financial limitations, or technological constraints, institutions should provide equitable alternatives that allow all learners to participate fully in educational activities (Jin et al., 2025). Such policies recognize that inclusion requires more than technological innovation; it also requires proactive efforts to ensure equitable access and participation.

Teacher preparation represents another essential component of inclusive AI implementation. Educators require sufficient knowledge and training to utilize AI technologies effectively within diverse learning environments. AI literacy should therefore include awareness of accessibility features, inclusive design principles, assistive technologies, and strategies for evaluating whether AI tools adequately support diverse learner populations. Teachers play a critical role in ensuring that technological innovations are implemented in ways that expand rather than restrict participation.

Recent scholarship increasingly advocates human-centered approaches to AI development that prioritize accessibility, inclusion, learner well-being,

and social responsibility (Alfredo et al., 2024; Ma et al., 2025). Such approaches recognize that technological effectiveness should be evaluated not only according to performance metrics but also according to the extent to which systems support meaningful participation and educational opportunity. In addition, researchers emphasize that high-risk AI applications should be subject to rigorous *documentation, evaluation,* and *certification processes* to ensure that they do not inadvertently marginalize already underserved populations (Nyaaba & Zhai, 2025; Zhai, Krajcik, & Pellegrino, 2021).

Therefore, AI possesses considerable potential to support more *inclusive* and *accessible language education*. Through adaptive learning, assistive technologies, personalized feedback, multilingual support, and specialized interventions for learners with disabilities and special educational needs, AI can help address barriers that have historically limited educational participation. However, realizing these benefits requires careful attention to accessibility, equity, design, governance, and implementation. *Inclusive AI* is not merely a technological achievement but an ethical commitment to ensuring that all learners—regardless of their abilities, backgrounds, or circumstances—can participate meaningfully in educational opportunities. The discussion of inclusion and accessibility therefore naturally leads to broader questions of governance and regulation. As AI technologies become increasingly integrated into educational systems, institutions and policymakers must establish frameworks that ensure ethical, equitable, and accountable implementation. The following section examines the policy, institutional guidelines, and regulatory approaches shaping the future of AI in language education.

12.5 Institutional AI Policies

The rapid proliferation of Artificial Intelligence (AI) and Generative Artificial Intelligence (GenAI) in education has compelled educational institutions worldwide to develop formal policies governing their use. As AI technologies become increasingly integrated into language teaching, learning, and assessment, universities, schools, and language centers face the challenge of balancing innovation with ethical responsibility, academic integrity, educational quality, and learner well-being. Institutional AI policies have therefore emerged as a critical mechanism for providing guidance,

establishing accountability, and ensuring that AI technologies are used in ways that align with educational objectives and values (Jin et al., 2025; Pan, 2024).

Initially, many educational institutions responded to the emergence of generative AI with uncertainty and concern. The public release of powerful language models such as ChatGPT generated widespread debates regarding *plagiarism, academic misconduct, authorship, assessment validity,* and *the future role of educators.* Early institutional responses often focused on *restriction* or *prohibition*. However, as AI technologies became increasingly accessible and embedded within everyday educational practices, many institutions shifted toward more nuanced approaches that emphasize *responsible, transparent, and pedagogically meaningful use* rather than outright bans (Kasneci et al., 2023). This transition reflects a growing recognition that AI is likely to remain a permanent component of educational ecosystems and that governance strategies must focus on adaptation rather than resistance.

Recent research indicates that universities worldwide are increasingly moving away from *reactionary restrictions* toward *adaptive policy frameworks* that support responsible innovation while protecting educational values (Belkina et al., 2025). Drawing upon the *Diffusion of Innovations Theory (DIT)*, institutional leaders increasingly recognize that successful AI adoption depends on factors such as compatibility with existing educational practices, opportunities for experimentation and trialability, and clear communication channels that support stakeholder engagement (Mustofa et al., 2025). As a result, effective institutional policies seek not only to regulate AI use but also to facilitate informed and constructive integration across teaching, learning, and assessment contexts.

Contemporary institutional AI policies generally pursue several interconnected objectives. *First,* they establish clear expectations regarding appropriate and inappropriate uses of AI technologies by students and educators. *Second*, they provide guidance concerning academic integrity, disclosure requirements, authorship, and attribution. *Third*, they address issues related to privacy, data protection, and ethical AI use. *Fourth*, they support the development of AI literacy among educational stakeholders by encouraging informed and critical engagement with emerging technologies (Jin et

al., 2025). Collectively, these objectives help institutions navigate the opportunities and risks associated with AI-enhanced education.

One of the most important functions of institutional policy is *clarifying the role of AI* within academic work. In language education, AI may be used for *idea generation, language practice, translation, grammar correction, feedback provision, content creation, and assessment preparation*. Without clear guidelines, learners may be uncertain about the extent to which AI assistance is acceptable. In turn, many institutions have adopted transparency-based approaches that require students to disclose AI use while allowing educators flexibility in determining appropriate levels of AI integration for specific assignments and learning outcomes (Pan, 2024). Such approaches recognize that AI use exists on a continuum rather than as a simple distinction between acceptable and unacceptable behavior.

For example, using AI to *brainstorm ideas, receive formative feedback,* or *generate practice exercises* may be viewed as pedagogically beneficial, whereas submitting AI-generated content as entirely original work may violate academic integrity standards. Effective institutional policies therefore distinguish among different forms of AI use rather than treating all AI-assisted activities identically. This differentiation enables institutions to promote innovation while maintaining academic standards and learner accountability.

Survey evidence indicates that the demand for such differentiated guidance is now shared across the institutional hierarchy rather than confined to administrators. A survey of 329 students and 14 faculty at a Portuguese university found genuinely divergent views on generative AI's effect on skill development—roughly half of each group disagreed that it harms competency development—yet found consensus, cutting across that disagreement, on the need for *clear institutional guidelines* (Fernandes et al., 2025). What such guidelines should prioritize is itself now measurable: a mixed-methods study of 628 stakeholders in Sichuan, China, found that *policy clarity, transparency, and human oversight* most strongly predicted *institutional readiness* for AI-enabled learning governance (Che & Perkins, 2026).

Academic integrity remains a central concern within institutional AI frameworks. Universities worldwide have developed policies that emphasize transparency regarding AI involvement in academic work. For instance,

Tecnológico de Monterrey requires students to disclose and appropriately cite any use of AI tools, while explicitly prohibiting the presentation of AI-generated content as wholly original work (Jin et al., 2025). Similar approaches have emerged across numerous institutions seeking to balance educational innovation with academic honesty. Such policies reflect a broader shift from attempting to eliminate AI use toward fostering ethical and transparent engagement with AI technologies.

The emergence of generative AI has also prompted institutions to reconsider traditional approaches to assessment. Concerns regarding "*chatting and cheating*" have accelerated discussions surrounding assessment redesign and evaluation practices (Cotton et al., 2023). Increasingly, educational institutions are moving away from assessment models that rely exclusively on final written products and toward *process-oriented, authentic, and performance-based approaches.* These revised assessment frameworks emphasize *critical thinking, reflective practice, collaborative learning,* and *evidence of learning processes* rather than solely evaluating completed outputs. Such changes align with broader efforts to ensure that assessment remains meaningful in AI-rich educational environments.

Data privacy and *information security* constitute another key dimension of institutional AI policy. Educational institutions bear significant responsibility for protecting learner information and ensuring compliance with ethical and legal standards governing data use. Policies increasingly address issues such as i*nformed consent, cybersecurity, data retention, third-party platform use,* and *learner privacy* (Alfredo et al., 2024). Given the widespread use of commercial AI platforms, institutions must carefully evaluate the privacy practices of technology providers and establish safeguards that protect educational data.

A notable example can be found at the University of California, Berkeley, where institutional guidance operates under the assumption that no sensitive, confidential, or personally identifiable information should be entered into commercial large language models (Jin et al., 2025). Such policies recognize the potential risks associated with transmitting educational data to external platforms and emphasize the importance of maintaining strict boundaries around sensitive information. Similar policies have emerged globally as institutions seek to balance the benefits of AI-supported learning with the need to safeguard learner rights and privacy.

Institutional policies also increasingly recognize the importance of *educator preparedness* and *professional development*. Teachers play a central role in the effective implementation of AI-enhanced pedagogies, yet many educators report uncertainty regarding AI capabilities, limitations, and ethical implications. For this reason, AI literacy development has become a major component of institutional AI strategies. Professional development initiatives seek to equip educators with the knowledge and skills necessary to evaluate AI-generated content, redesign assessments, guide students in responsible AI use, and critically interpret AI-supported educational data (Ng et al., 2021; Ma et al., 2025).

The role of institutional policy extends beyond regulation toward *pedagogical transformation.* Some universities explicitly view AI as an opportunity to re-evaluate and modernize educational practices. For example, Nanyang Technological University frames AI not simply as a challenge to existing assessment systems but as a catalyst for rethinking teaching, learning, and evaluation practices more broadly (Jin et al., 2025). This perspective reflects a growing understanding that AI adoption should not simply preserve traditional educational models but also support innovation and pedagogical renewal.

Crucially, institutional AI policies cannot be static documents. AI technologies continue to evolve rapidly, introducing new capabilities, opportunities, and ethical challenges. Policies that are overly restrictive may quickly become obsolete, while excessively permissive approaches may fail to address emerging risks. Researchers therefore advocate iterative policy development processes characterized by *regular review, stakeholder consultation,* and *ongoing evaluation* of AI-related practices (Jin et al., 2025; McDonald et al., 2025). Such adaptive approaches allow institutions to remain responsive to technological change while maintaining educational integrity and accountability.

Transparency and stakeholder participation are essential components of effective policy development. Successful institutional guidelines should incorporate input from educators, students, administrators, technology specialists, legal experts, and other relevant stakeholders. Such collaborative approaches help ensure that policies reflect diverse perspectives, address practical concerns, and remain responsive to the evolving needs of educational communities. At the same time, participatory policy development

can strengthen trust, legitimacy, and institutional acceptance of AI governance measures.

As we have already stated, institutional AI policies serve as a bridge between technological innovation and educational responsibility. By establishing clear expectations, supporting ethical practices, protecting learner rights, promoting transparency, and cultivating AI literacy, these policies help educational institutions navigate the opportunities and challenges associated with AI-enhanced language education. As AI becomes increasingly embedded within teaching, learning, and assessment, robust institutional frameworks will remain essential for ensuring that technological advancement contributes positively to educational quality, academic integrity, and learner development.

12.5.1 Assessment Regulations

The integration of AI into educational assessment has created significant regulatory challenges for educational institutions, accreditation agencies, and policymakers. Traditional assessment frameworks were designed around the assumption that student work was produced independently and that authorship could be clearly attributed to individual learners. The emergence of GenAI has disrupted these assumptions by enabling students to generate essays, translations, summaries, responses, and other forms of academic work with varying degrees of AI assistance. It follows that institutions are increasingly revising assessment regulations to address questions of authorship, transparency, validity, reliability, and academic integrity while recognizing the growing role of AI in contemporary learning environments (Pan, 2024; Jin et al., 2025).

One of the most significant regulatory concerns involves distinguishing between *legitimate educational support* and inappropriate academic assistance. AI technologies can provide grammar correction, vocabulary suggestions, translation support, brainstorming assistance, and formative feedback that may enhance learning outcomes. At the same time, these systems can generate complete assignments that learners may submit with minimal personal contribution. As a result, institutions face the challenge of determining which forms of AI use are pedagogically acceptable and which constitute violations of academic integrity policies. Increasingly, educational organizations are moving away from *blanket prohibitions* toward frameworks that emphasize transparency, disclosure, and responsible use (Kasneci et al., 2023).

Many institutions now require students to disclose their use of AI tools in assessment tasks. Such disclosure policies recognize that AI-assisted learning is becoming a normal component of educational practice while maintaining the importance of transparency in academic work. Similar to citation practices for traditional sources, students may be expected to identify when and how AI tools contributed to the completion of assignments. These approaches aim to *preserve academic honesty* while acknowledging that AI has become part of the broader educational ecosystem (Pan, 2024).

The rise of AI has also prompted renewed attention to *assessment validity*. If AI systems can successfully complete tasks originally intended to measure student learning, educators must reconsider whether those tasks continue to assess the intended knowledge, skills, and competencies. Traditional take-home essays, written assignments, and certain forms of online assessment may no longer provide reliable evidence of individual learner achievement when extensive AI assistance is available. Accordingly, researchers increasingly advocate assessment redesign strategies that prioritize authentic, process-oriented, and higher-order learning outcomes (Ng et al., 2025).

Language education provides a particularly relevant context for these developments. AI systems are now capable of generating coherent written texts, providing language corrections, translating content, simulating conversational interactions, and even evaluating learner performance. While these capabilities offer valuable learning opportunities, they also complicate efforts to evaluate learners' independent language proficiency. Assessment regulations therefore increasingly emphasize *process documentation, iterative drafting, oral defense components, reflective writing, portfolio assessment, and classroom-based performance tasks* that make learners' thinking processes more visible (Pan, 2024).

Process-oriented assessment approaches have gained considerable attention in AI-enhanced educational environments. Rather than focusing exclusively on final products, educators increasingly seek evidence of learners' decision-making processes, revision histories, prompt development strategies, and critical engagement with AI-generated content. Such approaches align with contemporary perspectives on assessment for learning and provide richer insights into learner development than product-focused evaluation alone (van Niekerk et al., 2025). Assessment regulations are therefore

evolving to accommodate new forms of evidence that reflect the realities of AI-assisted learning.

The emergence of Large Language Models (LLMs) has also transformed automated assessment practices. Recent advances in automated short-answer grading (ASAG), automated essay scoring, and AI-supported evaluation systems have demonstrated considerable potential for increasing efficiency and scalability (Jung et al., 2025; Rodrigues et al., 2025). AI-driven scoring engines can provide rapid feedback, reduce grading workloads, and deliver highly consistent evaluations, thereby addressing challenges associated with human fatigue and scoring variability (De Vrindt et al., 2025). These capabilities make AI particularly attractive for formative assessment contexts where timely feedback is essential.

However, questions concerning the validity and reliability of AI-generated scores remain significant. Empirical studies indicate that models such as GPT-4 frequently exhibit a tendency to overestimate student performance compared with human raters, resulting in only moderate levels of agreement with expert evaluations (Rodrigues et al., 2025). Although AI systems often demonstrate high internal consistency, consistency alone does not guarantee accurate measurement of learner competence. Consequently, assessment regulations increasingly emphasize that AI-generated scores should not be accepted uncritically, particularly in high-stakes educational contexts.

For this reason, researchers and policymakers advocate *human-centered regulatory frameworks* that require meaningful human oversight of AI-supported assessment processes. AI scoring systems may serve valuable functions in formative assessment, where the speed and frequency of feedback are prioritized over final grading decisions (Dai et al., 2025). In contrast, high-stakes summative assessments require rigorous calibration procedures, human review mechanisms, and final validation by qualified educators to ensure fairness, validity, and educational justice (Rodrigues et al., 2025). Such approaches align with broader human-in-the-loop models that position AI as a decision-support tool rather than an autonomous evaluator.

Another important regulatory issue concerns the reliability of AI-detection technologies. In response to concerns regarding AI-generated submissions, many institutions initially adopted AI-content detection tools as mechanisms for identifying potential misuse. However, a growing body of

research has demonstrated that these systems frequently produce unreliable results and may generate false positives, particularly for multilingual learners, non-native English speakers, and neurodiverse students (Liang et al., 2023; Pan, 2024; Turós et al., 2025). Simplified linguistic structures, formulaic writing styles, and language patterns common among second-language writers are often incorrectly classified as AI-generated, raising serious concerns regarding fairness and due process.

Henceforth, educational institutions are increasingly cautious about relying solely on automated detection tools when making academic integrity decisions. Emerging regulatory frameworks emphasize the importance of human review, procedural fairness, multiple forms of evidence, and transparent investigation procedures when allegations of AI-related misconduct arise. These safeguards help protect students from wrongful accusations while maintaining confidence in academic integrity processes.

Assessment regulations must also address issues of *equity* and *access*. Learners differ substantially in their access to AI technologies, digital literacy skills, technological resources, and familiarity with AI-assisted learning strategies. Policies that fail to acknowledge these differences risk creating unintended inequities. On this basis, equitable assessment regulations require institutions to consider issues of technological access, learner support, transparency, and fairness when establishing expectations regarding AI use (Belkina et al., 2025).

The growing adoption of AI has further encouraged institutions to reconsider the competencies they seek to assess. Increasingly, educators argue that learners should not only demonstrate disciplinary knowledge but also the ability to engage critically, ethically, and effectively with AI technologies. From this perspective, competencies such as AI literacy, critical evaluation of AI-generated outputs, prompt design, ethical decision-making, and responsible technology use become legitimate educational outcomes in their own right (Long & Magerko, 2020; Ma et al., 2025). Assessment regulations must therefore evolve to reflect changing conceptions of knowledge, expertise, and competence in AI-mediated learning environments.

Respectively, effective assessment regulation in the age of AI requires a careful balance between innovation and accountability. Educational institutions must protect academic integrity and assessment validity while recognizing the legitimate educational value of AI technologies. Rather than

attempting to eliminate AI from assessment contexts, contemporary regulatory approaches increasingly focus on transparency, authenticity, human oversight, process-oriented evaluation, and the cultivation of learners' critical AI competencies. Such frameworks acknowledge that AI is becoming an enduring component of education and that assessment practices must evolve accordingly to remain valid, reliable, equitable, and educationally meaningful in increasingly AI-mediated learning environments.

12.5.2 National and International Guidelines

The rapid expansion of Artificial Intelligence (AI) in education has prompted governments, international organizations, accreditation agencies, and regulatory bodies to develop policy frameworks that support innovation while addressing ethical, legal, and educational concerns. As AI increasingly influences language teaching, learning, and assessment, national and international guidelines have become essential mechanisms for promoting responsible implementation, protecting learner rights, and ensuring that technological developments align with broader educational goals. These frameworks provide strategic direction for educational institutions while helping stakeholders navigate the complex opportunities and challenges associated with AI-enhanced learning environments (Jin et al., 2025; Ng et al., 2025).

One of the most influential developments in recent years has been the growing recognition that AI governance must be fundamentally human-centered. International organizations increasingly emphasize that educational technologies should support human development, learner well-being, social inclusion, and educational equity rather than focusing exclusively on technological efficiency or economic competitiveness. Human-centered approaches position learners and educators as active participants in AI-enhanced educational environments and stress the importance of maintaining human agency, transparency, accountability, and ethical responsibility throughout AI implementation processes (Kasneci et al., 2023; Ma et al., 2025).

Among international organizations, UNESCO has played a particularly significant role in shaping global discussions of AI governance in education. UNESCO's recommendations on the ethics of Artificial Intelligence and its subsequent guidance on Generative AI in Education and Research emphasize ethical AI development, protection of human rights, equitable access to educational technologies, transparency, accountability, data

privacy, and inclusion (UNESCO, 2022; Jin et al., 2025). These principles are especially relevant for language education because AI systems frequently process highly personal forms of communication and interact with learners from diverse linguistic and cultural backgrounds. UNESCO's framework highlights the importance of ensuring that AI technologies respect cultural diversity, support multilingualism, and contribute to sustainable educational development rather than reinforcing existing inequalities.

Similarly, the Organisation for Economic Co-operation and Development (OECD) has promoted principles centered on trustworthy AI, transparency, accountability, robustness, fairness, and human oversight. OECD guidelines encourage policymakers and educational institutions to evaluate not only the technical performance of AI systems but also their broader social, ethical, and educational consequences. Such recommendations recognize that AI technologies can influence learning opportunities, educational decision-making, and social equity in ways that extend far beyond their immediate instructional functions. For language education, these considerations are particularly important because language learning is closely connected to issues of identity, culture, participation, and social mobility.

Beyond educational organizations, governments and supranational institutions have begun establishing comprehensive regulatory frameworks to govern AI deployment. One of the most significant developments has been the introduction of the European Union's AI Act, which establishes harmonized, risk-based regulations for artificial intelligence across member states (European Union, 2024). Notably, the AI Act classifies education as a high-risk domain due to the potential impact of AI systems on learners' educational opportunities and life outcomes. Seen in this light, educational AI applications are subject to heightened requirements concerning transparency, accountability, documentation, human oversight, and explainability (European Union, 2024; Hooshyar et al., 2025; Nyaaba & Zhai, 2025). Such regulations represent an important step toward ensuring that educational AI systems remain trustworthy, fair, and accountable.

Similarly, the United States has adopted a policy-oriented approach through executive governance mechanisms. The White House Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence emphasizes the importance of protecting citizens from potential harms while promoting innovation and economic growth (The

White House, 2023). Although educational policy remains largely decentralized in the United States, such initiatives provide strategic guidance for educational institutions seeking to balance technological innovation with ethical responsibility.

National governments increasingly translate international principles into educational strategies and policy frameworks. Across many countries, ministries of education are developing AI strategies that address curriculum reform, teacher preparation, digital infrastructure, assessment practices, ethical governance, and AI literacy development. While policy approaches vary considerably across national contexts, most contemporary frameworks emphasize several common priorities: responsible AI adoption, learner protection, educational innovation, digital inclusion, AI literacy, and equitable access to technological resources (Jin et al., 2025). These priorities reflect growing recognition that AI integration requires coordinated efforts across multiple dimensions of educational systems.

A central component of many national AI strategies involves the development of AI literacy. Policymakers increasingly regard AI literacy as a foundational competency for participation in contemporary society. Beyond technical understanding, AI literacy encompasses critical evaluation, ethical awareness, responsible use, digital citizenship, and the ability to collaborate effectively with intelligent technologies (Long & Magerko, 2020; Ng et al., 2021). As a result, several national frameworks advocate integrating AI-related competencies into school curricula, teacher education programs, and professional development initiatives. Such efforts aim to prepare learners and educators for meaningful participation in AI-mediated educational and professional environments.

Despite these advances, researchers have identified a substantial *implementation gap* between policy aspirations and educational practice. While many governments have mandated the inclusion of AI ethics, digital citizenship, and AI literacy within educational frameworks, these topics often receive limited instructional time and insufficient practical implementation within classrooms (Ma et al., 2025). Therefore, policy commitments do not always translate into meaningful educational experiences. This discrepancy highlights the need for stronger *institutional support, curriculum integration, and teacher preparation* to ensure that policy objectives are realized in practice.

The size of this gap is now being measured directly rather than only inferred. A professional-development intervention involving 452 K-12 educational professionals found that 88% expressed willingness to adopt AI chatbots in their teaching, even though only 31% had actually used them—a 57-point gap between stated intention and practice within a single sample (Bergdahl & Sjöberg, 2025). A comparative survey of 116 Chinese educators located a related but distinct gap: university teachers scored significantly higher than K-12 counterparts on AI-ethics literacy, with professional development emerging as the strongest predictor of ethical awareness across both groups (Zhang & Rhee, 2025). The pattern is not confined to education. An empirical analysis of Japanese corporate AI disclosures found a comparable gap between organizations' stated ethical commitments and their actual implementation, informing a proposed Tripartite Ethical Policy Framework for AI governance (Nakano, 2026); if a similar dynamic holds in education, the aspiration–practice gap identified above is not a transitional artifact of systems still catching up with new technology, but a more durable feature of how organizations everywhere adopt ethical AI commitments.

International guidelines increasingly emphasize multi-stakeholder collaboration as a prerequisite for successful AI governance. Effective implementation requires *cooperation among governments, educational institutions, accreditation bodies, technology developers, researchers, educators, learners,* and *civil society organizations* (Ng et al., 2025). Such collaborative approaches help ensure that AI governance reflects diverse perspectives and remains responsive to rapidly evolving technological developments.

Accreditation agencies and *quality assurance organizations* are also assuming more prominent roles in AI governance. Increasingly, these bodies evaluate institutions not only on the basis of technological infrastructure but also according to their capacity-building initiatives, faculty development programs, AI literacy training, ethical governance mechanisms, and institutional readiness for responsible AI adoption (Guan et al., 2025; Pei et al., 2025). This shift reflects a broader understanding that responsible AI implementation depends not merely on access to technology but also on organizational culture, professional expertise, and ethical leadership.

Data governance has emerged as another key area of national and international regulation. Given the extensive collection and processing of learner

data within AI-supported educational environments, policymakers increasingly focus on informed consent, data protection, cybersecurity, algorithmic transparency, and learner privacy (Akgun & Greenhow, 2022). Effective governance frameworks seek to balance educational innovation with the protection of fundamental rights related to autonomy, privacy, and human dignity.

Equity and *inclusion* remain equally important dimensions of AI governance. Policymakers increasingly recognize that AI technologies may either reduce or exacerbate educational inequalities depending on how they are designed and implemented. In turn, many national and international frameworks emphasize equitable access to digital resources, support for underrepresented populations, multilingual AI development, accessible educational design, and inclusive technological innovation (Song et al., 2024; Ma et al., 2025). These priorities are particularly relevant for language education, where linguistic diversity, cultural representation, and equitable participation constitute fundamental educational concerns.

Researchers further emphasize that institutional policies must clearly define the appropriate extent of AI usage within educational environments. Without clear guidance, students and educators may struggle to distinguish between acceptable educational assistance and inappropriate dependence on AI-generated outputs (Aldreabi et al., 2025). Effective guidelines therefore require explicit expectations regarding transparency, attribution, academic integrity, and ethical engagement with AI technologies.

Another increasingly important principle is *adaptive governance*. AI technologies evolve rapidly, often outpacing traditional regulatory mechanisms. Static policy frameworks may therefore become outdated quickly as new capabilities, risks, and educational applications emerge. Scholars increasingly advocate iterative governance models characterized by continuous monitoring, stakeholder consultation, evidence-based revision, and regulatory flexibility (Jin et al., 2025; Ng et al., 2025). Such approaches enable policymakers and institutions to respond proactively to technological developments while maintaining ethical safeguards and educational accountability.

Despite substantial progress, important challenges remain. Regulatory frameworks vary significantly across countries, creating inconsistencies in how AI is governed and implemented. Differences in technological

capacity, economic resources, educational priorities, legal traditions, and political contexts contribute to highly diverse policy landscapes. Beyond this, many educational institutions continue to struggle with translating broad ethical principles into practical classroom guidelines, assessment procedures, and institutional practices. These challenges underscore the continuing need for collaboration among policymakers, researchers, educators, technology developers, accreditation agencies, and international organizations.

As such, national and international guidelines provide an essential foundation for responsible AI implementation in language education. By establishing shared principles related to transparency, accountability, fairness, inclusion, human oversight, and educational equity, these frameworks help ensure that technological innovation serves educational rather than purely technological objectives. As AI continues to reshape language teaching, learning, and assessment, robust governance structures will remain critical for protecting learner interests, supporting educational quality, promoting ethical innovation, and ensuring equitable access to the benefits of technological advancement across diverse global educational contexts.

12.5.3 Role of Ministries, Universities, and Accreditation Bodies

The successful integration of Artificial Intelligence (AI) into language education depends not only on classroom practices and institutional policies but also on the broader governance structures that guide educational systems. Ministries of education, universities, accreditation agencies, and quality assurance bodies play critical roles in shaping how AI technologies are adopted, regulated, evaluated, and sustained. As AI continues to transform language teaching, learning, and assessment, these institutions increasingly function as key actors responsible for balancing innovation with ethical responsibility, educational quality, and public accountability (Jin et al., 2025; Pan, 2024).

At the *national level*, ministries of education are responsible for establishing strategic directions for AI integration across educational sectors. Their responsibilities include developing national AI-in-education frameworks, allocating resources for technological infrastructure, supporting teacher professional development, and ensuring that AI adoption aligns with broader educational goals. Ministries also play an essential role in reducing inequalities by promoting equitable access to digital technologies and

supporting institutions that may lack sufficient technological resources. In many countries, national AI strategies increasingly emphasize digital transformation, responsible AI governance, and the development of future-ready competencies that prepare learners to participate effectively in AI-mediated societies (Ng et al., 2025).

Beyond infrastructure and funding, ministries are responsible for establishing regulatory frameworks that protect learners, educators, and educational institutions. These frameworks frequently address issues such as data privacy, cybersecurity, ethical AI use, academic integrity, algorithmic accountability, and transparency. Because educational institutions increasingly rely on commercial AI platforms and third-party service providers, ministries must ensure that procurement and implementation processes comply with national legal standards and international ethical guidelines. Effective governance therefore requires collaboration among educational authorities, technology providers, legal experts, and researchers to develop policies that are both innovative and ethically sound (Akgun & Greenhow, 2022).

International organizations have further reinforced the strategic role of ministries in AI governance. UNESCO and the OECD advocate national frameworks that position AI literacy as a core educational competency and encourage governments to align technological innovation with principles of equity, human rights, transparency, and accountability (Ho & Lee, 2025; UNESCO, 2024). Such frameworks emphasize that AI literacy should extend beyond technical proficiency to include ethical reasoning, critical evaluation of AI outputs, and awareness of issues related to bias, privacy, and social responsibility (Kılıç et al., 2025; Cerratto Pargman et al., 2025). For this reason, ministries increasingly bear responsibility for ensuring that AI literacy becomes systematically embedded within school curricula, teacher education programs, and professional development initiatives.

Universities occupy a particularly influential position within AI governance because they serve simultaneously as educational institutions, research centers, innovation hubs, and professional training providers. In the context of language education, universities are often at the forefront of experimenting with AI-enhanced pedagogies, learning analytics systems, automated assessment tools, intelligent tutoring systems, and generative AI applications. It follows that they play a central role in generating evidence-

based practices that inform educational policy and technological innovation (Kasneci et al., 2023).

One of the most important responsibilities of universities is the development and implementation of institutional policies governing AI use. As discussed in previous sections, universities increasingly establish guidelines addressing academic integrity, assessment redesign, AI disclosure requirements, responsible technology use, and ethical engagement with generative AI systems. However, their role extends beyond regulation. Universities also contribute to AI governance through interdisciplinary research, curriculum development, faculty training, and professional learning initiatives that help educators and students develop AI literacy and critical technological awareness (Long & Magerko, 2020; Ma et al., 2025).

Teacher education represents a particularly important area of university responsibility. Future language teachers require the knowledge and skills necessary to integrate AI technologies effectively while addressing associated ethical, pedagogical, and social challenges. Researchers increasingly argue that AI literacy should become a core component of teacher preparation programs. Such training should extend beyond technical proficiency to include critical evaluation of AI systems, awareness of algorithmic bias, understanding of data governance issues, ethical decision-making, and strategies for promoting responsible AI use in educational settings (Ng et al., 2021; Mishra et al., 2023). In this respect, universities function as key institutions for preparing educators capable of navigating increasingly AI-mediated educational environments.

Universities also play an important role in bridging global policy frameworks and local educational realities. By adapting international recommendations to specific institutional and cultural contexts, universities can help ensure that AI implementation remains pedagogically meaningful and socially responsive. Through research partnerships, pilot projects, and community engagement initiatives, universities contribute to the development of AI practices that are evidence-based, context-sensitive, and aligned with educational values.

Accreditation bodies and quality assurance agencies constitute another crucial component of AI governance. These organizations establish standards that educational institutions must meet to demonstrate quality, accountability, and continuous improvement. As AI becomes increasingly embedded

in teaching and assessment, accreditation agencies are beginning to evaluate how AI-related practices should be incorporated into quality assurance frameworks. This includes examining institutional policies, assessment procedures, academic integrity measures, data governance practices, accessibility provisions, and support structures for AI literacy development.

The involvement of accreditation agencies is particularly important because AI technologies have significant implications for educational quality and assessment validity. Automated assessment systems, generative AI tools, and learning analytics platforms can influence how learning outcomes are measured, interpreted, and reported. Quality assurance frameworks must therefore ensure that AI-supported practices remain aligned with educational standards and do not compromise the reliability, validity, fairness, or transparency of assessment processes (Pack et al., 2024; Pan, 2024).

Recent studies suggests that accreditation bodies must move beyond evaluating final assessment products and begin recognizing the importance of human–AI collaborative learning processes. Traditional assessment models focused primarily on final outputs may no longer adequately capture learner competencies in AI-enhanced environments. Accordingly, quality assurance frameworks increasingly emphasize critical thinking, ethical decision-making, reflective practice, and learners' ability to effectively guide and evaluate AI-generated outputs (Oliveira et al., 2025). Such approaches align with broader educational trends that view AI as a collaborative learning partner rather than merely a technological tool.

Accreditation agencies can also encourage innovation by disseminating evidence-based recommendations, facilitating institutional learning, and supporting professional development initiatives. Rather than functioning solely as regulatory mechanisms, these bodies increasingly serve as facilitators of responsible innovation by promoting best practices, encouraging inter-institutional collaboration, and nurturing shared standards for AI governance. Such efforts contribute to greater consistency and accountability across educational systems while supporting institutions in adapting to rapidly evolving technological landscapes.

The governance of AI in language education increasingly requires collaboration among ministries, universities, accreditation agencies, technology developers, researchers, and international organizations. No single stakeholder possesses sufficient expertise or authority to address the full range

of challenges associated with AI implementation. Effective governance therefore depends on multi-stakeholder approaches that integrate technological, educational, ethical, legal, and sociocultural perspectives. Such collaboration supports the development of policies and practices that are responsive to emerging technologies while remaining grounded in educational values and societal needs.

Researchers increasingly advocate human-centered governance models that prioritize learner well-being, educational equity, social responsibility, and democratic participation (Alfredo et al., 2024; Ma et al., 2025). These models emphasize that AI should serve educational purposes rather than dictate them. Ministries, universities, and accreditation bodies therefore share a collective responsibility to ensure that AI technologies enhance learning opportunities while preserving human agency, professional judgment, and ethical accountability.

The successful regulation of AI ultimately requires coordinated action among all stakeholders. Universities cannot operate in isolation; they rely on ministries of education and accreditation bodies to provide strategic direction, regulatory guidance, and institutional support that align pedagogical innovation with ethical mandates (Jin et al., 2025). Similarly, policymakers depend on educational institutions and researchers to generate evidence that informs responsible governance. Building genuinely "AI-ready" educational systems therefore requires a socio-technical approach in which policymakers, educational leaders, researchers, and technology developers collaboratively design frameworks that ensure AI serves the diverse needs of learners across different linguistic, cultural, and socioeconomic contexts.

Ultimately, the future of AI in language education will be shaped not only by technological innovation but also by the effectiveness of the governance structures that guide its implementation. Ministries provide strategic leadership and regulatory oversight, universities contribute research and educational expertise, and accreditation agencies promote quality, accountability, and continuous improvement. Together, these institutions form the foundation of responsible AI governance, helping ensure that technological advancements contribute to equitable, ethical, transparent, and educationally meaningful outcomes for language learners worldwide.

12.6 Future Directions

The rapid evolution of Artificial Intelligence (AI) is transforming language education at a pace that challenges existing pedagogical practices, ethical frameworks, and policy structures. As AI technologies continue to advance, future developments will likely extend far beyond the current applications of automated feedback, intelligent tutoring systems, learning analytics, and generative AI. Consequently, the future of AI in language education will depend not only on technological innovation but also on the extent to which educational stakeholders can develop sustainable, ethical, and inclusive approaches to AI integration. Emerging research increasingly emphasizes that the long-term success of AI in education requires balancing innovation with human values, educational equity, and social responsibility (Kasneci et al., 2023; Ma et al., 2025).

One of the most important future directions involves the development of *ethical-by-design educational AI systems.* Traditionally, ethical concerns have often been addressed after technologies have already been deployed. However, researchers increasingly argue that ethical considerations should be embedded throughout the design process itself. Ethical-by-design approaches seek to integrate principles such as transparency, accountability, fairness, privacy protection, and inclusivity directly into AI systems from their earliest stages of development (Akgun & Greenhow, 2022). In language education, this means creating technologies that actively support learner autonomy, protect personal data, and minimize algorithmic bias rather than merely responding to ethical concerns after implementation.

A second critical direction concerns the advancement of *multilingual and culturally responsive AI systems.* Much of the current AI ecosystem remains heavily dominated by English-language data and Western educational perspectives. As discussed earlier in this chapter, such imbalances risk reinforcing linguistic inequalities and marginalizing less-resourced languages. Future AI development must therefore prioritize broader linguistic representation and cultural inclusivity. Researchers increasingly advocate the creation of AI models trained on diverse linguistic datasets that better reflect the multilingual realities of global education (Song et al., 2024; Ma et al., 2025). Such efforts would help ensure that AI technologies contribute to language preservation and linguistic diversity rather than accelerating language homogenization.

Future developments are also likely to strengthen the role of *human-centered AI* in educational environments. Rather than pursuing complete automation, contemporary research increasingly supports collaborative models in which humans and AI systems work together to enhance learning outcomes. Human-centered AI emphasizes that technological systems should augment rather than replace human expertise, judgment, and interpersonal interaction (Alfredo et al., 2024). In language education, this approach aligns with the recognition that language learning involves social, cultural, emotional, and communicative dimensions that cannot be fully captured through automated processes alone. Future educational systems will therefore likely emphasize human–AI partnerships that combine the efficiency of AI with the contextual understanding and empathy of human educators.

Another emerging priority is the expansion of *AI literacy as a core educational competency.* As AI technologies become increasingly integrated into everyday life, the ability to understand, evaluate, and use AI responsibly is becoming an essential form of literacy. Future language curricula may therefore incorporate AI literacy alongside traditional linguistic competencies. Learners will need to develop skills related to prompt design, critical evaluation of AI-generated content, ethical technology use, data awareness, and responsible participation in AI-mediated communication environments (Long & Magerko, 2020; Ng et al., 2021). Similarly, educators will require ongoing professional development to ensure that they can effectively integrate AI into pedagogically meaningful and ethically responsible practices.

The future of AI in language education will also be shaped by advances in *learning analytics and adaptive learning systems.* Emerging technologies are increasingly capable of analyzing learner interactions in real time and providing personalized support based on individual needs, preferences, and learning trajectories. Such systems may enable more responsive instructional environments that support diverse learners and facilitate continuous formative assessment (Alfredo et al., 2024). However, the expansion of these capabilities must be accompanied by robust safeguards related to privacy, transparency, and learner agency to prevent excessive surveillance and data misuse.

Policy development will remain another crucial area of future growth. As AI technologies evolve, governments, educational institutions, and international organizations will need to continuously revise regulatory frameworks to

address emerging challenges. Future policies will likely place greater emphasis on data governance, algorithmic accountability, AI-assisted assessment, intellectual property, and educational equity (Jin et al., 2025). Flexible and adaptive governance models will be particularly important because technological change often occurs more rapidly than traditional policy development processes.

Equity considerations will also remain central to future discussions. While AI offers significant opportunities for expanding access to educational resources, disparities in infrastructure, connectivity, digital literacy, and technological access continue to affect learners worldwide. Future efforts must therefore focus on reducing digital divides and ensuring that AI benefits are distributed equitably across different regions, institutions, and populations (Belkina et al., 2025). Addressing these challenges requires sustained investment in infrastructure, teacher training, and inclusive technology design.

Finally, future research will need to move beyond questions of technological effectiveness toward broader examinations of *AI's educational* and *societal impact.* Scholars increasingly call for interdisciplinary investigations that explore how AI influences learner identity, language practices, cultural representation, social relationships, and educational equity. Such research is essential for understanding the long-term implications of AI integration and for ensuring that educational technologies remain aligned with human and societal needs.

In the final analysis, the future of AI in language education will depend on the choices made by educators, policymakers, researchers, and technology developers today. Technological advancement alone cannot guarantee positive educational outcomes. Rather, the most successful future will be one in which AI systems are developed and implemented according to principles of *ethical responsibility, inclusivity, transparency, and human-centered design*. By prioritizing these values, language education can harness the transformative potential of AI while safeguarding the educational, cultural, and social dimensions that remain fundamentally human.

12.7 Conclusion

Artificial Intelligence has emerged as one of the most influential forces shaping contemporary language education. Across teaching, learning, and assessment, AI technologies offer unprecedented opportunities for personalization, efficiency, accessibility, and learner support. From intelligent tutoring systems and automated feedback mechanisms to generative AI applications and learning analytics platforms, AI has expanded the possibilities for language education in ways that continue to redefine educational practice.

However, as this chapter has demonstrated, the integration of AI extends far beyond questions of technological capability. The growing presence of AI in educational environments raises complex *ethical, social, cultural, and regulatory challenges* that require careful consideration. Issues related to *data privacy, educational surveillance, algorithmic bias, learner data ownership, and responsible AI use* illustrate that technological innovation must be accompanied by robust ethical safeguards. Without such safeguards, AI risks reproducing existing inequalities and compromising fundamental educational values.

The chapter has also highlighted the importance of *equity* and *inclusion*. While AI can support personalized learning and accessibility, unequal access to technological resources, linguistic representation, and digital literacy continues to shape educational opportunities worldwide. Digital divides, linguistic imperialism, and disparities between Global North and Global South contexts demonstrate that equitable AI implementation requires sustained attention to diversity, inclusion, and social justice. Similarly, the future of AI-enhanced language education depends on ensuring that technological innovation benefits learners from all linguistic, cultural, and socioeconomic backgrounds.

Policy and *governance* play an equally important role in shaping responsible AI adoption. Institutional guidelines, assessment regulations, national strategies, and international frameworks provide essential mechanisms for balancing innovation with accountability. Ministries of education, universities, accreditation agencies, and international organizations must work collaboratively to establish governance structures that protect learner rights, promote educational quality, and promote ethical AI implementation.

Looking ahead, the most promising future for AI in language education lies not in automation alone but in *human-centered approaches* that combine technological capabilities with human expertise, critical judgment, and ethical responsibility. Human–AI collaboration, multilingual and inclusive AI development, ethical-by-design systems, and widespread AI literacy will be essential components of sustainable educational innovation. Ultimately, AI should be viewed not as a replacement for educators or learners but as a tool that supports human development, communication, and learning.

As language education continues to evolve within increasingly AI-mediated environments, maintaining a commitment to fairness, transparency, inclusion, and educational integrity will be crucial. By embracing these principles, educators and institutions can harness the transformative potential of AI while ensuring that language education remains fundamentally human, equitable, and socially responsible.

Ensuring educational equity in the AI era requires a transition from "*support as substitution*" to "*support as augmentation*" (Hooshyar et al., 2025; Yan et al., 2025). Success depends on developing inclusive AI curricula that affirm diverse student identities and linguistic backgrounds while maintaining robust regulatory oversight and ethical transparency.

Pedagogical Companion for Chapter 12

Key Concepts: data privacy and surveillance; algorithmic bias; data ownership; responsible classroom AI use; digital divides; linguistic imperialism; Global South perspectives; accessibility and special needs; institutional policy; international guidelines.

Discussion Questions

1. Map the data flows of one AI tool used in your institution: what is collected, where does it go, and who could answer these questions if a learner asked?
2. The chapter links AI to linguistic imperialism. By what mechanisms could language AI strengthen—or weaken—the position of dominant languages?

3. Whose interests does an institutional AI policy protect in your context: learners, teachers, the institution, or vendors? Rank them by how the policy's actual text reads.
4. For which learners with special needs does AI currently deliver the greatest genuine benefit, and where does accessibility rhetoric outrun the evidence?
5. If you could add one binding requirement to international AI-in-education guidance, what would it be, and how would compliance be verified?

Further Reading

- Miao, F., Holmes, W., Huang, R., & Zhang, H. (2021). *AI and education: Guidance for policy-makers*. UNESCO.
- Prinsloo, P. (2020). Data frontiers and frontiers of power in (higher) education. *Teaching in Higher Education, 25*(4).
- Evangeline, S. I. (2025). Regulatory frameworks for agentic AI in education. IGI Global.

Chapter 13. Conclusion: Towards Responsible and Pedagogically Grounded AI Use

13.1 Introduction

The integration of Artificial Intelligence (AI) into language teaching, learning, and assessment represents one of the most significant educational developments of the 21st century. Over the past several decades, language education has experienced successive waves of technological innovation, beginning with Computer-Assisted Language Learning (CALL), progressing through Mobile-Assisted Language Learning (MALL) and Intelligent Computer-Assisted Language Learning (ICALL), and culminating in the recent emergence of Generative Artificial Intelligence (GenAI) and Large Language Models (LLMs). Each technological shift has expanded the possibilities for language instruction, learner support, and assessment. However, the rapid development and widespread adoption of AI technologies have introduced changes of a different magnitude, challenging fundamental assumptions about language learning, teaching expertise, educational assessment, and the nature of human communication itself (Kasneci et al., 2023; Pan, 2024).

The chapters presented throughout this volume have illustrated both the remarkable opportunities and the complex challenges associated with AI-enhanced language education. AI technologies now possess capabilities that only a few years ago appeared unattainable. Contemporary systems can generate coherent texts, simulate human conversation, provide personalized feedback, assess language production, translate across languages, support adaptive learning pathways, and analyze large volumes of educational data in real time (Alfredo et al., 2024; Ng et al., 2025). These capabilities have created new possibilities for *individualized learning, increased accessibility, enhanced learner autonomy, and more efficient educational practices*. At the same time, they have raised profound questions regarding *ethics, academic integrity, algorithmic bias, data privacy, educational equity, and the future role of teachers* in increasingly AI-mediated learning environments (Akgun & Greenhow, 2022; Ma et al., 2025).

One of the central themes emerging from this book is that AI should not be understood merely as a collection of technological tools. Rather, AI represents a *transformative force* that is reshaping educational ecosystems at

multiple levels simultaneously. AI influences not only classroom practices but also curriculum design, assessment frameworks, teacher preparation, institutional governance, policy development, and educational research. For this reason, discussions surrounding AI in language education extend beyond questions of technological effectiveness to encompass broader pedagogical, social, ethical, and philosophical concerns. The fundamental challenge facing educators is no longer whether AI can perform educational tasks, but how educational systems should respond to the growing presence of intelligent technologies within learning environments.

Throughout this book, a recurring argument has been that the future of language education depends not on technological sophistication alone but on the quality of *pedagogical decisions* that guide AI implementation. Although AI systems continue to demonstrate increasingly advanced capabilities, educational value cannot be reduced to automation, efficiency, or computational power. *Language learning is inherently human*. It involves communication, identity formation, cultural understanding, emotional expression, critical reflection, and social interaction. These dimensions of learning cannot be fully captured by technological systems, regardless of their sophistication. It follows that successful AI integration requires approaches that position technology as a means of supporting human learning rather than replacing human educational relationships (Hooshyar et al., 2025; Wiese et al., 2025).

Another important insight emerging from the preceding chapters concerns the need to move beyond simplistic narratives that portray AI either as an educational revolution or as a threat to educational integrity. Such binary perspectives fail to acknowledge the complexity of contemporary educational realities. AI technologies simultaneously create opportunities and risks. They can support learner autonomy while encouraging dependency. They can enhance accessibility while reproducing inequalities. They can facilitate personalized instruction while raising concerns about surveillance and privacy. They can strengthen assessment practices while challenging traditional notions of authorship and originality. Understanding these tensions requires nuanced and evidence-based approaches that recognize both the *strengths and limitations of AI* within educational contexts (Kasneci et al., 2023; Song et al., 2024).

Equally significant is the growing recognition that AI is transforming not only how language education is delivered but also what competencies learners require for participation in contemporary society. Increasingly, language learners must develop skills that extend beyond linguistic proficiency alone. Critical AI literacy, ethical awareness, information evaluation, digital citizenship, and the ability to collaborate effectively with intelligent technologies are becoming essential competencies in AI-mediated educational and professional environments (Long & Magerko, 2020; Ng et al., 2021). As a result, language education itself must evolve to prepare learners for futures in which communication increasingly occurs through interactions involving both humans and intelligent systems.

The discussions presented throughout this book also highlight the importance of equity and inclusion as foundational principles for AI implementation. While AI has the potential to expand educational opportunities, its benefits are not distributed equally across contexts. Significant disparities continue to exist regarding technological access, infrastructure, digital literacy, language representation, and institutional resources. Moreover, the dominance of English and other high-resource languages within AI development raises concerns regarding linguistic diversity and the marginalization of underrepresented languages and communities (Pan, 2024; Ma et al., 2025). Accordingly, responsible AI integration requires deliberate efforts to ensure that technological innovation contributes to educational justice rather than reinforcing existing inequalities.

As AI technologies continue to evolve, educators, researchers, institutions, and policymakers face a critical responsibility. The future of language education will not be determined solely by advances in machine learning, natural language processing, or multimodal systems. Rather, it will be shaped by human decisions regarding how these technologies are designed, implemented, regulated, and evaluated. Questions concerning *educational purpose, human agency, ethical responsibility,* and *social inclusion* therefore remain central to discussions of AI-enhanced learning.

Against this background, this concluding chapter synthesizes the major insights emerging from the preceding chapters and reflects on the future directions of AI in language education. First, it identifies key lessons learned across teaching, learning, and assessment contexts. Second, it proposes a set of principles for meaningful and pedagogically grounded AI integration.

Finally, it offers reflections for educators, researchers, institutions, and policymakers who seek to navigate the opportunities and challenges of AI-enhanced language education responsibly.

Thus, the central argument of this chapter is that the future of language education should not be understood as a choice between human intelligence and artificial intelligence. Rather, it involves the creation of *productive, ethical, and educationally meaningful partnerships* between humans and intelligent technologies. The success of AI in language education will not be measured solely by technological innovation, but by the extent to which these innovations enhance human learning, promote equity, support linguistic diversity, and contribute to the broader goals of education. In this sense, the future of AI in language education is fundamentally a human future—one that requires thoughtful, responsible, and pedagogically grounded engagement with the technologies that increasingly shape our educational landscapes.

13.2 Key Insights Across Teaching, Learning, and Assessment

The rapid integration of Artificial Intelligence (AI) into language education has generated one of the most profound transformations in the history of language teaching, learning, and assessment. Across the chapters of this volume, a recurring theme has emerged: AI is neither a simple technological tool nor an inevitable replacement for human educational practice. Rather, it represents a complex *socio-technical development* that simultaneously creates new opportunities, new responsibilities, and new challenges for educators, learners, institutions, and policymakers. As language education enters an increasingly AI-mediated era, it becomes necessary to synthesize the key insights emerging across teaching, learning, and assessment in order to understand both the possibilities and limitations of this transformation.

A central finding throughout the literature is that AI has considerably *expanded the capacity for personalization in language education*. Traditional language instruction has often struggled to accommodate the diversity of learner needs present within contemporary classrooms. Differences in proficiency levels, learning styles, motivation, prior knowledge, linguistic backgrounds, and learning pace frequently challenge educators attempting

to provide individualized support. AI-enhanced educational systems offer new mechanisms for addressing these challenges through adaptive learning pathways, intelligent tutoring systems, automated feedback, conversational agents, and learning analytics (Alfredo et al., 2024; Pan, 2024). By continuously analyzing learner interactions and performance data, AI systems can provide targeted recommendations, identify learning difficulties, and deliver personalized instructional support that would be difficult to achieve at scale through traditional instructional models alone.

Research reviewed throughout this book suggests that *AI-supported personalization* can positively influence *learner engagement* and *language development* when implemented appropriately. Generative AI systems can provide immediate explanations, vocabulary support, grammar feedback, and conversational practice opportunities that extend learning beyond classroom boundaries (Kasneci et al., 2023). AI-powered conversational agents have demonstrated particular promise in supporting speaking practice and reducing foreign language anxiety by providing learners with low-stakes communicative environments in which to experiment with language use. Such systems allow learners to engage in repeated interactions, receive immediate responses, and build confidence without the social pressures that often accompany face-to-face classroom participation.

At the same time, the literature emphasizes that per*sonalization should not be equated with automation.* Learners benefit most when AI functions as *a scaffold for self-regulated learning* rather than as a substitute for cognitive effort (Anders & Speltz, 2025; Kong & Yang, 2024). Studies examining learner interactions with generative AI indicate that excessive dependence on automated support can result in *reduced critical reflection, weakened writing originality, and diminished metacognitive engagement.* Consequently, effective AI-enhanced learning environments require pedagogical structures that encourage learners to question, evaluate, and refine AI-generated outputs rather than simply accept them. The educational value of personalization therefore depends not only on technological sophistication but also on how successfully AI supports *learner agency* and *intellectual growth.*

The evidence reviewed throughout this volume further suggests that AI is *reshaping the role of teachers* rather than eliminating it. Early public discussions frequently portrayed AI as a potential replacement for educators.

However, contemporary research increasingly supports a different interpretation. Generative AI functions most effectively as a *pedagogical partner* that augments instructional practice rather than replacing professional expertise. AI systems can assist teachers by *generating instructional materials, creating differentiated learning resources, producing vocabulary activities, supporting lesson planning, and streamlining administrative tasks.* Yet the effectiveness of these applications depends heavily on the pedagogical judgment, technological competence, and ethical awareness of educators themselves (Mishra et al., 2023; Guan et al., 2025).

On this basis, language teachers are increasingly transitioning from traditional knowledge transmitters to *learning designers* and *AI orchestrators*. Their responsibilities now extend beyond content delivery to include facilitating critical engagement with AI, monitoring the quality of AI-generated outputs, supporting ethical technology use, and guiding learners through increasingly complex human-AI interactions. This emerging teacher role requires new forms of professional expertise that combine pedagogical knowledge with AI literacy, data literacy, and ethical reasoning.

Feedback represents another domain in which AI has generated significant pedagogical innovation. Providing *timely, individualized, and actionable feedback* has long been recognized as one of the most influential factors in language learning. Yet it remains one of the most resource-intensive instructional responsibilities. AI technologies now offer the capacity to provide *automated feedback on writing, speaking, pronunciation, vocabulary, and grammar with unprecedented speed and scalability* (Latif & Zhai, 2024; Pack et al., 2024). Learners can receive immediate responses, revise their work iteratively, and engage in multiple cycles of improvement without waiting for instructor intervention.

Recent studies suggest that such feedback can contribute positively to language development, particularly when learners actively engage with revision opportunities (Dai et al., 2025; Zhuang et al., 2025). Nevertheless, research consistently cautions against viewing AI-generated feedback as inherently accurate or pedagogically sufficient. Automated systems may overlook pragmatic appropriateness, intercultural considerations, rhetorical nuance, or contextual meaning. They may also reinforce dominant linguistic norms at the expense of legitimate linguistic variation (Song et al., 2024; Wang et al., 2025). Seen in this light, the most effective feedback models

appear to be hybrid systems in which AI enhances efficiency while human educators provide interpretation, contextualization, and pedagogical guidance (Hooshyar et al., 2025).

The rise of generative AI has also altered fundamental conceptions of language learning itself. Historically, language education often prioritized grammatical accuracy, vocabulary acquisition, and the production of correct linguistic forms. Today, AI systems can generate grammatically sophisticated texts, perform translations, summarize information, and create coherent written responses almost instantaneously. These capabilities challenge traditional assumptions regarding what constitutes meaningful language competence and what educational assessments should measure (Pan, 2024).

As a result, educational priorities are increasingly shifting toward *higher-order competencies* that remain difficult to automate. *Critical thinking, argumentation, intercultural communication, creativity, ethical reasoning,* and *information evaluation* are becoming central educational objectives because they require learners to move beyond reproduction toward interpretation and judgment (Ng et al., 2021; Ma et al., 2025). Language learning is therefore evolving from a focus on producing language toward a broader emphasis on understanding, evaluating, adapting, and ethically using language within complex communicative environments.

The literature further highlights the growing importance of AI literacy as a *foundational educational competency*. AI literacy extends beyond technical proficiency and includes understanding how AI systems function, recognizing their limitations, evaluating their outputs critically, and using them responsibly (Long & Magerko, 2020; Ng et al., 2021). Recent frameworks emphasize that both educators and learners must develop the capacity to interrogate AI-generated information rather than accepting it uncritically (Jin et al., 2025; Wiese et al., 2025). This competency is especially important in language education, where AI-generated texts may appear fluent and authoritative despite containing factual inaccuracies, cultural biases, or problematic assumptions.

Assessment represents perhaps the most contested area of AI integration. AI-powered assessment systems offer substantial benefits in terms of scalability, consistency, efficiency, and rapid feedback. Automated essay scoring, speech evaluation technologies, and large language model-based

assessment systems can process large quantities of learner data while reducing instructor workload (Pack et al., 2024; Rodrigues et al., 2025). Advances in deep learning, including Bidirectional Long Short-Term Memory (BiLSTM) architectures and Fuzzy Inference Systems (FIS), have further improved the capacity of educational systems to evaluate complex linguistic performance while reducing scoring variability.

Despite these advantages, significant concerns remain regarding *validity, fairness, transparency, and accountability*. Research indicates that AI assessment systems may exhibit biases against non-native writing styles, regional dialects, and underrepresented linguistic groups (Liang et al., 2023; Baker & Hawn, 2022). Even highly advanced systems such as GPT-4 demonstrate inconsistencies when compared with expert human evaluators in high-stakes contexts (Rodrigues et al., 2025; Wang et al., 2025). As a result, the future of language assessment is unlikely to involve complete automation. Instead, evidence increasingly supports human-in-the-loop approaches in which AI contributes analytical efficiency while educators retain responsibility for interpretation, judgment, and accountability (Hooshyar et al., 2025).

Finally, one of the most significant insights emerging across teaching, learning, and assessment concerns the relationship between AI and educational equity. AI possesses extraordinary potential to expand access to educational resources, personalize instruction, and support multilingual learning. Yet it also risks reinforcing existing inequalities related to technological infrastructure, linguistic representation, digital literacy, and resource distribution (Agyare et al., 2025; Belkina et al., 2025). Most current AI systems continue to privilege English and other high-resource languages, raising concerns regarding linguistic justice and cultural inclusion.

Nevertheless, emerging *multilingual* and *multimodal technologies* provide reasons for cautious optimism. Advances in multilingual language models, bilingual conversational agents, multimodal learning environments, and culturally responsive AI systems suggest that future developments may better support *linguistic diversity* and *educational inclusion* (META, 2022; Xiao et al., 2025; Nyaaba & Zhai, 2025). These innovations indicate that AI can contribute to educational equity when guided by principles of *accessibility, cultural responsiveness,* and *social responsibility*.

Considered together, the evidence reviewed throughout this volume demonstrates that AI is not merely introducing new technologies into language education; it is prompting a fundamental reconsideration of teaching, learning, and assessment themselves. The future of language education will not be determined solely by advances in artificial intelligence but by the ways educational communities choose to integrate these technologies into pedagogical practice. The most promising futures are those in which AI enhances human learning without diminishing *human agency*, expands educational opportunities without deepening inequality, and supports innovation while preserving the ethical, relational, and human foundations that define meaningful language education.

The evidentiary base behind this synthesis is narrower than it may appear. Much of the personalization and feedback research surveyed in this volume comes from higher-education and EFL settings in China, East Asia, and the Anglophone world, using quasi-experimental or self-report designs; a systematic review of 22 empirical studies, drawn from 189 screened records, found consistent gains across reading, writing, listening, speaking, and vocabulary in school-level EFL contexts, while noting that digital-divide and teacher-readiness challenges persisted (Kundu & Bej, 2025). The optimistic reading of personalization is further complicated by evidence that performance gains and learning gains are not interchangeable: in a randomized experiment with 117 university students, the ChatGPT group produced the highest-scoring essays yet showed no advantage in knowledge gain or transfer, a pattern the authors attribute to metacognitive laziness (Fan et al., 2024)—the same distinction a Nature Reviews Psychology commentary warns is easily lost when task performance is treated as a proxy for deep cognitive engagement (Yan et al., 2025). These findings do not overturn the chapter's claims, but they discipline them.

13.3 Principles for Meaningful AI Integration

As Artificial Intelligence (AI) becomes increasingly embedded in language teaching, learning, and assessment, the central challenge facing educators is no longer whether AI should be used, but how it should be used. Throughout this volume, a consistent theme has emerged: technological innovation alone does not guarantee educational improvement. The effectiveness of AI depends largely on the *pedagogical principles, ethical*

frameworks, and institutional practices that guide its implementation. In turn, meaningful AI integration requires moving beyond technology-centered approaches toward pedagogically grounded models that place learning, human development, and educational values at the center of decision-making.

The literature reviewed across this book suggests that *responsible AI adoption* should be guided by a set of foundational principles. These principles do not represent rigid rules but rather interconnected commitments that help educators, institutions, and policymakers navigate the opportunities and challenges associated with AI-enhanced language education. Together, they provide a framework for ensuring that AI serves educational purposes rather than allowing educational practices to become subordinate to technological capabilities.

13.3.1 Pedagogy Before Technology

Perhaps the most important principle emerging from contemporary scholarship is that *pedagogy must precede technology*. Educational history contains numerous examples of technological innovations that generated considerable enthusiasm but ultimately failed to produce meaningful learning improvements because pedagogical considerations were overlooked. From language laboratories and multimedia software to early computer-assisted language learning systems, technological advancement alone has never guaranteed educational effectiveness. Artificial Intelligence (AI) is no exception. The educational value of AI does not reside in the technology itself but in how it is integrated into teaching, learning, and assessment processes (Kasneci et al., 2023; Pan, 2024).

As AI becomes increasingly embedded within language education, there is a risk that institutions and educators may focus excessively on technological capabilities while neglecting pedagogical purpose. Generative AI systems can produce texts, generate exercises, provide feedback, and simulate conversations with remarkable efficiency. However, the existence of these capabilities does not automatically justify their use in educational contexts. Meaningful AI integration begins not with the question "*What can this technology do?"* but rather with questions such as "*What are learners expected to achieve?*", "*How does this technology contribute to language development?*", and "*What pedagogical problem does it solve?*" Educational goals must therefore drive technological decisions rather than the reverse.

This principle is particularly important in language education because language learning is a *multidimensional process involving cognitive, social, cultural, emotional, and communicative development.* Language acquisition extends beyond the accumulation of vocabulary and grammatical structures; it requires *opportunities for interaction, meaning-making, reflection, negotiation, and identity construction*. For this reason, AI systems should be evaluated not only according to their technical performance but also according to their capacity to support broader educational objectives such as *communicative competence, intercultural understanding, critical literacy, learner autonomy,* and *lifelong learning* (Ng et al., 2021; Mishra et al., 2023).

Contemporary pedagogical theories provide valuable guidance for AI implementation. *Constructivist approaches* emphasize that learning occurs through active engagement, knowledge construction, and meaningful interaction rather than passive information reception. Similarly, *sociocultural theories* highlight the importance of social participation, collaboration, scaffolding, and mediated learning experiences in language development. These perspectives suggest that AI should not function merely as an automated information provider but as a tool that supports learners' active engagement with language, ideas, and communities of practice.

Within this context, *Conversational frameworks* offer a particularly useful lens for evaluating AI-enhanced language learning environments. According to this framework, effective learning emerges through *iterative cycles of dialogue, feedback, adaptation, and reflection between teachers, learners, and learning resources*. AI technologies can contribute to these conversational processes by providing *immediate feedback, generating alternative explanations, and supporting practice opportunities*. However, their educational value depends on whether they facilitate *meaningful knowledge negotiation* and *learner reflection* rather than encouraging passive consumption of machine-generated content. AI should therefore function as a responsive *pedagogical infrastructure* that supports interaction and inquiry rather than replacing the conversational dimensions of learning.

The principle of pedagogy before technology also highlights the importance of *instructional design*. Linguistic platforms should never be implemented merely because of their novel generative capabilities. Instead, educators should design *structured learning activities* that deliberately

align AI use with established pedagogical objectives and learning theories. For example, AI-supported writing tasks should encourage *drafting, revision, reflection, and critical evaluation* rather than simply producing polished final products. Similarly, conversational agents should be employed to facilitate communicative practice and strategic language use rather than merely generating grammatically correct responses. Such approaches ensure that AI serves educational development rather than technological convenience.

Recent research further suggests that *pedagogically grounded AI integration* supports *learner agency* and *self-regulated learning* more effectively than technology-centered approaches. When learners are encouraged to evaluate AI outputs, justify decisions, revise generated content, and reflect on their interactions with intelligent systems, AI becomes a catalyst for deeper cognitive engagement rather than a shortcut that bypasses learning processes (Anders & Speltz, 2025; Kong & Yang, 2024). In this sense, the educational value of AI emerges not from automation itself but from the quality of the learning experiences that educators design around it.

Large-scale evidence corroborates this warning about sequencing. A PRISMA- and TPACK-guided review of 125 empirical studies conducted between 2021 and 2025 found only a negligible correlation between the scale of institutional investment in faculty AI development and any resulting pedagogical transformation, alongside a persistent divide in outcomes between Global North and Global South institutions (Altamirano et al., 2026). Investment in tools and training, in other words, does not translate into instructional change unless pedagogical redesign is the explicit object of that investment rather than an assumed byproduct of exposure to the technology.

As we have stated above, the principle of pedagogy before technology reminds educators that AI is a means rather than an end. Educational innovation should be guided by theories of learning, disciplinary goals, and learner needs rather than by technological novelty. As language education continues to evolve in increasingly AI-mediated environments, the most successful implementations will be those that remain firmly anchored in pedagogical purpose. Technology may transform the tools available to educators, but pedagogy must continue to determine how those tools are used in the service of meaningful language learning.

13.3.2 Human-Centered Design and Human Agency

A *second foundational principle* is the preservation of *human agency*. Across recent AI-in-education research, scholars consistently argue that educational technologies should augment rather than replace human decision-making (Alfredo et al., 2024; Hooshyar et al., 2025). *Human-centered AI* recognizes that learners and educators must remain active participants in educational processes rather than passive recipients of algorithmic recommendations.

In language education, *human agency* manifests in multiple ways. Learners should maintain control over their learning goals, learning strategies, and interactions with AI systems. Similarly, educators should retain authority over instructional decisions, assessment judgments, and pedagogical interventions. AI may provide recommendations, analyzes, and feedback, but final educational decisions should remain human responsibilities.

This perspective aligns closely with emerging *"human-in-the-loop"* approaches that combine the computational strengths of AI with the contextual understanding, ethical reasoning, and professional judgment of educators (Mosqueira-Rey et al., 2023). Such models acknowledge that effective education depends not only on information processing but also on empathy, trust, mentorship, and social relationships that remain uniquely human dimensions of learning.

Empirical grounding for this principle remains limited, and what exists complicates as much as it confirms. Comparing human-factors, psychology, and computer-science perspectives, one influential synthesis mapped substantial disciplinary disagreement over what "human-centered AI" means even in general work contexts, a caution worth carrying into education, where the term is invoked just as freely (Mazarakis et al., 2023). Agency preservation cannot, moreover, be assumed to be self-sustaining once designed in; in a 114-student experiment, learners preferred human-tutor feedback overall, but their preference for ChatGPT's free-dialogue interface rose measurably after continued use, a pattern the study's authors note raises questions about over-reliance and algorithm-aversion (Le et al., 2025). Human-centered design, on this evidence, is a standard that must be actively maintained rather than a property that survives on its own once a system is launched.

13.3.3 Transparency and Explainability

Meaningful AI integration also requires transparency. Educational stakeholders must understand when AI is being used, how it operates, what data it relies upon, and what limitations it possesses. Yet many contemporary AI systems function as opaque "black boxes" whose internal processes remain difficult to interpret even for experts (Hooshyar et al., 2025; Wiese et al., 2025).

Transparency is especially important in language education because AI systems increasingly influence feedback, assessment, recommendations, and learning opportunities. Learners have a right to understand how educational decisions affecting them are made. Similarly, educators need sufficient insight into AI-generated outputs to evaluate their accuracy and appropriateness.

Explainable AI contributes to trust and accountability by making algorithmic decisions more understandable and contestable. In assessment contexts, for example, learners should be able to understand why a particular score or recommendation was generated. Without transparency, AI risks undermining both educational fairness and learner confidence.

The empirical record supports this concern directly. A mixed-methods study of 628 stakeholders across higher-education institutions in Sichuan, China, found that policy clarity, transparency, and human oversight were the strongest predictors of institutional readiness for AI-enabled learning governance (Che & Perkins, 2026). Institutions that treat transparency as a compliance afterthought rather than a design requirement should therefore expect readiness problems that no amount of additional infrastructure will resolve on its own.

13.3.4 Ethical Data Governance and Privacy Protection

The increasing reliance on learner data has made *ethical data governance* one of the defining challenges of AI-enhanced education. Contemporary AI systems depend heavily on the collection, analysis, and interpretation of learner data to personalize instruction, provide adaptive feedback, monitor progress, and generate predictive insights regarding learning outcomes. These capabilities have undoubtedly expanded the possibilities for individualized learning support. However, they have also introduced significant concerns regarding *privacy, surveillance, data ownership, informed consent,*

transparency, and security (Akgun & Greenhow, 2022; Jin et al., 2025). As AI becomes more deeply embedded within language teaching, learning, and assessment, the responsible management of learner data must be regarded as a foundational educational principle rather than a secondary technical consideration.

Language education presents particular challenges in this regard because language learning activities frequently involve highly personal forms of communication. Learners routinely produce written reflections, spoken narratives, conversational interactions, discussion posts, and creative texts that may reveal personal experiences, beliefs, identities, emotions, and cultural perspectives. When these interactions are captured, analyzed, and stored by AI systems, important ethical questions emerge regarding who owns the resulting data, how they are protected, and whether learners fully understand the implications of their participation. It follows that data governance in language education extends beyond technical compliance to encompass broader concerns related to human dignity, learner autonomy, and educational trust.

Meaningful AI integration therefore requires *robust data governance frameworks* that prioritize *learner rights, institutional accountability, and ethical transparency.* Educational institutions must clearly communicate what data are collected, how they are processed, where they are stored, who can access them, and for what purposes they may be used. *Transparency* is particularly important because many AI-powered educational platforms operate through complex data ecosystems that are often invisible to learners and educators alike. Ethical implementation requires that learners are treated as informed participants in these processes rather than passive sources of data extraction.

A central principle within responsible data governance is *informed consent*. Learners should possess a clear understanding of how their interactions with AI systems contribute to educational analytics, system improvement, or broader technological development. Consent should not be reduced to lengthy terms-of-service agreements that are rarely read or understood. Instead, educational institutions should adopt accessible and learner-friendly approaches that explain data practices in transparent and comprehensible ways. Such practices strengthen trust and reinforce learners' rights to make informed decisions regarding their personal information.

The principle of *data minimization* is equally important. Educational systems should collect only the information necessary to achieve clearly defined pedagogical objectives rather than accumulating extensive behavioral data simply because technological capabilities make such collection possible. Advances in learning analytics and AI increasingly enable institutions to track clicks, browsing behaviors, engagement patterns, eye movements, speech characteristics, and other forms of learner activity. While these data may offer valuable educational insights, indiscriminate collection risks creating environments of continuous surveillance that undermine learner autonomy and psychological safety. Ethical governance therefore requires balancing personalization with privacy and innovation with individual rights.

Closely related to data minimization is the principle of *learner agency*. Contemporary approaches to responsible AI increasingly emphasize that learners should retain meaningful control over the information they generate. This includes opportunities to access, review, correct, and in some cases delete personal data stored within educational systems. Human-centered models of AI governance argue that learners should be viewed not simply as data providers but as active stakeholders whose interests and rights remain central throughout the educational process (Hooshyar et al., 2025). Such perspectives align closely with broader principles of educational ethics and democratic participation.

The deployment of AI in education also raises important concerns regarding *surveillance* and *algorithmic monitoring.* AI-powered systems can track learner behaviors in ways that were previously impossible, generating detailed profiles of participation, engagement, performance, and communication patterns. While these capabilities may support early intervention and personalized learning, excessive monitoring risks transforming educational environments into spaces of constant observation. Scholars increasingly warn that educational surveillance can influence learner behavior, increase anxiety, and reduce opportunities for intellectual risk-taking and authentic self-expression (Holmes & Tuomi, 2022). Accordingly, institutions must carefully evaluate whether specific forms of data collection genuinely serve educational purposes or merely expand institutional oversight.

Beyond *privacy* concerns, ethical data governance must also address questions of *fairness* and *representation*. AI systems are trained on data that

often reflect historical inequalities, cultural assumptions, and dominant linguistic norms. Without careful oversight, such systems may reproduce forms of algorithmic bias that disadvantage learners from underrepresented linguistic, cultural, or socioeconomic backgrounds. Researchers have highlighted concerns regarding dialect prejudice, linguistic discrimination, and the marginalization of non-dominant language varieties within AI-powered educational systems (Kizilcec & Lee, 2022). Responsible governance therefore requires ongoing auditing, evaluation, and critical examination of how AI systems affect diverse learner populations.

The risk is not hypothetical: large language models have been shown to generate covertly racist decisions about people based on their dialect (Hofmann et al., 2024). Because such discrimination operates beneath the level of overt content, it can survive exactly the kind of surface-level moderation many institutions rely on—which is why dialect-aware auditing belongs in data governance as a requirement, not an optional refinement.

For language education specifically, this challenge extends to questions of *linguistic justice*. AI systems should not privilege standardized forms of language at the expense of linguistic diversity, regional dialects, multilingual practices, or culturally situated forms of communication. Educational technologies must be designed and implemented in ways that respect pluralistic Englishes, multilingual identities, and diverse cultural perspectives. Failure to do so risks reinforcing existing linguistic hierarchies and reproducing forms of digital linguistic inequality. *Ethical AI* integration therefore requires educators and institutions to actively interrogate AI outputs, ensuring that language technologies support inclusion rather than assimilation.

In this context, *ethical data governance* and *privacy protection* are not merely technical or legal concerns; they are fundamentally educational responsibilities. *Responsible AI* integration depends on establishing systems that respect *learner autonomy, protect personal information, promote transparency, and safeguard human rights*. As AI becomes increasingly influential in language teaching, learning, and assessment, institutions must ensure that technological innovation remains aligned with educational values. The future of AI-enhanced language education will depend not only on what technologies can achieve but also on how responsibly they are governed. Meaningful AI integration therefore requires a commitment to

privacy, transparency, accountability, and *linguistic justice* as foundational principles guiding the development of educational futures.

13.3.5 Equity, Accessibility, and Inclusion

A recurring concern throughout this volume has been the uneven distribution of AI's benefits and risks. *Digital divides, linguistic inequalities, algorithmic bias, and unequal access to technological resources* threaten to exacerbate existing educational disparities if left unaddressed (Song et al., 2024; Belkina et al., 2025).

Meaningful AI integration therefore requires a deliberate commitment to equity and inclusion. Educational technologies should be designed to support diverse learners, including multilingual students, learners with disabilities, students from underrepresented communities, and those with limited access to technological resources.

Accessibility should not be viewed as an optional feature but as a fundamental design requirement. AI systems should align with principles of *Universal Design for Learning (UDL)*, providing multiple pathways for engagement, representation, and expression (Ng et al., 2021). What is more, institutions should ensure that students are not disadvantaged because of financial limitations, technological barriers, or linguistic background.

In language education specifically, promoting *multilingualism* is essential. Future AI systems should support linguistic diversity rather than reinforcing the dominance of a small number of global languages. Efforts to expand multilingual datasets, support indigenous languages, and develop culturally responsive technologies represent important steps toward more equitable educational futures (Nyaaba & Zhai, 2025).

Concrete evidence shows how easily this commitment can be undercut by the underlying technology itself. Benchmark testing of Gemini 2.5 Flash found refusal rates near 90% in English compared with only 35% to 55% across Yoruba, Hausa, Igbo, and Igala, showing severe degradation of cross-lingual safety behavior in precisely the lower-resourced languages this book has argued deserve equal protection (Faruna, 2026). Multilingual promise carries its own tensions even where safety is not at issue: a survey of Basque speakers found that growing reliance on machine translation, convenient as it is, raises concerns for the carefully planned language-revitalization process on which Basque's future depends (Aranberri &

Iñurrieta, 2024). Equity in AI-mediated language education, on this evidence, cannot be secured merely by adding more languages to a model's training data; the safety and sociolinguistic consequences of that inclusion vary enormously by language and must be evaluated language by language rather than assumed.

13.3.6 Assessment Integrity and Human Oversight

The widespread availability of Generative AI has challenged traditional approaches to assessment and academic integrity. Consequently, meaningful AI integration requires assessment frameworks that remain valid, fair, and educationally meaningful in AI-mediated environments (Pan, 2024).

Recent research suggests that the solution is not to eliminate AI from assessment contexts but to redesign assessment practices. Process-oriented evaluation, reflective learning tasks, portfolio assessment, authentic performance activities, and collaborative projects provide richer evidence of learner development than assessments focused solely on final products (van Niekerk et al., 2025).

At the same time, automated assessment systems should operate under human supervision. AI may assist with scoring, feedback, and analytics, but high-stakes educational decisions should remain subject to human review (Wang et al., 2025). Such oversight helps mitigate algorithmic bias, ensure fairness, and maintain public trust in educational assessment systems.

The rationale for retaining human oversight is strengthened by evidence about how that oversight can be quietly eroded. A conceptual analysis of generative-AI-driven automated essay scoring argues that such systems risk diminishing teachers' evaluative agency unless an explicit ecological framework requires teachers to critically mediate AI-generated scores, both to mitigate bias and to protect data privacy (Barrot, 2026). The threat is not only institutional but psychological: research on student attitudes toward AI-based grading finds that reluctance to accept an AI-generated grade decreases when students are dissatisfied with the status quo or when the AI's outcome happens to favor them (Jones-Jang et al., 2025), suggesting that acceptance of automated assessment can track self-interest as readily as it tracks a considered judgment about fairness. Human oversight, in other words, is not a safeguard that institutions can assume stakeholders will demand for principled reasons alone.

13.3.7 AI Literacy as a Core Educational Competency

One of the most significant lessons emerging from contemporary research on Artificial Intelligence (AI) in education is that meaningful and responsible AI integration depends fundamentally on *AI literacy*. Throughout previous technological transformations, educational systems recognized the importance of developing *digital literacy, information literacy, and media literacy* as essential competencies for participation in increasingly technology-mediated societies. Today, a similar transition is occurring with AI. As intelligent systems become embedded in everyday educational, professional, and social practices, AI literacy is increasingly viewed as a foundational competency for both educators and learners (Long & Magerko, 2020; Ng et al., 2021).

The importance of AI literacy is particularly evident in language education, where learners *interact directly with AI-generated texts, automated feedback systems, conversational agents, translation tools, adaptive learning platforms, and multimodal educational environments*. These technologies are no longer peripheral resources; they are becoming integral components of language learning ecosystems. On this basis, learners and teachers must develop the knowledge and critical capacities necessary to engage with AI effectively, ethically, and responsibly. Without such competencies, educational users risk becoming passive consumers of AI-generated outputs rather than informed participants capable of evaluating, questioning, and directing intelligent technologies.

AI literacy extends far beyond technical proficiency. It is not simply the ability to operate AI tools or generate effective prompts. Rather, AI literacy encompasses understanding how AI systems function, recognizing their strengths and limitations, evaluating AI-generated outputs critically, identifying potential biases and inaccuracies, understanding ethical implications, and using AI in ways that support learning and human development (Long & Magerko, 2020; Ng et al., 2021). In this sense, AI literacy represents a multidimensional competency that integrates technical knowledge, critical thinking, ethical awareness, and responsible decision-making.

This multidimensional definition is no longer only aspirational; validated instruments now exist to measure it. A Critical Artificial Intelligence Literacy Scale, piloted with 57 doctoral students and then validated with 314 first-year student teachers, yielded a reliable 24-item, four-dimension

instrument that operationalizes the construct rather than merely describing it (Ranieri et al., 2025). Complementing this psychometric approach, a widely cited competency framework—cited by 133 subsequent works—defines twelve generative-AI-literacy competencies spanning foundational understanding, prompt engineering, and ethics, spanning individuals, educators, and policymakers (Annapureddy et al., 2025). Together, these instruments make it possible to ask not only whether a curriculum claims to teach AI literacy but whether learners' scores on a validated scale actually move as a result.

For language learners, AI literacy involves developing the ability to engage critically with AI-generated language. Modern generative systems often produce highly fluent, persuasive, and grammatically sophisticated outputs that can create an illusion of authority. However, fluency should not be confused with accuracy. AI-generated texts may contain factual errors, cultural misrepresentations, biased assumptions, or misleading information despite appearing credible and coherent. Learners must therefore acquire the skills necessary to verify information, compare multiple sources, evaluate evidence, and recognize the limitations of machine-generated knowledge. Such competencies are increasingly essential in a world where AI-mediated communication is becoming commonplace.

The development of AI literacy also requires understanding how algorithms shape information environments. AI systems are not neutral technologies; they are constructed from training data, design decisions, and underlying assumptions that influence their outputs. Learners should therefore understand how algorithmic bias emerges, how data representation affects system performance, and how AI technologies may reinforce social, cultural, and linguistic inequalities. In language education, this awareness is particularly important because AI systems often reflect dominant linguistic norms and may privilege certain varieties of language while marginalizing others. Critical AI literacy enables learners to recognize these patterns and engage with AI technologies more thoughtfully and responsibly.

For educators, AI literacy encompasses an even broader set of competencies. Teachers must understand not only how AI systems function but also how they can be integrated into pedagogically meaningful learning experiences. This includes *evaluating educational applications, selecting appropriate tools, designing AI-supported learning activities, interpreting AI-*

generated feedback, and assessing the educational value of AI-enhanced instruction (Ma et al., 2025). Teachers must also understand the implications of AI for assessment, academic integrity, learner autonomy, and educational equity. As AI becomes increasingly integrated into classrooms, educators require the expertise to guide students in navigating both the opportunities and the risks associated with intelligent technologies.

Recent frameworks increasingly emphasize that AI literacy should become a core component of teacher education and professional development programs (Ma et al., 2025; UNESCO, 2024). Future language teachers will need competencies that extend beyond traditional technological skills. They must understand prompt engineering strategies, algorithmic decision-making processes, ethical governance principles, data protection regulations, and the sociocultural implications of AI adoption. Such knowledge enables educators to function not just as users of AI but as informed mediators who can critically evaluate AI systems and support responsible student engagement.

A key consideration is that fostering technical proficiency alone is insufficient. Contemporary scholarship increasingly advocates for Responsible AI Literacy, a broader framework that integrates ethical, social, and human-centered dimensions into AI education (Ma et al., 2025). Responsible AI Literacy emphasizes questions such as: *Who benefits from AI systems? Who may be disadvantaged? How are learner data collected and used? What forms of bias may be embedded within AI outputs? How can AI be employed in ways that promote fairness, inclusion, and human well-being?* These questions encourage learners and educators to move beyond operational competence toward deeper forms of critical and ethical engagement.

Such perspectives align closely with emerging international frameworks. UNESCO's *AI Competency Framework for Teachers* (Miao & Cukurova, 2024), for example, highlights the importance of combining technological understanding with ethical reasoning, professional judgment, learner protection, and social responsibility. These frameworks recognize that successful AI integration requires educators who can evaluate AI systems not only according to efficiency or performance but also according to their broader educational and societal consequences. Similarly, recent research suggests that AI literacy should be embedded across curricula rather than confined to isolated technology courses, allowing learners to encounter AI-

related questions within authentic disciplinary contexts (UNESCO, 2024; Ma et al., 2025).

Recent studies have also moved to differentiate AI literacy by stakeholder rather than treating it as a single competency ladder. One widely cited proposal (56 citing works) moves beyond the coding-focused literacy models that dominated K-12 practice, proposing instead a stakeholder-first approach in which the ethical understanding expected of different educational audiences is treated as distinct rather than interchangeable (Domínguez Figaredo & Stoyanovich, 2023). A practical instantiation of this differentiation is an AI literacy course-design evaluation tool, built on a four-level international Delphi-study framework and demonstrated on a professional micro-credential, giving educators a concrete instrument for evaluating whether a course actually embeds AI literacy rather than merely gesturing toward it (Parsons et al., 2024). For language education specifically, focus-group research with Spanish applied-linguistics students found that their own working definition of critical AI literacy already combined technical understanding and ethical awareness with an insistence that AI augment rather than replace human expertise (Pérez-Paredes et al., 2025)—a definition arrived at independently by the learners themselves, and one that converges closely with the augmentation commitment this book has defended throughout.

Language education offers particularly rich opportunities for developing AI literacy because language itself serves as the primary medium through which many AI systems operate. Students can *analyze AI-generated texts, evaluate machine translations, compare human and AI-produced writing, investigate linguistic biases, and reflect on the ethical implications of machine-mediated communication*. Such activities simultaneously support language development and critical technological awareness, creating opportunities for interdisciplinary learning that connects language education with broader digital citizenship goals.

Furthermore, AI literacy contributes directly to *learner agency*. Research suggests that individuals who understand how AI systems operate are better positioned to collaborate effectively with intelligent technologies rather than becoming dependent on them (Ng et al., 2021). AI-literate learners are more likely to use AI strategically, question machine-generated recommendations, and maintain control over their learning processes. In contrast,

learners who lack AI literacy may become overly reliant on automated systems, accept outputs uncritically, or fail to recognize algorithmic errors and biases. Developing AI literacy therefore serves as a safeguard against technological dependency while promoting informed and autonomous engagement.

Another important dimension of AI literacy involves *data awareness*. As AI systems increasingly rely on learner data to personalize instruction and generate recommendations, students and educators must understand how data are collected, analyzed, and utilized. Knowledge of data governance, privacy rights, surveillance concerns, and algorithmic profiling helps individuals make informed decisions regarding their participation in AI-enhanced educational environments. Such awareness is essential for protecting learner autonomy and supporting ethical educational practices.

Thus AI literacy represents far more than a technological skill set. It is a critical *educational competency* that enables learners and educators to participate meaningfully in AI-mediated societies. Responsible AI integration cannot be achieved simply by providing access to advanced technologies. Instead, educational institutions must cultivate the knowledge, skills, values, and critical dispositions necessary to engage with AI thoughtfully and ethically. This requires systematic curricular integration, continuous professional development, and sustained attention to the ethical, social, and pedagogical dimensions of AI.

As language education continues to evolve alongside intelligent technologies, AI literacy will become increasingly central to educational success. The goal is more than merely to teach individuals how to use AI tools but to prepare them to understand, evaluate, question, and shape the role of AI in society. By cultivating critical awareness, ethical responsibility, and informed agency, AI literacy provides the foundation upon which meaningful, equitable, and pedagogically grounded AI integration can be built.

13.3.8 Continuous Reflection, Adaptation, and Responsible Innovation

A final principle concerns the need for *ongoing reflection* and *adaptation*. AI technologies are evolving rapidly, often faster than educational institutions can develop policies or pedagogical frameworks. Seen in this light, meaningful AI integration cannot be treated as a one-time implementation

process. Rather, it requires continuous evaluation, stakeholder engagement, and iterative refinement (Jin et al., 2025).

Responsible innovation involves asking not only whether a technology works but also whether it contributes positively to *educational goals*, *learner well-being*, and *social equity*. Educators and institutions must remain willing to revise practices as new evidence emerges. Technologies that appear promising today may reveal unforeseen limitations tomorrow, while future innovations may create opportunities that are currently difficult to anticipate.

This commitment to reflection is especially important because language education occupies a unique position at the intersection of communication, culture, identity, and human development. Decisions regarding AI adoption therefore have implications that extend far beyond instructional efficiency. They influence how learners construct knowledge, interact with others, and understand themselves within increasingly digital societies.

Comparative evidence supports treating this as an ongoing institutional obligation rather than a policy that, once written, can be left alone. A comparative policy analysis spanning the European Union, the United States, the United Kingdom, Australia, China, and Canada found that AI regulation in higher education is converging in its stated principles while persistent enforcement gaps remain around academic integrity and authorship—a divergence between principle and practice that only sustained, iterative review is likely to close (Uslu & Bozkurt, 2026). The cost of treating adoption as a one-time decision is illustrated by a cross-national analysis of industry-academia AI micro-credentials in the United States, the European Union, and Singapore, which found a 40% decline in faculty autonomy under market-driven governance arrangements as instructors shifted toward moderator roles (Soomro et al., 2026). Continuous reflection, on this evidence, functions less as a virtue than as a maintenance requirement for keeping governance aligned with a fast-moving technology.

13.3.9 The PAGE Framework: An Integrative Model for Responsible AI Integration

The eight principles developed in this chapter are not a checklist to be satisfied item by item; they cluster into a smaller set of load-bearing commitments that can guide decisions at every level of an educational system. This

book consolidates them into an integrative model referred to here as the **PAGE framework**—**P**edagogy-first, **A**gency-preserving, **G**overnance-bound, **E**quity-oriented AI integration. The name is intentional, in language education, turning the page is both a literacy act and a metaphor for moving beyond automation-first thinking.

Each dimension of PAGE absorbs two of the eight principles, translates them into a guiding question, and specifies what must be observably true in practice for the dimension to be satisfied.

Pedagogy-first (absorbing *Pedagogy Before Technology* and *Assessment Integrity and Human Oversight*) asks of every integration decision: *what learning problem does this solve, and how will we know?* In practice, the dimension is satisfied when AI adoption begins from an instructional or assessment need rather than from tool availability; when every AI-supported task carries an explicit statement of its learning objective; and when assessment designs specify the permitted level of AI involvement in advance rather than policing it after the fact.

Agency-preserving (absorbing *Human-Centered Design and Human Agency* and *AI Literacy as a Core Educational Competency*) asks: *who decides, and does the human capacity to decide grow or shrink over time?* The dimension is satisfied when teachers retain authority over pedagogical judgments and learners retain ownership of their learning goals; when systems are configured to expand rather than replace human skill—the augmentation test that runs through this book; and when AI literacy development is treated as a curricular obligation for both teachers and learners, not an optional enrichment.

Governance-bound (absorbing *Transparency and Explainability* and *Ethical Data Governance and Privacy Protection*) asks: *can we explain, audit, and answer for what the system does with learners and their data?* The dimension is satisfied when institutions can state what data AI systems collect, where it flows, and who is accountable for decisions informed by it; when consequential outputs (grades, placements, progress decisions) are explainable to those affected; and when AI use is disclosed honestly at every level, from student assignments to institutional reporting.

Equity-oriented (absorbing *Equity, Accessibility, and Inclusion* and *Continuous Reflection, Adaptation, and Responsible Innovation*) asks: *who benefits, who is burdened, and who is checking?* The dimension is satisfied

when access to AI-supported learning does not track existing socioeconomic divides; when systems are audited for performance across language varieties, accents, and learner populations; and when integration is treated as a provisional arrangement subject to scheduled review rather than a one-time adoption decision.

The four-part architecture proposed here is not without precedent in the broader literature on multidimensional integration frameworks for AI in language education, even though PAGE itself is offered as this book's own synthesis rather than as an instrument that has been separately tested. A comparable five-dimension framework for ethical, inclusive, and pedagogically-driven AI-assisted language education—covering intelligent tutoring systems, NLP-based chatbots, and AI-generated content—was validated through a survey of 48 language educators, who registered moderate-to-strong agreement across all five of its dimensions (Paul et al., 2026). That independent convergence on a multidimensional, educator-facing structure lends the general approach taken here some plausibility, even as the specific naming and boundaries of PAGE's four dimensions remain this book's own proposal.

The framework becomes operational when its four dimensions are crossed with the three educational functions that structure this book—teaching, learning, and assessment—and read through the Teacher–AI–Student triad developed in Chapters 1 and 3. Table 7. presents this operationalization: each cell states the core question a practitioner, institution, or policymaker should be able to answer affirmatively.

Dimension	In teaching	In learning	In assessment
Pedagogy-first	Does AI planning support serve stated objectives?	Do AI tools target identified learning needs?	Is AI's role defined by the construct, not convenience?
Agency-preserving	Does the teacher retain instructional authority?	Does learner self-regulation grow with use?	Do humans make and own consequential judgments?
Govern-ance-bound	Are AI-drafted materials disclosed and checkable?	Is learner data use transparent and minimal?	Are scores explainable and appealable?

Dimension	In teaching	In learning	In assessment
Equity-oriented	Do all learners access comparable support?	Are diverse varieties and profiles served fairly?	Are error rates audited across groups?

Table 7. The PAGE framework operationalized across teaching, learning, and assessment.

Three properties distinguish PAGE from a principle list. First, it is *diagnostic*: because each cell is phrased as an answerable question, the framework can be used to locate precisely where an integration effort is weak—a school may be pedagogy-first in teaching yet governance-poor in assessment. Second, it is *level-independent*: the same sixteen questions apply to a single lesson, an institutional policy, or a national program, changing only in scope. Third, it is *sequenced*: the dimensions are ordered deliberately, pedagogical purpose before agency, agency before governance, governance before equity audits—reflecting the argument, developed across Chapter 12, that ethical safeguards are meaningless when the underlying activity lacks educational value, and that equity monitoring presupposes governance structures capable of acting on what it finds.

The PAGE framework also names the standard against which this book's central thesis should be tested. *Augmentation*, as defended throughout the preceding chapters, is not an attitude but an achievable design condition: an AI integration that satisfies the twelve questions of Table 7 will, by construction, augment rather than automate—because it will have preserved the pedagogical purposes, human agencies, accountability structures, and equity commitments that automation-first adoption erodes. Chapter 14 applies this test prospectively to the emerging generation of agentic AI platforms, where autonomy on the machine side makes each PAGE question more consequential, not less.

13.4 Final Reflections for Educators and Researchers

Throughout this volume, AI has been examined as a pedagogical resource, an assessment technology, a catalyst for innovation, and a source of ethical and social debate. Collectively, the chapters demonstrate that AI is no longer a peripheral educational technology but an increasingly influential

force shaping how languages are taught, learned, assessed, and experienced. Yet the central message emerging from this book is clear: the future of language education will not be determined by technology alone. Rather, it will be shaped by the pedagogical, ethical, and institutional choices that educators, researchers, and policymakers make regarding the integration of AI into educational practice.

One of the most significant insights emerging from contemporary studies concerns the *evolving role of teachers*. Early narratives often portrayed AI as a potential replacement for human educators. Yet the evidence reviewed throughout this volume suggests a different trajectory. Rather than eliminating the need for teachers, AI increasingly amplifies the importance of uniquely human educational functions. As routine instructional tasks become increasingly automated, educators are positioned to *devote greater attention to critical thinking, intercultural dialogue, ethical reasoning, creativity, learner well-being, and identity development* (Hooshyar et al., 2025; Cukurova, 2024). The teacher of the future is therefore unlikely to be a transmitter of information but rather a designer of learning experiences, a facilitator of inquiry, and a mediator of human–AI interaction.

For researchers, the emergence of AI presents an equally transformative agenda. While early studies focused largely on technological effectiveness and system performance, future scholarship must increasingly address broader questions of *equity, inclusion, learner agency, and long-term educational impact* (Feng et al., 2025; Ma et al., 2025). Researchers must move beyond asking whether AI works to investigating for whom it works, under what conditions it is beneficial, and what unintended consequences may accompany its adoption. Particular attention should be devoted to multilingual contexts, indigenous language communities, under-resourced educational systems, and Global South perspectives that remain underrepresented in current AI research (Nyaaba & Zhai, 2025; Jin et al., 2025).

At the same time, the educational community must resist the temptation to view AI as a universal solution to complex educational challenges. Educational history demonstrates that technological innovations are often accompanied by unrealistic expectations regarding their transformative power. While AI can support learning in powerful ways, it cannot independently resolve structural inequalities, inadequate funding, curriculum limitations, or broader social disparities. Technology should therefore be understood as

one component within a larger educational ecosystem rather than as a substitute for sound pedagogy, institutional support, or educational policy.

A recurring theme throughout this volume has been the growing importance of *AI literacy*. Both educators and learners require the knowledge and skills necessary to understand how AI systems function, critically evaluate their outputs, recognize their limitations, and engage with them ethically and responsibly (Long & Magerko, 2020; Ng et al., 2021). AI literacy is rapidly becoming a foundational competency because participation in contemporary educational and professional environments increasingly requires the ability to navigate AI-mediated forms of communication, information, and decision-making. Educational institutions therefore have a responsibility not only to provide access to AI technologies but also to cultivate the critical capacities needed to use them wisely.

Perhaps the most important conclusion emerging from this book is that language education remains fundamentally a *human endeavor*. Language learning is not merely the acquisition of vocabulary and grammar; it is a process through which individuals construct identities, negotiate meaning, develop relationships, and participate in cultural communities. These dimensions of language cannot be fully reduced to algorithms, data analytics, or automated outputs. Even in highly AI-mediated environments, learners continue to require trust, empathy, mentorship, dialogue, and authentic human interaction (Carless, 2013; Rodgers & Raider-Roth, 2006).

As we move toward increasingly intelligent educational systems, the goal should not be to maximize automation but to maximize human flourishing. AI should be employed to reduce administrative burdens, enhance accessibility, support personalization, and expand educational opportunities. In doing so, it can create more space for the aspects of teaching and learning that remain uniquely human. The future of language education therefore lies not in replacing educators with machines, but in leveraging intelligent technologies to strengthen the relational, ethical, creative, and intercultural dimensions of educational practice.

In the end, the most important question is not what AI will become, but what educators, researchers, institutions, and societies choose to become in response to it. Artificial intelligence offers powerful tools, but it does not determine educational destiny. That responsibility remains firmly within human hands. The future of language education will therefore be shaped

not by artificial intelligence alone, but by the collective wisdom, ethical commitment, and pedagogical imagination with which humanity chooses to employ it.

In this sense, the future remains open. If guided by principles of responsibility, inclusion, critical reflection, and human-centered design, AI has the potential to enrich language education in ways previously unimaginable. If guided solely by efficiency, automation, or commercial interests, however, it risks reproducing many of the inequalities and limitations that education seeks to overcome. The challenge before us is therefore not simply technological but profoundly educational: to ensure that AI serves learning, humanity, and linguistic diversity rather than the reverse. The future of language education will be defined by how successfully we meet that challenge.

Pedagogical Companion for Chapter 13

Key Concepts: pedagogy before technology; human agency; transparency; data governance; equity; assessment integrity; AI literacy; continuous reflection; the PAGE framework.

Discussion Questions

1. Apply the twelve questions of Table 7 to one AI integration context you know well. Where does it fail first, and is the failure a design flaw or a governance flaw?
2. The PAGE dimensions are deliberately sequenced—pedagogy before agency before governance before equity. Construct a case where this ordering gives the wrong answer, if you can.
3. Which of the eight principles of Section 13.3 is most often honored only rhetorically in your context, and what would honoring it substantively cost?
4. The chapter claims augmentation is "an achievable design condition," not an attitude. What makes this claim falsifiable?
5. Write the AI-use declaration you believe should accompany your own institution's next strategy document.

Further Reading

- Miao, F., & Cukurova, M. (2024). *AI competency framework for teachers*. UNESCO.
- Cerratto Pargman, T., McGrath, C., & Milrad, M. (2025). Towards responsible AI in education. *CAEAI, 9*.
- Long, D., & Magerko, B. (2020). What is AI literacy? *CHI '20*.

Chapter 14. From Generative to Agentic: AI Agents and the Next Generation of Language Learning Platforms

14.1 Introduction: The Agentic Turn

The chapters of this book have examined a language education landscape transformed by generative AI—systems that respond, on request, with fluent text, speech, and feedback. As this book goes to press, a further shift is underway. Large language models are increasingly deployed not as conversational endpoints but as *agents*: systems that pursue goals over multiple steps, plan and revise their own actions, invoke external tools, retain memory across interactions, and coordinate with other agents (Xi et al., 2023). Landmark demonstrations such as generative agents that simulate believable social behavior over extended periods (Park et al., 2023) and orchestration frameworks that allow several specialized agents to collaborate on a task while interacting with human users (Q. Wu et al., 2023) established the technical pattern; the term *agentic AI* now designates this class of systems, distinguished from chatbots by *autonomy, proactivity, and multi-step goal pursuit*.

The definitional shift is sharper than the label alone suggests. *Agentic AI* is now distinguished from two adjacent categories with which it is easily conflated: from conventional, rule-bound AI agents, which execute narrowly scoped tasks without persistent goals, and from generative AI proper, which produces fluent output on request but does not act, plan, or monitor its own progress across time (Dwivedi et al., 2025). What marks a system as agentic, on this account, is the *combination of planning that spans multiple steps, memory that persists and accumulates across sessions, and the capacity to invoke tools in pursuit of a goal the system itself is tracking*, rather than one supplied fresh with each turn. Applied to education generally, this combination yields a wider taxonomy than the tutoring-chatbot model this book has largely examined so far: reviews of agentic AI in education distinguish intelligent-tutoring functions from adaptive testing and from autonomous learning companions that act on the learner's behalf without being prompted anew each time (Kostopoulos et al., 2025). Language learning, as the remainder of this chapter shows, is where these functions are beginning to converge within single deployed systems.

The distinction matters pedagogically, not simply technically. A chatbot waits; an agent initiates. A chatbot answers a learner's question; an agentic tutoring system may notice a persistent error pattern, schedule a review session, generate targeted practice, evaluate the learner's performance against a rubric, and decide—*on its own initiative*—what to do next. Every affordance and every risk documented in this book is thereby amplified. The personalization promised by adaptive systems (Chapters 7 and 8) becomes continuous and self-directing; the erosion of learner agency and productive struggle warned of in Chapters 8 and 12 acquires a mechanism that operates without waiting to be asked (Woollaston, Flanagan, & Wijerathne, 2026). This closing chapter therefore examines the emerging research on agentic AI in language learning through a structured review and then turns to the design questions facing the coming generation of agentic language learning platforms, using the PAGE framework of Chapter 13 as its evaluative lens.

The distinction carries particular weight for language education specifically, where the risk is not only cognitive but cultural. Conceptual analysis warns that language higher education adopting agentic AI passively rather than critically risks *cognitive dependency* and *cultural homogenization* (Sargsyan, 2025). The mechanism is not hard to see: a proactive system optimized for efficient progression has every incentive to standardize toward its own dominant training-language norms rather than toward the cultural and linguistic variation language teaching is meant to cultivate. That risk is a consequence of how a system is built, however, not an inevitable property of agentic design as such: a curriculum-aligned conversational agent developed and evaluated for Ghanaian teacher education through design science research shows that cultural and curricular responsiveness can be written into an agentic system's design brief from the outset rather than treated as an afterthought (Nyaaba et al., 2026). Whether a *proactive agentic platform* serves a learner's own linguistic and cultural development or steers it toward its training data's defaults is accordingly a design choice, not a foregone conclusion of the technology—the augmentation-over-automation commitment this book has argued throughout, now posed at the level of self-initiating interactions rather than of discrete, individually rejectable outputs.

14.2 Reviewing an Emerging Literature: Method

Because *agentic language learning* is a young research area, this chapter's claims rest on a deliberately transparent search procedure rather than an exhaustive systematic review. Literature was identified through structured queries of the Scite citation database combining agent-related vocabulary ("agentic AI," "AI agents," "multi-agent," "autonomous agents") with education- and language-specific terms ("language learning," "language education," "pedagogy," "teaching"), restricted to 2023–2026 to capture the post-LLM period. The most relevant sources were read in full or in extended excerpt, every cited record was checked for retraction and correction notices, and no source is cited in this chapter that could not be verified bibliographically. The resulting corpus is small—a handful of platform architectures, early empirical studies, and conceptual analyzes—and the synthesis below should be read accordingly: as a map of an emerging territory, not a settled body of findings.

14.3 What the Early Research Shows

14.3.1 Architectures: The Division of Pedagogical Labor

The clearest pattern across early systems is the decomposition of pedagogical work into specialized, coordinated agents. A multi-agent chatbot architecture for *English–Arabic language learning assigns separate agents to translation, example retrieval, review of the learner's own translations, and learning feedback, combining retrieval-based and generative models under a shared orchestration layer* (Aleedy, Atwell, & Meshoul, 2025). A system for Luxembourgish—a low-resource language poorly served by mainstream tools—uses *cooperating agents over an LLM to deliver conversational practice and feedback that single-model systems could not sustain* (Nouzri et al., 2024). In adjacent professional-skills training, the same triadic pattern recurs: one *agent generates the practice scenario, one conducts the persona-driven spoken interaction, and one evaluates the performance against explicit standards and produces structured feedback* (De Marez et al., 2025). At platform scale, structured agent specification becomes an engineering discipline of its own: profile-driven systems define hundreds of educational agents through explicit declarations of role, skills, tools, and pedagogical expertise, arguing that educational capability grows

with the structural richness of these specifications as much as with the size of the underlying model (Wu et al., 2026).

The pattern extends beyond text-based tutoring. An *adaptive gamified architecture layering multiple agents over an experience-tracking (xAPI) and blockchain infrastructure produced significantly higher language proficiency and motivation* than a control group, suggesting the division of labor generalizes across delivery formats rather than being an artifact of chatbot design specifically (Hammami & Alomair, 2026). In *mixed reality*, a system in which multiple agents jointly sustain a group conversation was tested with 21 second-language learners, increasing engagement and willingness to communicate while easing speaking-related fear (Fernandez-Espinosa et al., 2025)—evidence that the division of labor can be distributed across *simulated interlocutors*, not only across pedagogical functions. Conceptual proposals push the decomposition further still: one *architecture assigns agents to consulting, lecturing, methodology, advising, and support roles modeled on a human language-teaching team* (Slyusar, 2025), and a separate theory-driven proposal specifies evidentiary rules by which a multi-agent system should decide, from accumulating evidence of a learner's state, when and how to scaffold, building on the intelligent-tutoring-system lineage of Chapter 7 to replace undifferentiated support with a principled division of labor among agents (Cohn et al., 2026). The direction of travel across these designs is consistent: *as architectures mature, the number of distinct agent roles grows, and the roles increasingly map onto functions historically distributed among a teacher, a curriculum designer, and an assessor, rather than remaining confined to a single generation-interaction-evaluation triad.*

Review-level evidence corroborates what these individual designs suggest. A PRISMA-guided synthesis of multi-agent cognitive architectures for adaptive tutoring found individualized feedback superior to conventional methods across the studies reviewed, while identifying neuroadaptive and emotion-responsive tutoring as capabilities such architectures still lack (Gupta & Heggond, 2026).

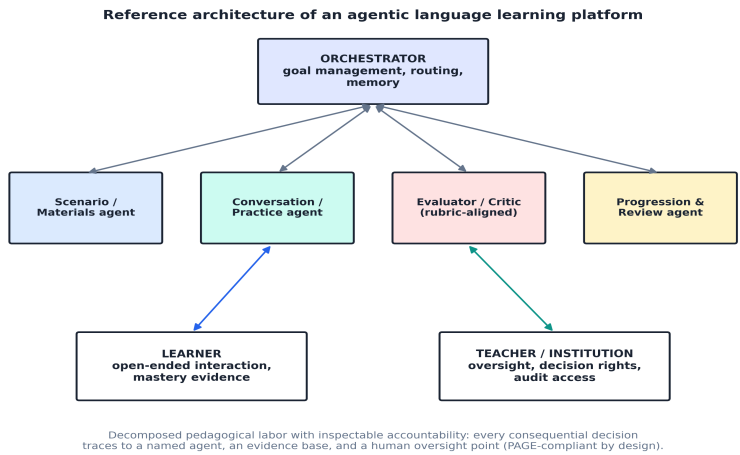


Figure 4. Reference architecture of an agentic language learning platform

Two features of these architectures deserve emphasis. First, they instantiate—in software—the separation of concerns that this book has argued should govern human–AI collaboration: generation, interaction, and evaluation are distinct functions with distinct accountability requirements, and multi-agent designs make that separation inspectable. Second, the *critic* or *evaluator* agent, present in nearly every reviewed architecture, is where assessment theory enters the system: rubric alignment, scoring reliability, and feedback quality are no longer properties of a monolithic model's improvisation but design commitments of a dedicated component (De Marez et al., 2025; Scaria, Kennedy, & Subramani, 2026).

These architectural gains are not free. Coordinating several agents multiplies the points at which a pedagogical interaction can fail: latency accumulates across sequential agent calls, disagreement between agents must be resolved by some further mechanism that carries its own failure modes, and no single component holds a unified model of the learner, so an error in one agent's output can propagate into another's decision without an obvious point of interception. A systematic review of 143 studies on agentic AI architectures—spanning the planning, memory, and reflection mechanisms on which such systems depend—catalogues exactly these categories of cost at a *general level: technical, ethical, and security challenges* that a pedagogical ensemble inherits along with the architecture, even though

none of them is specific to education (Bandi et al., 2025). The trade-off between opacity and inspectability that Chapter 2's account of AI system archetypes already raised recurs here in sharpened form: a monolithic model is opaque but has one locus of accountability; a multi-agent system is inspectable component by component, yet the ensemble's overall behavior—which agent's judgment prevailed, and why—can be harder to reconstruct than a single model's, precisely because the labor has been divided.

14.3.2 Pedagogical Functions: From Practice Partner to Progression Engine

Functionally, the reviewed systems extend AI's role along a continuum. At one end lies *conversational practice*—the *digital interlocutor* familiar from Chapters 5 and 7, now *more robust and more proactive*. At the other end lies something qualitatively new: agentic systems that *manage learning trajectories*. The most developed example grounds progression decisions in demonstrated conversational competence rather than quiz recall: role-specialized agents independently score a learner's open-ended conversation for grammar, vocabulary, and interactive communication against *CEFR-aligned rubrics*, debate their disagreements, synthesize consensus scores, and then decide what the learner should review, scheduling spaced re-practice of weak items and requiring demonstrated mastery before advancement (Scaria et al., 2026). An eight-week study with 180 A2-level learners found that this combination of rubric-aligned scoring, targeted recommendation, spaced review, and mastery-based progression produced better learning outcomes than feedback alone (Scaria et al., 2026).

The *practice-partner* end of this continuum is itself evolving, not static. A conversation-analytic comparison of human tutors and an autonomous-turn-taking AI agent across role-play tasks with 75 Japanese university students found the agent capable of managing turn allocation on its own initiative rather than only responding to learner turns, even as it handled repair sequences—central to interactional competence—differently from a human interlocutor (Eguchi et al., 2025). Between this proactive-but-local behavior and Scaria et al.'s (2026) full progression engine sit systems that manage one competency without yet managing the learner's overall trajectory: a generative AI agent built on the E.F.F.E.C.T. framework, piloted with Hong Kong primary students, scaffolded self-regulated L2 writing and reduced writing anxiety—proactive within a

single skill, but not extended to sequencing what the learner studies next (Wong, 2026).

Placed against this continuum, the corpus surveyed in this chapter clusters heavily toward the practice-partner end. Of the individual architectures and studies introduced across this section, only Scaria et al.'s (2026) system closes the loop from evaluation to a binding decision about what the learner does next and gates advancement on demonstrated mastery; the rest extend proactivity within a bounded task—a turn taken, a hint offered, a scaffold adjusted—without a persisting decision that reorganizes the learner's subsequent path. On the evidence reviewed here, the *progression-engine function* is not where the field currently sits; it is the direction one documented system is testing, and the direction toward which this chapter's design and governance questions are ultimately addressed.

This progression-engine function is the *pedagogically consequential frontier*. It moves agentic systems into territory that this book has consistently reserved for *human judgment*: deciding what a learner should do next. The early evidence suggests the decision quality can be respectable when the deciding is disciplined—*validated rubrics, multi-agent cross-examination of scores, expert-annotated benchmarks* (Scaria et al., 2026). But it equally demonstrates the stakes: *an autonomous progression engine* with poor construct validity would automate mislearning at scale, which is why the assessment-theoretic apparatus of Chapters 10 and 11—*construct validity, fairness auditing, explainability, washback*—applies to agentic platforms with full force.

The stakes are not hypothetical, and they do not require full autonomy to materialize. A randomized experiment with 117 university students found that a ChatGPT-assisted group improved essay scores the most of any condition yet showed no corresponding advantage in knowledge gain or transfer, evidence of metacognitive laziness (Fan et al., 2024)—and this occurred under generative assistance far short of autonomous progression management. If output quality can already decouple from learning this sharply under a system that only responds when asked, a system that also decides what the learner reviews next, and declares mastery achieved, closes off the very struggle on which the transfer measured in that study depends. The construct-validity safeguards this chapter invokes below are therefore not a precaution against a remote failure mode but a response to a mechanism already observed at the generative stage of the very continuum this section has traced.

14.3.3 Empirical Evidence: Promising, Thin, and Uneven

The empirical base is developing rapidly but remains thin. A randomized field experiment on generative AI agents in language learning provides rare causal evidence from a real deployment context (Choi, Lee, & Baek, 2025). Usability and effectiveness evaluations of the English–Arabic multi-agent system with 40 undergraduate students and their instructors report *enhanced engagement* and *contextually appropriate language* support (Aleedy et al., 2025). The mastery-progression study described above offers the strongest learning-outcome evidence to date (Scaria et al., 2026). Beyond these, much of the literature consists of *architecture papers with small-scale pilots* (De Marez et al., 2025; Nouzri et al., 2024) or *conceptual and regulatory analyzes* (Woollaston et al., 2026; Evangeline, 2025).

Points of agreement across this early corpus are nonetheless discernible: multi-agent decomposition outperforms monolithic prompting for pedagogically structured tasks; *rubric-anchored evaluator agents* can approach acceptable agreement with expert human judgment under careful protocol design; and learner-facing benefits concentrate in engagement, practice volume, and feedback immediacy. The open contradictions are equally instructive. Where platform-engineering work treats greater agent autonomy and richer capability profiles as straightforwardly desirable (Wu et al., 2026), pedagogically grounded analyzes argue that proactive assistance can dissolve precisely the desirable difficulties—ambiguity, retrieval effort, productive struggle—on which durable learning depends, and call for *intentional friction*, dynamic fading of scaffolds, and human oversight as first-class design requirements (Woollaston et al., 2026). This tension—capability scaling versus pedagogical restraint—is the agentic-era form of the augmentation-versus-automation argument that has organized this book.

This thinness is worth stating precisely, since the asymmetry it discloses is itself a finding. What counts as *agentic-specific evidence* is, with one exception, a set of architecture papers validated by a small user study: 40 undergraduates and 4 faculty for the English–Arabic system (Aleedy et al., 2025), 21 learners for a mixed-reality multi-agent conversation tool (Fernandez-Espinosa et al., 2025), 75 Japanese university students contributing 24 analyzed interactions to a qualitative comparison of human- and AI-agent-delivered roleplay (Eguchi et al., 2025), a single exploratory micro-

genetic pilot scaffolding L2 writing among Hong Kong primary pupils (Wong, 2026). Even the randomized field experiment already noted above remains a single instance; Scaria et al.'s (2026) 180-learner, eight-week trial is the only other study in this corpus approaching trajectory-scale, causal design. A bibliometric mapping of 774 publications on intelligent-agent-supported language learning between 2005 and 2025 traces two decades of growth in the wider field of chatbots and virtual tutors (Liu, 2026); the specifically agentic slice of that literature, by the chapter's own search parameters, is barely three years old. This is not the evidentiary condition this book's chatbot-era claims have rested on. Those draw on a systematic review and meta-analysis of ChatGPT's effects on student engagement (Heung & Chiu, 2025), a multi-level meta-analysis of automated writing feedback (Fleckenstein et al., 2023), a dedicated meta-analysis of AI in L2 learning (Wu, 2024), and randomized comparisons with samples of 117 (Fan et al., 2024) and 114 university students (Le et al., 2025). A reader who has extended this book's cautious confidence in chatbot-level generative AI claims should not extend the same confidence to agentic claims at this stage; the two literatures are not at the same evidentiary maturity, and the sections that follow should be read with that calibration in mind.

14.3.4 Governance: Autonomy Raises the Stakes

Because agentic systems act, monitor, and decide in real time, they intensify every governance concern examined in Chapter 12: continuous data collection becomes constitutive of the system's operation; accountability for decisions distributes across agent ensembles; and the opacity of a single model is compounded by the emergent behavior of several models interacting. Early regulatory analysis argues that general data-protection law and generic AI ethics codes underdetermine these systems, and proposes multi-layered governance—legal accountability, data governance, bias mitigation, algorithmic audit, and stakeholder involvement—tailored to agentic autonomy in educational settings (Evangeline, 2025). For language education specifically, the fairness findings of Chapter 11 carry over with added urgency: an agent ensemble that penalizes non-standard varieties would not just score unfairly on one occasion but would steer learning trajectories unfairly over months.

Each of these three intensifications deserves unpacking, since each maps onto a distinct failure mode. *Distributed accountability* is the most

immediately consequential: when an orchestrator, a scenario-generation agent, an interaction agent, and a critic agent jointly produce a placement decision or a progression gate, responsibility for a wrong or harmful decision has nowhere obvious to sit—not with any single model's output, not with the human who configured the ensemble, and not with the institution that purchased it. A survey of 75 university educators examines exactly this terrain: perceptions of agentic AI's applications, challenges, governance mechanisms, and institutional readiness for adoption in higher education (Brohi & Jhanjhi, 2026); a qualitative systematic review of 71 studies from 2020–2024, using thematic coding and SWOT synthesis, maps the resulting governance tensions across global education (Mariyono, 2026). Neither source resolves who is accountable when an ensemble, rather than a single model, gets a learner's trajectory wrong, because that resolution does not yet exist in the literature this chapter has reviewed. A sharper challenge comes from Mishra and Henriksen (2026), who argue that "*agentic AI*" marketing borrows the moral vocabulary of learner agency while such platforms function as institutional "*double agents*," eroding the vocabulary available to defend student and teacher self-determination. That critique lands on this book's own terms directly: a platform billed as agency-preserving in the PAGE sense must be distinguishable, in practice, from one that merely borrows the language of agency to license autonomous institutional action.

Continuous data collection is not an incidental privacy cost of agentic systems but a condition of their operation: an agent that plans, remembers, and revises its own actions across sessions can only do so by holding a continuously updated model of the learner, which puts the data-minimization principle this book has urged since Chapter 12 in structural tension with the very mechanism that makes agentic personalization work. *Compounded opacity* follows from the same architecture. A systematic review of agentic AI in autonomous tutoring systems finds real progress in multi-agent architectures and emotion-responsive instruction alongside persistent gaps in transparency, data protection, and equitable access (Meti, 2026), and a scoping review and thematic analysis of agentic AI and large language models across six domains, including education, catalogues goal misalignment and limited human oversight as risks multi-agent systems add on top of the bias and misinformation risks they inherit from their underlying models (Brohi et al., 2025). Where a single model's opacity is already a documented

problem for *explainability* and *washback* (Chapters 10 and 11), an ensemble whose emergent behavior need not trace to any individually opaque component multiplies that problem rather than merely repeating it. Regulatory responses built on multi-layered governance (Evangeline, 2025) are necessary, but on the evidence reviewed here they remain proposals: no source in this corpus reports an implemented audit trail or decision log for a deployed agentic language-learning platform.

14.3.5 Evaluation Methodology: Validating the Validators

The review implies a question it never poses directly: what would it actually take to validate an evaluator agent, or an autonomous progression engine, against the assessment-theoretic apparatus of Chapters 10 and 11? The architectures reviewed above already treat evaluation as a design commitment rather than an afterthought—a multi-agent English-teaching architecture specifies evaluation metrics for its consulting, lecturing, and advising agents (Slyusar, 2025), and an *exploratory human-AgenticAI architecture* builds assessment into its coordination layer, where AI assistants help learners and tutors co-create and assess outcomes (Williams, 2025)—but proposing a metric is not validating one, and neither source reports the reliability or validity evidence Chapter 10 treats as a precondition for a score that carries consequences for a learner.

A defensible validation program would need at least three elements this literature does not yet supply. First, *inter-agent reliability* benchmarked against human judgment, not merely against other agents: Scaria et al.'s (2026) multi-agent debate framework reports 0.23 score-variation agreement among its own scoring agents and 90.91% recommendation acceptability, but no study in this corpus compares an evaluator agent's scores against certified human CEFR raters using the generalizability-theory or many-facet Rasch designs Chapter 10 treats as baseline for automated scoring. Second, *construct validation* of the progression decision itself, independent of the scoring that feeds it: evidence that a mastery threshold set by a critic agent tracks a defensible construct of readiness, rather than whatever threshold happens to produce an acceptable-looking recommendation-acceptance rate. Third, *fairness auditing* conducted on the deployed ensemble rather than on any one component in isolation, since bias can enter or amplify at the orchestration layer even when every individual agent passes

an isolated audit—the ensemble-level counterpart to the dialect-based unfairness Chapter 11 documents in single-model scoring.

None of this indicts the systems reviewed here so much as it names a gap the field has not had time to close. Theory-driven proposals such as the *Evidence-Decision-Feedback framework*, which grounds multi-agent scaffolding decisions in explicit evidentiary rules (Cohn et al., 2026), point toward the kind of specification this validation would need to test—but a specification is a hypothesis, not a psychometric result. Until evaluator agents and progression engines undergo the reliability, validity, and fairness scrutiny outlined above, the Governance-bound cell of the PAGE framework cannot be marked satisfied for any agentic language-learning platform, however thoroughly its architecture is documented.

14.4 Case: ApiLingua—Designing an Agentic Language Learning Platform

The design questions surfaced by this review are concrete for the platforms now being built. ApiLingua, an agentic AI-supported foreign language learning platform under development in Türkiye illustrates how the architectural patterns of Section 14.3.1 translate into product decisions. The platform follows the division-of-labor pattern documented across the reviewed systems: an orchestrator coordinating specialized agents for conversational practice, targeted skill exercises, error-pattern analysis, and rubric-based formative assessment. Its progression logic combines mastery evidence from open-ended interaction with spaced review of diagnosed weaknesses, the approach for which Scaria et al. (2026) provide the strongest current evidence.

Read through the PAGE framework, the platform's design checklist becomes explicit. *Pedagogy-first:* each agent's intervention must trace to a stated learning objective and a defensible construct—particularly for any agent that scores or gates progression. *Agency-preserving:* proactive moves (initiating review, adjusting difficulty) should default to proposal rather than imposition, keeping the learner's and—in institutional deployments—the teacher's decision rights intact, with scaffolding that demonstrably fades as competence grows. *Governance-bound:* the platform must be able to answer, for every consequential decision, which agent decided, on what

evidence, and under whose accountability; learner data flows must be minimal, disclosed, and auditable. *Equity-oriented:* scoring and speech-facing agents must be audited across accents, varieties, and learner profiles before high-stakes use—the lesson of Chapter 11 and of the dialect-prejudice findings reviewed there (Hofmann et al., 2024).

Presented this way, ApiLingua serves the argument of this book not as a promotional exhibit but as a demonstration that the augmentation standard is buildable: the same agentic capabilities that could automate a learner's journey can be architected to augment it, if the twelve questions of Table 7 are treated as requirements rather than aspirations.

14.5 A Research Agenda for Agentic Language Learning

The review above implies a research agenda that extends Chapter 13's into the agentic era. Five priorities stand out, and the evidence gap described in this chapter now gives each a specific design. First, *causal evidence at trajectory scale*: cluster-randomized trials assigning intact classes, not individual learners, to agent-managed, teacher-managed, and self-managed progression across a full course, with learning outcomes, self-regulation, and autonomy measured as joint, pre-registered dependent variables rather than the single post-test typical of the studies reviewed above. Second, *construct validation of evaluator agents*, taken up directly in Section 14.3.5: generalizability-theory or many-facet Rasch comparisons of agent scores against certified human raters, independent validation of the progression thresholds those scores feed, and fairness audits run on the deployed ensemble—the full psychometric treatment that Chapters 10 and 11 develop for automated assessment generally, in place of the recommendation-acceptance rates the literature currently reports. Third, *agency dynamics*: whether sustained exposure to proactive agents strengthens or atrophies learner self-regulation is an open empirical question on which the augmentation argument ultimately rests (Woollaston et al., 2026) and answering it needs longitudinal designs with behavioral measures of self-regulation, not the correlational design already used with 280 Saudi higher-education students (Alqurni, 2026). Fourth, *low-resource and equity contexts*: the Luxembourgish case (Nouzri et al., 2024) hints that agentic architectures may serve languages and populations that mainstream tools neglect; testing whether this potential is realized, or whether agentic sophistication instead

concentrates in high-resource languages, is a question of consequence for the Global South concerns raised in Chapter 12, and it calls for controlled comparisons on the model Hammami and Alomair (2026) report, deliberately replicated in typologically diverse, lower-resource languages rather than in already well-resourced ones. Fifth, *governance instrumentation*: audit trails, decision logs, and explanation interfaces for agent ensembles need to move from regulatory proposal (Evangeline, 2025) and educator survey (Brohi & Jhanjhi, 2026) to field trials that test whether such instruments actually let an institution answer, after the fact, which agent decided what and on what evidence—the question Section 14.3.4 shows current platforms cannot yet answer.

14.6 Conclusion: The Page That Turns Itself?

Agentic AI gives the central question of this book its sharpest form. A book, like a lesson, waits for its reader; an agent does not wait. Systems that plan, act, and decide on the learner's behalf can either absorb the pedagogical labor that makes learning happen—or be designed to hand that labor back, at the right moments, in the right amounts, to the humans whose growth is the point. The early literature reviewed here shows both futures under construction. The architectures are ready; the evidence is young; the design choices are still open. What this book has offered—*the augmentation thesis, the Teacher–AI–Student triad, the assessment-theoretic safeguards,* and *the PAGE framework*—is a way of making those choices deliberately. The page does not turn itself. Turning it well remains, and should remain, *a human act.*

Pedagogical Companion for Chapter 14

Key Concepts: agentic AI; autonomy and proactivity; multi-agent architectures; orchestrator and critic agents; mastery-based progression; spaced review; intentional friction; scaffolding fading; governance of agent ensembles.

Discussion Questions

1. A chatbot waits; an agent initiates. Identify three classroom moments where machine initiative would help learning, and three where it would harm it.
2. The reviewed architectures separate scenario generation, interaction, and evaluation into distinct agents. Why does this separation matter for accountability, not just engineering?
3. Mastery-based progression governed by evaluator agents moves "what should the learner do next?" into software. Under what validation conditions, if any, is this delegation defensible?
4. Design the "intentional friction" for an agentic language platform: where exactly should the system slow the learner down, and how would you know the friction is productive rather than merely frustrating?
5. Apply the PAGE framework to any agentic learning product you know. Which of the sixteen cells is hardest for an autonomous system to satisfy in principle?

Further Reading

- Xi, Z., et al. (2023). The rise and potential of large language model based agents: A survey. arXiv.
- Scaria, N., Kennedy, S. J. J., & Subramani, D. (2026). Learning in blocks. arXiv.
- Woollaston, S., Flanagan, B., & Wijerathne, I. (2026). Agentic AI and pedagogical best practice. arXiv.

References

Abrusci, L., Dabaghi, K., D'Urso, S., & Sciarrone, F. (2025). AI4Design: A generative AI-based system to improve creativity in design–A field evaluation. Computers and Education: Artificial Intelligence, 8, Article 100401. https://doi.org/10.1016/j.caeai.2025.100401

Abualrob, M. (2025). From reflection to re-reflection: ChatGPT and transformative learning in preservice elementary science education. *International Journal of Learning, Teaching and Educational Research*, 24(10), 769-793. https://doi.org/10.26803/ijlter.24.10.37

Acosta-Enriquez, B. G., Arbulú Ballesteros, M. A., Guzman Valle, M. de los A., Morales Angaspilco, J. E., Aquino Lalupú, J. del R., Bravo Jaico, J. L., Germán Reyes, N. C., Alarcón García, R. E., & Janampa Castillo, W. E. (2025). The mediating role of academic stress, critical thinking and performance expectations in the influence of academic self-efficacy on AI dependence: Case study in college students. Computers and Education: Artificial Intelligence, 8, Article 100381. https://doi.org/10.1016/j.caeai.2025.100381

Acuna, R., & Bansal, A. (2022). Using programming autograder formative data to understand student growth. In Proceedings of the 2022 IEEE Frontiers in Education Conference (FIE) (pp. 1–8). IEEE. https://doi.org/10.1109/FIE56618.2022.9962650

Adams, C., Pente, P., Lemermeyer, G., & Rockwell, G. (2023). Ethical principles for artificial intelligence in K–12 education. Computers and Education: Artificial Intelligence, 4, Article 100131. https://doi.org/10.1016/j.caeai.2023.100131

Adil, J. G., Jr. (2025). AI in education: A systematic literature review of emerging trends, benefits, and challenges. *Seminars in Medical Writing and Education*, 4, 795. https://doi.org/10.56294/mw2025795

Agyare, B., Asare, J., Kraishan, A., Nkrumah, I., & Adjekum, D. K (2025). A cross-national assessment of artificial intelligence (AI) Chatbot user perceptions in collegiate physics education. Computers and Education: Artificial Intelligence, 8, Article 100365. https://doi.org/10.1016/j.caeai.2025.100365

Ahimsa, A. F. H., & Khawa, D (2025). The impact of artificial intelligence on English language teaching: Opportunities and challenges in technology era. IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature. https://doi.org/10.24256/ideas.v13i1.6270

Akgun, S., & Greenhow, C. (2022). Artificial intelligence in education: Addressing ethical challenges in K–12 settings. AI and Ethics, 2(3), 431–440. https://doi.org/10.1007/s43681-021-00096-7

Al-Aghbari, N., Alkharusi, H., & Al Fazari, M. (2025). AI-based formative assessment and its impact on the academic achievement and perceptions of the classroom assessment environment. *An-Najah University Journal for Research - B (Humanities)*, 39(11), 893–906. https://doi.org/10.35552/0247.39.11.2484

Al Darayseh, A. (2023). Acceptance of artificial intelligence in teaching science: Science teachers' perspective. Computers and Education: Artificial Intelligence, 4, Article 100132. https://doi.org/10.1016/j.caeai.2023.100132

Al-khresheh, M. H. (2024). Bridging technology and pedagogy from a global lens: Teachers' perspectives on integrating ChatGPT in English language teaching. Computers and Education: Artificial Intelligence, 6, Article 100218. https://doi.org/10.1016/j.caeai.2024.100218

Alcazar-Gonzales, J. (2026). Redefiniendo el construtivismo pedagógico como consecuencia de la inteligência artificial [Redefining pedagogical constructivism as a consequence of artificial intelligence]. *Millenium - Journal of Education, Technologies, and Health, 2*(21e), e41959. https://doi.org/10.29352/mill0221e.41959

Aldreabi, H., Alhur, M., Al-Okaily, M., Qasim, D., Dahdoul, N. K., & Shiyyab, F. S (2025). Unveiling the dynamics of generative AI adoption: A business intelligence analysis through topic modeling-based bibliometric study. Computers and Education: Artificial Intelligence, 9, Article 100456. https://doi.org/10.1016/j.caeai.2025.100456

Aldreabi, H., Dahdoul, N. K. S., Alhur, M., Alzboun, N., & Alsalhi, N. R. (2025). Determinants of student adoption of generative AI in higher education. Electronic Journal of e-Learning, 23(1), 15–33. https://doi.org/10.34190/ejel.23.1.3599

Aleedy, M., Atwell, E., & Meshoul, S. (2025). A multi-agent chatbot architecture for AI-driven language learning. Applied Sciences, 15(19), 10634. https://doi.org/10.3390/app151910634

Alfredo, R., Echeverria, V., Jin, Y., Yan, L., Swiecki, Z., Gašević, D., & Martinez-Maldonado, R. (2024). Human-centred learning analytics and AI in education: A systematic literature review. Computers & Education: Artificial Intelligence, 6, Article 100215. https://doi.org/10.1016/j.caeai.2024.100215

Alhazmi, A. A., & Muftah, M. (2025). Evaluating ChatGPT's reliability in second language acquisition (SLA): Insights on language skills and technology's role. *The Electronic Journal of e-Learning*, 23(1), 96-112. https://doi.org/10.34190/ejel.23.1.3762

Alhusaiyan, E. A. (2024). A systematic review of current trends in artificial intelligence in foreign language learning. *Saudi Journal of Language Studies*, 5(1), 1-16. https://doi.org/10.1108/sjls-07-2024-0039

Ali, O., Murray, P. A., Momin, M., Dwivedi, Y. K., & Malik, T. (2024). The effects of artificial intelligence applications in educational settings: Challenges and strategies. Technological Forecasting and Social Change, 199, Article 123076. https://doi.org/10.1016/j.techfore.2023.123076

Alkhutaba, M., & Alnajjar, K. (2026). Artificial intelligence and child second language learning. *An-Najah University Journal for Research - B (Humanities)*, *40*(3), 309–322. https://doi.org/10.35552/0247.40.2.2748

Allehyani, S. H., & Algamdi, M. A. (2023). Digital competences: Early childhood teachers' beliefs and perceptions of ChatGPT application in teaching English as a second language (ESL). *International Journal of Learning, Teaching and Educational Research*, 22(11), 343-363. https://doi.org/10.26803/ijlter.22.11.18

Alqurni, J. (2026). Exploring the role of agentic AI in fostering self-efficacy, autonomy support, and self-learning motivation in higher education. *Frontiers in Artificial Intelligence*, 9, Article 1738774. https://doi.org/10.3389/frai.2026.1738774

Alshorman, S. (2024). The readiness to use AI in teaching science: Science teachers' perspective. *Journal of Baltic Science Education*, 23(3), 432-448. https://doi.org/10.33225/jbse/24.23.432

Altamirano, S. F. S., Maquen-Niño, G. L. E., Roldán, C. M. G., Herrera, A. P., & Cárdenas, D. M. C. (2026). Generative artificial intelligence in higher education: A systematic review of perceptions, implementation and pedagogical transformation. *Review of Education*, 14(1), Article e70152. https://doi.org/10.1002/rev3.70152

Alzahrani, A., & Rafi, A. S. M. (2026). Emergent gendered AI learning ecologies in AI-mediated writing tasks: A critical comparative case study. *Interactive Learning Environments*, 1–15. https://doi.org/10.1080/10494820.2026.2683034

Amouri, H., Haroud, S., Ouchaouka, L., & Saqri, N. (2025). Acceptability of artificial intelligence in inclusive education: A TAM2-based study among preservice teachers. *Frontiers in Artificial Intelligence*, 8. https://doi.org/10.3389/frai.2025.1616327

Anders, A. D., & Speltz, E. D. (2025). Developing generative AI literacies through self-regulated learning: A human-centered approach. Computers and Education: Artificial Intelligence, 9, Article 100482. https://doi.org/10.1016/j.caeai.2025.100482

Annapureddy, R., Fornaroli, A., & Gatica-Perez, D. (2025). Generative AI literacy: Twelve defining competencies. Digital Government: Research and Practice, 6(1), 1-21. https://doi.org/10.1145/3685680

Aranberri, N., & Iñurrieta, U. (2024). When minoritized languages encounter MT. *The Journal of Specialised Translation*, (41), 179-205. https://doi.org/10.26034/cm.jostrans.2024.4718

Armas-Cervantes, A., Abedin, E., & Taymouri, F. (2024). Dusting for fingerprints: Tracking online student engagement. Computers & Education: Artificial Intelligence, 6, Article 100232. https://doi.org/10.1016/j.caeai.2024.100232

Aryadoust, V., Zakaria, A., & Jia, Y (2024). Investigating the affordances of OpenAI's large language model in developing listening assessments. Computers and Education: Artificial Intelligence, 6, 100204. https://doi.org/10.1016/j.caeai.2024.100204

Aslan, S., Alyuz, N., Li, B., Durham, L. M., Shi, M., Sharma, S., & Nachman, L. (2025). An early investigation of collaborative problem solving in conversational AI-mediated learning environments. Computers and Education: Artificial Intelligence, 8, Article 100393. https://doi.org/10.1016/j.caeai.2025.100393

Atre, S. (2026). Legal challenges of agentic AI systems in education and employment decision-making. *Indian Journal of Computer Science and Technology*, 44-47. https://doi.org/10.59256/indjcst.20260501007

Attia, A. A., Liu, J., Ai, W., Demszky, D., & Espy-Wilson, C. (2024). Kid-Whisper: Towards bridging the performance gap in automatic speech recognition for children vs. adults. https://doi.org/10.48550/arXiv.2309.07927

Avetisyan, H., & Broneske, D. (2023). Large language models and low-resource languages: An examination of Armenian NLP. In *Findings of the Association for Computational Linguistics: IJCNLP-AACL 2023* (pp. 199-210). https://doi.org/10.18653/v1/2023.findings-ijcnlp.18

Awidi, I. T. (2024). Comparing expert tutor evaluation of reflective essays with marking by generative artificial intelligence (AI) tool. Computers & Education: Artificial Intelligence, 6, Article 100226. https://doi.org/10.1016/j.caeai.2024.100226

Ayedoun, E., Hayashi, Y., & Seta, K. (2019). Adding communicative and affective strategies to an embodied conversational agent to enhance second language learners' willingness to communicate. International Journal of Artificial Intelligence in Education, 29, 29–57. https://doi.org/10.1007/s40593-018-0171-6

Bah, Y. M., & Artaria, M. D. (2020). Corona virus (COVID-19) and education for all achievement: Artificial intelligence and special education needs—achievements and challenges. *COUNS-EDU: The International Journal of Counseling and Education*, 5(2), 64-70. https://doi.org/10.23916/0020200528630

Baker, R. S., & Hawn, A (2022). Algorithmic bias in education. International Journal of Artificial Intelligence in Education, 32(4), 1052-1092. https://doi.org/10.1007/s40593-021-00285-9

Bandi, A., Kongari, B., Naguru, R., Pasnoor, S., & Vilipala, S. V. (2025). The rise of agentic AI: A review of definitions, frameworks, architectures, applications, evaluation metrics, and challenges. *Future Internet*, 17(9), Article 404. https://doi.org/10.3390/fi17090404

Bandura, A (1997). Self-efficacy: The exercise of control. W. H. Freeman.

Barrot, J. S. (2026). Generative artificial intelligence for automated essay scoring: Exploring teacher agency through an ecological perspective. *Assessing Writing*, 67, 100990. https://doi.org/10.1016/j.asw.2025.100990

Bax, S. (2003). CALL—past, present and future. *System*, 31(1), 13-28. https://doi.org/10.1016/S0346-251X(02)00071-4

Belkbir, R. (2024). Integrating the 6Cs in Moroccan higher education: A framework for inclusive AI learning design. *Arab World English Journal*, 15(3), 217-234. https://doi.org/10.24093/awej/vol15no3.13

Belkina, M., Daniel, S., Nikolic, S., et al. (2025). Implementing generative AI (GenAI) in higher education: A systematic review of case studies. Computers and Education: Artificial Intelligence, 8, Article 100407. https://doi.org/10.1016/j.caeai.2025.100407

Ben-Zion, Y., Yakov, S., Abramovitch, E., Balter, G., & Davidovitch, N. (2025). AI-based teaching evaluations: How well do they reflect student perceptions? Computers and Education: Artificial Intelligence, 9, Article 100448. https://doi.org/10.1016/j.caeai.2025.100448

Benedetto, L., Gaudeau, G., Caines, A., & Buttery, P. (2025). Assessing how accurately large language models encode and apply the Common European Framework of Reference for Languages. Computers and Education: Artificial Intelligence, 8, Article 100353. https://doi.org/10.1016/j.caeai.2024.100353

Bengsch, G. (2024). Redefining traditional pedagogy: The integration of machine learning in the contemporary language education classroom. In F. Pan (Ed.), AI in language teaching, learning, and assessment (pp. 195–221). IGI Global. https://www.igi-global.com/gateway/chapter/339097

Benson, P. (2013). Teaching and researching autonomy in language learning (2nd ed.). Routledge.

Bergdahl, N., & Sjöberg, J. (2025). Transformation, support needs and AI, in K-12 education. *Education and Information Technologies*, 31(1), 191-212. https://doi.org/10.1007/s10639-025-13762-8

Bernal Parraga, A. P., Coronel Ramírez, E. A., Aldas Macias, K. J., Carvajal Madrid, C. A., Valarezo Espinoza, B. D. C., Vera Alcivar, J. G., & Chávez Cedeño, J. U. (2025). The impact of artificial intelligence on personalized learning in English language education. *Ciencia Latina Revista Científica Multidisciplinar*, 9(1), 5500–5518. https://doi.org/10.37811/cl_rcm.v9i1.16234

Bognár, L., & Khine, M. S. (2025). The shifting landscape of student engagement: A pre-post semester analysis in AI-enhanced classrooms. Computers and Education: Artificial Intelligence, 8, Article 100395. https://doi.org/10.1016/j.caeai.2025.100395

Bolender, B., Vispoel, S., Converse, G., Koprowicz, N., Song, D., & Osaro, S. (2024). Generative AI in K12: Analytics from early adoption. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*. https://doi.org/10.21031/epod.1539710

Bonner, E., Lege, R., & Frazier, E. (2023). Large language model–based artificial intelligence in the language classroom: Practical ideas for teaching. Teaching English with Technology, 23(1), 23–41. https://doi.org/10.56297/BKAM1691/WIEO1749

Borg, S. (2005). Teacher cognition in language teaching. In *Expertise in second language learning and teaching* (pp. 190-209). London: Palgrave Macmillan UK.

Boulhrir, T., & Hamash, M. (2025). Unpacking artificial intelligence in elementary education: A comprehensive thematic analysis systematic review. Computers and Education: Artificial Intelligence, 9, Article 100442. https://doi.org/10.1016/j.caeai.2025.100442

Brohi, S., & Jhanjhi, N. Z. (2026). Agentic AI in higher education: A survey of educators' perceptions of applications, challenges, governance and adoption readiness. *SSRN*. https://doi.org/10.2139/ssrn.7230198

Brohi, S., Mastoi, Q., Jhanjhi, N. Z., & Pillai, T. R. (2025). A research landscape of agentic AI and large language models: Applications, challenges and future directions. *Algorithms*, 18(8), Article 499. https://doi.org/10.3390/a18080499

Budiyanto, D., Tridinanti, G., Manurung, J. E., & Setiawan, H. (2026). Blending human and AI in translation teaching for 21st-century EFL classroom. *JEE (Journal of English Education)*, 12(1), 365–387. https://doi.org/10.30606/jee.v12i1.4468

Bulut, M. A. (2026). "Aidemics" examined analytically through the connectivist theory: Institutional AI-mediation in higher education via an artificial intelligence center (CILT-AI). *Interactive Learning Environments*, *34*(2), 972–1001. https://doi.org/10.1080/10494820.2026.2647982

Bulut, R. (2026). Generative artificial intelligence in teacher education: Examining the role of AI self-efficacy and digital literacy in technology acceptance. *Research Square*. https://doi.org/10.21203/rs.3.rs-9996155/v1

Cabero-Almenara, J., Palacios-Rodríguez, A., Loaiza-Aguirre, M. I., & Rivas-Manzano, M. D. R. D. (2024). Acceptance of educational artificial intelligence by teachers and its relationship with some variables and pedagogical beliefs. *Education Sciences*, 14(7), 740. https://doi.org/10.3390/educsci14070740

Caccavale, F., Gargalo, C. L., Kager, J., & Krühne, U. (2025). ChatGMP: A case of AI chatbots in chemical engineering education towards the automation of repetitive tasks. Computers and Education: Artificial Intelligence, 8, Article 100354. https://doi.org/10.1016/j.caeai.2024.100354

Can, T., (2020). Training Pre-Service English Language Teachers with 3-D Machinima. *The Turkish Online Journal of Educational Technology - TOJET* , vol.19, no.1, 53-65.

Can, T., & Mangır, C. (2024). Empowering language educators: Integrating artificial intelligence in classrooms and beyond. Y. Özkan, H. Tuncer, & E. Bada (Eds.), *An approach to digitalization in language teaching from different perspectives: A reflection of empirical work* (pp. 45–60). NOVA Science Publishers.

Can, T., & Karacan, C. G (2025). Assessing teacher digital competences in pre-service language teachers: Adaptation and validation of the DigCompEdu scale. Journal of Research on Technology in Education, 58(5), 1114-1133. https://doi.org/10.1080/15391523.2025.2504354

Can, T., & Rey, A (2022). Training language teachers for English-medium instruction (EMI) contexts through the use of augmented reality. Springer (edited-volume chapter). https://doi.org/10.1007/978-3-030-88597-7_10

Can, T., & Simsek, I (2015). The use of 3D virtual learning environments in training foreign language pre-service teachers. Turkish Online Journal of Distance Education, 16(4). https://doi.org/10.17718/tojde.53012

Can, T., & Simsek, I (2020). Using tablets for technology integration in classroom differentiation. IntechOpen (book chapter, The Role of Technology in Education). https://doi.org/10.5772/intechopen.85713

Can, T., Elia, A., & Şimşek, İ. (2017). Dil eğitimini düşünmek: İstanbul Üniversitesi'nde sürükleyici oyunla kelime öğretimi [Immersive game-based vocabulary teaching at Istanbul University]. *RumeliDE Dil ve Edebiyat Araştırmaları Dergisi*, (10), 40–49. https://doi.org/10.29000/rumelide.360626

Can, T., Simsek, I., & Kucuk, S (2025). Teachers experience as designers and practitioners of differentiation activities using mobile devices across grades and subjects. Culture and Education, 37(4), 1099-1132. https://doi.org/10.1177/11356405251393048

Cao, X., Xu, S., Chen, X., Song, Y., Luo, Q., & He, T. (2026). Leveraging generative AI agent to promote teaching reflection in a K–12 AI course: Effects on teachers' reflection self-efficacy, instructional design, and reflective thinking. *IEEE Transactions on Learning Technologies*, 19, 127–140. https://doi.org/10.1109/tlt.2026.3668051

Carless, D. (2013). Trust and its role in facilitating dialogic feedback. In D. Boud & E. Molloy (Eds.), *Feedback in higher and professional education: Understanding it and doing it well* (pp. 90–103). Routledge.

Cengiz, B. C., Bilki, Z., Ataş, A. H., & Celik, B. (2025). Exploring second language writers' engagement with ChatGPT feedback: Revision behaviors and perceptions. *System*, 134, 103837. https://doi.org/10.1016/j.system.2025.103837

Cerratto Pargman, T., McGrath, C., & Milrad, M. (2025). Towards responsible AI in education: Challenges and implications for research and practice. Computers and Education: Artificial Intelligence, 9, Article 100345. https://doi.org/10.1016/j.caeai.2024.100345

Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. International Journal of Educational Technology in Higher Education, 20(1), Article 38. https://doi.org/10.1186/s41239-023-00408-3

Chan, C. K. Y., & Lee, K. K. W. (2023). The AI generation gap: Are Gen Z students more interested in adopting generative AI such as ChatGPT in teaching and learning than their Gen X and millennial generation teachers? Smart Learning Environments, 10(1), Article 29. https://doi.org/10.1186/s40561-023-00269-3

Chan, K. L. R. (2024). The construct of English native speaker in Hong Kong. Journal of English and Applied Linguistics, 3(1), 1–7. https://doi.org/10.59588/2961-3094.1074

Chan, K. L. R. (2025). Exploring generative artificial intelligence-aided language learning tools in language classrooms: Insights from English language teachers in Hong Kong. Technology in Language Teaching & Learning, 7(2). https://doi.org/10.29140/tltl.v7n2.103217

Chan, K. W., Ali, F., Park, J., Sham, K. S. B., Tan, E. Y. T., Chong, F. W. C., Qian, K., & Sze, G. K. (2025). Automatic item generation in various STEM subjects using large language

model prompting. *Computers and Education: Artificial Intelligence*, 8, Article 100344. https://doi.org/10.1016/j.caeai.2024.100344

Chang, M. A., Tissenbaum, M., Philip, T. M., & D'Mello, S (2025). Co-designing AI with youth partners: Enabling ideal classroom relationships through a novel AI relational privacy ethical framework. Computers and Education: Artificial Intelligence, 8, Article 100364. https://doi.org/10.1016/j.caeai.2025.100364

Chang, M. A., Wong, R. Y., Breideband, T., Philip, T. M., McKoy, A., Cortez, A., & D'Mello, S. K. (2024). Co-design partners as transformative learners: Imagining ideal technology for schools by centering speculative relationships. In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (pp. 1–15). https://doi.org/10.1145/3613904.3642559

Chapelle, C. A (2001). Computer applications in second language acquisition: Foundations for teaching, testing and research. Cambridge University Press. https://doi.org/10.1017/CBO9781139524681

Chapelle, C. A., & Douglas, D (2006). Assessing language through computer technology. Cambridge University Press. https://doi.org/10.1017/CBO9780511733116

Che, C., & Perkins, B. (2026). Academic stewardship through AI-enabled learning governance: Ethical systems for inclusive learner support in higher education. *Research Square*. https://doi.org/10.21203/rs.3.rs-9759894/v1

Cheah, Y. H., Lu, J., & Kim, J. (2025). Integrating generative artificial intelligence in K–12 education: Examining teachers' preparedness, practices, and barriers. Computers and Education: Artificial Intelligence, 8, Article 100363. https://doi.org/10.1016/j.caeai.2025.100363

Chen, M., Miao, Z., & Ahmad, N. (2026). Managing behavior in AI-mediated English education: The interplay of teachers, technology, and communication willingness. *Acta Psychologica*, 265, 106765. https://doi.org/10.1016/j.actpsy.2026.106765

Chen, X., & Hu, H. (2026). AI-smart classroom for English translation: An adaptive HMM-based framework. *Computational Intelligence*, *42*(1), Article e70170. https://doi.org/10.1111/coin.70170

Chen, Y. (2024). Effects of technology-enhanced language learning on reducing EFL learners' public speaking anxiety. Computer Assisted Language Learning, 37(4), 789–813. https://doi.org/10.1080/09588221.2022.2055083

Chen, Y. (2025). Evaluating the potential of ChatGPT-reformulated essays as written feedback in L2 writing. Computers and Education: Artificial Intelligence, 9, Article 100500. https://doi.org/10.1016/j.caeai.2025.100500

Chen, Y., Zhi, Y., & Derakhshan, A. (2025). Integrating artificial intelligence (AI) into the English as a foreign language classroom: Exploring its impact on students' achievement emotions and willingness to communicate (WTC). *European Journal of Education*, 60(3). https://doi.org/10.1111/ejed.70157

Cheng, E. C. K., & Wang, T. (2023). Leading digital transformation and eliminating barriers for teachers to incorporate artificial intelligence in basic education in Hong Kong. Computers and Education: Artificial Intelligence, 5, Article 100171. https://doi.org/10.1016/j.caeai.2023.100171

Cheng, J., Zhao D'Antilio, Y., Chen, X., & Bernstein, J. (2014). Automatic assessment of the speech of young English learners. In Proceedings of the Ninth Workshop on Innovative Use

of NLP for Building Educational Applications (pp. 12–21). Association for Computational Linguistics. https://aclanthology.org/W14-1802. https://doi.org/10.3115/v1/W14-1802

Cheng, L. (2025). AI- and big data-driven innovation in vocational education: A case study of "Belt and Road" language service talent development. *Research Square*. https://doi.org/10.21203/rs.3.rs-7901067/v1

Cheng, Y., Fan, Y., Li, X., Chen, G., Gašević, D., & Swiecki, Z (2025). Asking generative artificial intelligence the right questions improves writing performance. Computers and Education: Artificial Intelligence, 8, Article 100374. https://doi.org/10.1016/j.caeai.2025.100374

Chetveryk, V. (2026). The role of generative artificial intelligence in foreign language learning during educational transformation. NGO Scientific and Educational Innovation Center of Social Transformations. https://doi.org/10.54929/reicst-conf-2026-02-25-03-06

Choi, K. Y., Wu, C., & Moorhouse, B. L. (2025). Exploring the use of generative artificial intelligence in English language teaching: Voices from in-service teachers at an early-adopting Hong Kong secondary school. Technology in Language Teaching & Learning, 7(3), Article 102516. https://doi.org/10.29140/tltl.v7n3.102516

Choi, Y., Lee, D., & Baek, J. (2025). Generative AI agents in language learning: A randomized field experiment. In Proceedings of the 58th Hawaii International Conference on System Sciences. HICSS. https://doi.org/10.24251/hicss.2025.596

Chon, Y. V., & Shin, D. (2026). AI chatbot as a revision aid in second language writing: From error correction to lexical sophistication. *Journal of Computer Assisted Learning*, 42(3). https://doi.org/10.1002/jcal.70259

Chow, W. W.-Y., Peng, S., & Atiq, A. (2026). "AI enhanced my critical thinking". *Pacific Journal of Technology Enhanced Learning*, 8(1), 13-14. https://doi.org/10.24135/pjtel.v8i1.246

Cobbold, C., & Wright, L. (2021). Use of formative feedback to enhance summative performance. Anatolian Journal of Education, 6(1), 109–116. https://doi.org/10.29333/aje.2021.619a

Cohn, C., Guo, S., & Rayala, S. (2026). Evidence-decision-feedback: Theory-driven adaptive scaffolding for LLM agents. *arXiv*. https://doi.org/10.48550/arxiv.2602.01415

Coman, C., Gherheș, V., Bucs, A., Coman, E., Cioca, V.-R., Bódi, D.-C., & Șerbănescu, R.-M. (2026). Navigating opportunities and challenges of generative AI in higher education: Towards responsible, equitable, and human-centered integration. *Frontiers in Artificial Intelligence*, 9, 1750978. https://doi.org/10.3389/frai.2026.1750978

Consoli, T., & Petko, D. (2025). Which educational approaches predict students' generative AI confidence and responsibility? Computers and Education: Artificial Intelligence, 9, Article 100431. https://doi.org/10.1016/j.caeai.2025.100431

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. Innovations in Education and Teaching International, 61(2), 228–239. https://doi.org/10.1080/14703297.2023.2190148

Council of Europe. (2018). Common European framework of reference for languages: Learning, teaching, assessment. Companion volume with new descriptors. Council of Europe. https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989

Cowie, N., & Alizadeh, M. (2025). Virtual reality for language learning. *ELT Journal*, 80(3), 388-396. https://doi.org/10.1093/elt/ccaf025

Crompton, H., Edmett, A., Ichaporia, N., & Burke, D. (2024). AI and English language teaching: Affordances and challenges. British Journal of Educational Technology, 55(6), 2503–2529. https://doi.org/10.1111/bjet.13460

Cukurova, M. (2024). The interplay of learning, analytics, and artificial intelligence in education: A vision for hybrid intelligence (arXiv:2403.16081). arXiv. https://doi.org/10.48550/arXiv.2403.16081

Cukurova, M., Suraworachet, W., & Zhou, Q. (2025). Towards synergistic teacher-AI interactions with generative artificial intelligence. *arXiv*. https://doi.org/10.48550/arxiv.2511.19580

Czerkawski, B. C. (2024). Designing language learning experiences with generative AI tools. In F. Pan (Ed.), AI in language teaching, learning, and assessment (pp. 324–341). IGI Global. https://doi.org/10.4018/979-8-3693-0872-1.ch015

da Silva, G. C., Rodrigues, R. L., Amorim, A. N., Jeon, L., Albuquerque, E. X. S., Silva, V. C., da Silva, V. F., Pinheiro, A. L. A., Nunes, J. P. J. R., de Souza, S. X. M. G., Silva, M. S., Mauro, I., & Maciel, A. M. A. (2025). Assessing reading fluency in elementary grades: A machine learning approach. Computers and Education: Artificial Intelligence, 8, Article 100411. https://doi.org/10.1016/j.caeai.2025.100411

Dai, W., Cheng, Y., Aldino, A. A., Tsai, Y.-S., Gašević, D., & Chen, G. (2025). Evaluating the capability of large language models in characterising relational feedback: A comparative analysis of prompting strategies. Computers and Education: Artificial Intelligence, 8, Article 100427. https://doi.org/10.1016/j.caeai.2025.100427

Dakakni, D., & Safa, N. (2023). Artificial intelligence in the L2 classroom: Implications and challenges on ethics and equity in higher education: A 21st century Pandora's box. Computers and Education: Artificial Intelligence, 5, Article 100179. https://doi.org/10.1016/j.caeai.2023.100179

Dandan, J., Abdul Rahman, S. B., & Jaafar, J. R. (2026). AI literacy and self-regulated learning in the agentic AI era: A conceptual review of EFL learning. *Journal of Nusantara Studies (JONUS)*, 11(2), 326–342. https://doi.org/10.24200/jonus.vol11iss2pp326-342

De Marez, V., Van Nooten, J., De Bruyne, L., et al. (2025). An agentic AI framework for training general practitioner student skills. arXiv. https://doi.org/10.48550/arXiv.2512.18440

De Vrindt, M., Tack, A., Van den Noortgate, W., Lesterhuis, M., & Bouwer, R. (2025). Opportunities of natural language processing for comparative judgment assessment of essays. Computers and Education: Artificial Intelligence, 8, Article 100414. https://doi.org/10.1016/j.caeai.2025.100414

de Winter, J. C. F. (2024). Can ChatGPT pass high school exams on English language comprehension? International Journal of Artificial Intelligence in Education, 34, 915–930. https://doi.org/10.1007/s40593-023-00372-z

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. Psychological Inquiry, 11(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01

DeepSeek-AI. (2024). *DeepSeek LLM: Scaling open-source language models with longtermism*. arXiv. https://doi.org/10.48550/arXiv.2401.02954

Deepshikha, D. (2025). A comprehensive review of AI-powered grading and tailored feedback in universities. *Discover Artificial Intelligence*, 5(1). https://doi.org/10.1007/s44163-025-00517-0

DeKeyser, R. M. (Ed.) (2007). Practice in a second language: Perspectives from applied linguistics and cognitive psychology.

Desaire, H., Chua, A. E., Isom, M., Jarosova, R., & Hua, D (2023). Distinguishing academic science writing from humans or ChatGPT with over 99% accuracy using off-the-shelf machine learning tools. Cell Reports Physical Science, 4(6), 101426. https://doi.org/10.1016/j.xcrp.2023.101426

Dicken, S., DeLeon, N., Gutierrez, E. J., & Trujillo, J. M. (2026). Evaluating the quality of artificial intelligence–generated rubrics and grading reliability in student-written reflection assignments. *American Journal of Pharmaceutical Education*, 90(8), 102014. https://doi.org/10.1016/j.ajpe.2026.102014

Dinh, T. B. H. (2025). Factors affecting ICT integration of EFL teachers at a Vietnamese university. International Journal of Artificial Intelligence in Language Education, 2(1), 55–73. https://doi.org/10.54855/ijaile.25214

Divya, T. J., & Alamelu, C. (2026). Integrating DeepL Write and Claude AI to enhance argumentative writing competence of engineering students in India. *Frontiers in Artificial Intelligence*, 9. https://doi.org/10.3389/frai.2026.1838463

Domínguez Figaredo, D., & Stoyanovich, J. (2023). Responsible AI literacy: A stakeholder-first approach. *Big Data & Society*, 10(2). https://doi.org/10.1177/20539517231219958

Dong, L. (2025). Development and validation of a multidimensional ChatGPT feedback engagement scale in second language writing. *System*, 135, 103875. https://doi.org/10.1016/j.system.2025.103875

Dong, L., Tang, X., & Wang, X. (2025). Examining the effect of artificial intelligence in relation to students' academic achievement: A meta-analysis. Computers and Education: Artificial Intelligence, 8, Article 100400. https://doi.org/10.1016/j.caeai.2025.100400

Dong, X., Wang, Z., & Han, S. (2025). Mitigating learning burnout caused by generative artificial intelligence misuse in higher education: A case study in programming language teaching. Informatics, 12(2), 51. https://doi.org/10.3390/informatics12020051

Du, Q. (2025). How artificially intelligent conversational agents influence EFL learners' self-regulated learning and retention. *Education and Information Technologies*, 30(15), 21635–21701. https://doi.org/10.1007/s10639-025-13602-9

Dwivedi, Y. K., Helal, M. Y., & Elgendy, I. A. (2025). Agentic AI systems: What it is and isn't. *Global Business and Organizational Excellence*, 45(3), 253-263. https://doi.org/10.1002/joe.70018

Edmett, A., Ichaporia, N., Crompton, H., & Crichton, R. (2024). Artificial intelligence and English language teaching: Preparing for the future (2nd ed.). British Council. https://doi.org/10.57884/78EA-3C69

Eguchi, M., Takizawa, K., & Saeki, M. (2025). Human- versus artificial intelligence-delivered roleplay tasks for assessing interactional competence: An applied conversation analytic study. *TESOL Quarterly*, 59(S1). https://doi.org/10.1002/tesq.70028

Elmotri, B., Abd Elwahab Abd Elmajid Osman, W., Ahmed Al-humari, M., Amri, F., M. Elyasa, Y., Hussain Malik, F., Garrouri, S., & H. S. Hassanein, O. (2025). Possibilities for improving ESP curriculum design and assessment strategies for Saudi universities with ChatGPT. *Arab World English Journal*, (1), 234–250. https://doi.org/10.24093/awej/ai.13

Erdős, F. G., Németh, R., Afrifa, S., et al. (2025). Comparative analysis of NLP-driven MCQ generators from text sources. Computers and Education: Artificial Intelligence, 9, Article 100440. https://doi.org/10.1016/j.caeai.2025.100440

Ericsson, E., & Johansson, S. (2023). English speaking practice with conversational AI: Lower secondary students' educational experiences over time. *Computers and Education: Artificial Intelligence*, 5, 100164. https://doi.org/10.1016/j.caeai.2023.100164

European Commission. (2022). Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for educators. Publications Office of the European Union. https://op.europa.eu/en/publication-detail/-/publication/d81a0d54-5348-11ed-92ed-01aa75ed71a1/language-en

European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence. http://data.europa.eu/eli/reg/2024/1689/oj/eng

Evangeline, S. I. (2025). Regulatory frameworks for agentic AI in education. In Agentic AI in education (pp. 179–210). IGI Global. https://doi.org/10.4018/979-8-3373-4621-2.ch009

Faiz, H., & Hazilan, W. F (2024). ChatGPT in English as a second language education: A review. Global Journal of Educational Thoughts, 34-40. https://gsf.info/wp-content/uploads/2024/05/ChatGPT-in-English-as-A-Second-Language-Education-A-Review.pdf

Falcon, G., & Carillo, K. (2026). Responsible AI by design: Embedding diversity, equity and inclusion values into AI development practices. *Behaviour and Information Technology*, 1–16. https://doi.org/10.1080/0144929x.2026.2657460

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. British Journal of Educational Technology, 56(2), 489–530. https://doi.org/10.1111/bjet.13544

Faruna, G. A. (2026). LSR: Linguistic safety robustness benchmark for low-resource West African languages. *arXiv*. https://doi.org/10.48550/arxiv.2603.19273

Feng, L. (2025). Investigating the effects of artificial intelligence-assisted language learning strategies on cognitive load and learning outcomes: A comparative study. *Journal of Educational Computing Research*, 62(8), 1741–1774. https://doi.org/10.1177/07356331241268349

Feng, S., Zhang, H., & Gašević, D. (2025). Mapping the evolution of AI in education: Toward a co-adaptive and human-centered paradigm. Computers and Education: Artificial Intelligence, 9, Article 100513. https://doi.org/10.1016/j.caeai.2025.100513

Fernandes, S., Araújo, A. M., Rodrigues, P., et al. (2025). The use of generative AI in academic work: Insights from university students and faculty. PAEE. https://doi.org/10.5281/zenodo.15880821

Fernandez-Espinosa, M., Zhang, K., & Bendarkawi, J. (2025). Practicing a second language without fear: Mixed reality agents for interactive group conversation. *arXiv*. https://doi.org/10.48550/arxiv.2510.08227

Fitria, T. N (2021). The use technology based on artificial intelligence in English teaching and learning. ELT Echo: The Journal of English Language Teaching in Foreign Language Context. https://doi.org/10.24235/eltecho.v6i2.9299

Fleckenstein, J., Liebenow, L. W., & Meyer, J. (2023). Automated feedback and writing: A multi-level meta-analysis of effects on students' performance. Frontiers in Artificial Intelligence, 6, Article 1162454. https://doi.org/10.3389/frai.2023.1162454

Floridi, L., & Cowls, J. (2022). A unified framework of five principles for AI in society. In Machine learning and the city (pp. 535–545). John Wiley & Sons.

Franco, A. (2026). The attribution-efficacy paradox: A scoping review of GenAI interaction and academic self-efficacy in higher education. *ResearchSquare*. https://doi.org/10.21203/rs.3.rs-10362021/v1

Gao, H. (2025). Optimization method for academic English content based on generative adversarial networks and data augmentation. Computers and Education: Artificial Intelligence, 9, Article 100492. https://doi.org/10.1016/j.caeai.2025.100492

Gao, J., & Yang, W. (2026). Faculty adoption of AI-assisted teaching tools in Chinese higher education: An integrated UTAUT2-TAM framework at Shanghai University. *Frontiers in Artificial Intelligence*, 9. https://doi.org/10.3389/frai.2026.1847439

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, H., & Wang, H. (2023). Retrieval-augmented generation for large language models: A survey (arXiv:2312.10997). arXiv. https://doi.org/10.48550/arXiv.2312.10997

Garg, A., Nisumba Soodhani, K., & Rajendran, R. (2025). Enhancing data analysis and programming skills through structured prompt training: The impact of generative AI in engineering education. Computers and Education: Artificial Intelligence. https://doi.org/10.1016/j.caeai.2025.100380

Gayed, J. M., Carlon, M. K. J., Oriola, A. M., & Cross, J. S (2022). Exploring an AI-based writing Assistant's impact on English language learners. Computers and Education: Artificial Intelligence, 3, 100055. https://doi.org/10.1016/j.caeai.2022.100055

Ghiasvand, F., & Kogani, M. (2024). "I'm not ready for this metamorphosis": An ecological approach to Iranian and Italian EFL teachers' readiness for artificial intelligence-mediated instruction. *Teaching English with Technology*, *2024*(3). https://doi.org/10.56297/vaca6841//bffo7057/iski2001

Ghiasvand, F., & Seyri, H. (2025). A collaborative reflection on the synergy of artificial intelligence (AI) and language teacher identity reconstruction. *Teaching and Teacher Education*, 160, 105022. https://doi.org/10.1016/j.tate.2025.105022

Giannakos, M., & Sharma, K. (2020). Multimodal data capabilities for learning: What can multimodal data tell us about learning? British Journal of Educational Technology, 51(5), 1450–1484. https://doi.org/10.1111/bjet.12993

Giannakos, M., Spikol, D., Di Mitri, D., Sharma, K., Ochoa, X., & Hammad, R. (2022). Introduction to multimodal learning analytics. In The multimodal learning analytics handbook (pp. 3–28). Springer. https://doi.org/10.1007/978-3-031-08076-0_1

Giannakos, M. N., & Cukurova, M. (2023). The role of learning theory in multimodal learning analytics. *British Journal of Educational Technology*, 54(5), 1246-1267. https://doi.org/10.1111/bjet.13320

Gligorea, I., Cioca, M., Oancea, R., Gorski, A.-T., Gorski, H., & Tudorache, P (2023). Adaptive learning using artificial intelligence in e-learning: A literature review. Education Sciences, 13(12), 1216. https://doi.org/10.3390/educsci13121216

Godwin-Jones, R (2024). Distributed agency in second language learning and teaching through generative AI. Language Learning & Technology, 5-31. https://doi.org/10.64152/10125/73570

Goellner, S., Tropmann-Frick, M., & Brumen, B. (2024). Responsible artificial intelligence: A structured literature review. arXiv. https://arxiv.org/abs/2403.06910. https://doi.org/10.48550/arXiv.2403.06910

Goh, C. C. M., & Aryadoust, V. (2025). Developing and assessing second language listening and speaking: Does AI make it better? *Annual Review of Applied Linguistics*, 45, 179–199. https://doi.org/10.1017/s0267190525100111

Grab, M. Ö. (2025). Integrated AI chatbot practice: A pathway to improved ESL speaking skills. *Social Sciences & Humanities Open*, 12, 101933. https://doi.org/10.1016/j.ssaho.2025.101933

Graesser, A. C., Chipman, P., Haynes, B. C., & Olney, A (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. IEEE Transactions on Education, 48(4), 612-618. https://doi.org/10.1109/TE.2005.856149

Guan, L., Lee, J. C. K., Zhang, Y., & Gu, M. M. (2025). Investigating the tripartite interaction among teachers, students, and generative AI in EFL education: A mixed-methods study. Computers and Education: Artificial Intelligence, 8, Article 100384. https://doi.org/10.1016/j.caeai.2025.100384

Guan, L., Zhang, Y., & Gu, M. M. (2025). Pre-service teachers preparedness for AI-integrated education: An investigation from perceptions, capabilities, and teachers' identity changes. Computers and Education: Artificial Intelligence, 8, Article 100341. https://doi.org/10.1016/j.caeai.2024.100341

Guibangguibang, H. R. B., & Hesula, K. C. (2026). Can AI replace teachers? Comparing AI and human teachers' ratings and feedback on students' essays. *Research Square*. https://doi.org/10.21203/rs.3.rs-10135718/v1

Gunasekara, L., El-Haber, N., Nagpal, S., Moraliyage, H., Issadeen, Z., Manic, M., & De Silva, D. (2025). A systematic review of responsible artificial intelligence principles and practice. *Applied System Innovation*, 8(4), 97. https://doi.org/10.3390/asi8040097

Güneyli, A., Burgul, N. S., Dericioğlu, S., Cenkova, N., Becan, S., Şimşek, Ş. E., & Güneralp, H. (2024). Exploring teacher awareness of artificial intelligence in education: A case study from Northern Cyprus. *European Journal of Investigation in Health, Psychology and Education*, 14(8), 2358–2376. https://doi.org/10.3390/ejihpe14080156

Guo, H., & Erdenebold, T. (2025). Factors influencing intention to adopt an AI chatbot for learning in higher education: An integrated PLS-SEM, IPMA, and ANN approach. Computers and Education: Artificial Intelligence, 9, Article 100477. https://doi.org/10.1016/j.caeai.2025.100477

Gupta, N., & Heggond, S. (2026). Agentic artificial intelligence-driven tutoring: A multi-agent cognitive architecture for personalized adaptive learning in education. *International Journal of Applied Resilience and Sustainability*, *2*(2), 572–598. https://doi.org/10.70593/deep-sci.0202022

Gyekye, E. K. (2025). From users to designers; high school students as co-designers and critical agents in ChatGPT-integrated learning. *Research Square*. https://doi.org/10.21203/rs.3.rs-7151671/v1

Haidar, R., & Tassis, A. (2025). Power, pedagogy, and algorithms: An exploratory study of EAP instructors' critical engagement with generative AI. *Research Square*. https://doi.org/10.21203/rs.3.rs-7014667/v1

Hammami, S., & Alomair, Y. (2026). Enhancing language learning through adaptive gamification and a multi-agent approach with experience application programming interface and blockchain integration. *PeerJ Computer Science*, 12, Article e3508. https://doi.org/10.7717/peerj-cs.3508

Hao, T., Zhao, X., & Lu, Z (2025). Spontaneous use of generative artificial intelligence in preparing for College English Test: GenAI adoption factors, learners' motivation, and self-efficacy. Technology in Language Teaching & Learning. https://doi.org/10.29140/tltl.v7n3.102640

Heung, Y. M. E., & Chiu, T. K. F. (2025). How ChatGPT impacts student engagement from a systematic review and meta-analysis study. Computers and Education: Artificial Intelligence, 8, Article 100361. https://doi.org/10.1016/j.caeai.2025.100361

Heydarnejad, T. (2025). Unmasking the impacts of self-evaluation in AI-supported writing instruction on EFL learners' emotion regulation, self-competence, motivation, and writing achievement. Computers and Education: Artificial Intelligence, 9, Article 100494. https://doi.org/10.1016/j.caeai.2025.100494

Ho, C. S. M., & Lee, J. C.-K. (2025). From intuition to action: Exploring teachers' ethical awareness in the use of AI tools in education. Computers and Education: Artificial Intelligence, 9, Article 100502. https://doi.org/10.1016/j.caeai.2025.100502

Ho, M.-T., Mantello, P., & Vuong, Q.-H. (2024). Emotional AI in education and toys: Investigating moral risk awareness in the acceptance of AI technologies from a cross-sectional survey of the Japanese population. *Heliyon*, 10(16), e36251. https://doi.org/10.1016/j.heliyon.2024.e36251

Hofmann, V., Kalluri, P. R., Jurafsky, D., & King, S (2024). AI generates covertly racist decisions about people based on their dialect. Nature, 633(8028), 147-154. https://doi.org/10.1038/s41586-024-07856-5

Holec, H. (1981). Autonomy and foreign language learning. Pergamon.

Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. European Journal of Education, 57(4), 542–570. https://doi.org/10.1111/ejed.12533

Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., et al. (2022). Ethics of AI in education: Towards a community-wide framework. International Journal of Artificial Intelligence in Education, 32(3), 504–526. https://doi.org/10.1007/s40593-021-00239-1

Hooshyar, D., Šír, G., Yang, Y., Kikas, E., Hämäläinen, R., Kärkkäinen, T., Gašević, D., & Azevedo, R. (2025). Towards responsible AI for education: Hybrid human-AI to confront the elephant in the room. Computers & Education: Artificial Intelligence, 9, Article 100524. https://doi.org/10.1016/j.caeai.2025.100524

Horbach, A., Pehlke, J., Laarmann-Quante, R., & Ding, Y. (2024). Crosslingual content scoring in five languages using machine-translation and multilingual transformer models. International Journal of Artificial Intelligence in Education, 34(4), 1294-1320. https://doi.org/10.1007/s40593-023-00370-1

Hossain, S., & Nawmi, T. A. (2026). Measuring how students rely on generative AI in academic writing: Development and multi-source validation of the Generative AI Reliance Types Scale (GenAI-RTS). *arXiv*. https://doi.org/10.48550/arxiv.2607.14301

Huang, G., & Chen, H. (2026). Beyond grammar: How ChatGPT-4.0 shapes writing self-efficacy and performance in CSL writing. *Language Learning & Technology*, 30(1), 1-23. https://doi.org/10.64152/10125/73706

Huang, W., Abdul Aziz, A., & Kutty, F. M. (2025). Integrating AI into Asian tertiary EFL learners' speaking instruction: A systematic literature review. Forum for Linguistic Studies, 7(3), 104–119. https://doi.org/10.30564/fls.v7i3.8263

Huang, X., Zou, D., Cheng, G., Chen, X., & Xie, H. (2023). Trends, research issues and applications of artificial intelligence in language education. Educational Technology & Society, 26(1), 112–131. https://doi.org/10.30191/ETS.202301_26(1).0009

Huang, Z., Nazar, N., Abbas, G., & Jiang, Z. (2025). Assessing AI literacy in second language writing: A scale development and validation study. *Education and Information Technologies*, 30(18), 25909–25935. https://doi.org/10.1007/s10639-025-13767-3

Hubbard, P (2008). CALL and the future of language teacher education. CALICO Journal, 25(2), 175-188. https://doi.org/10.1558/cj.v25i2.175-188

Hubbard, P., & Schulze, M. (2025). AI and the future of language teaching: Motivating sustained integrated professional development. *International Journal of Computer-Assisted Language Learning and Teaching*, 15(1), 1-17. https://doi.org/10.4018/IJCALLT.378304

Huertas-Abril, C. A., & Palacios-Hidalgo, F. J. (2023). New possibilities of artificial intelligence-assisted language learning (AIALL): Comparing visions from the East and the West. *Education Sciences*, 13(12), 1234. https://doi.org/10.3390/educsci13121234

Hyland, K. (2025). EFL teachers and feedback fatigue: AI to the rescue? *Language Teaching*, 59(3), 310–326. https://doi.org/10.1017/s0261444825101018

Ibrahim, H. S (2025). The robustness of AI-classifiers in the face of AI-assisted plagiarism: The case of Turnitin AI content detector. https://doi.org/10.4018/IJCALLT.375428

Iqbal, M., Singh, K., Khan, S., Osho, O., Sidnam-Mauch, E., Bannister, N., Caine, K., & Knijnenburg, B. P. (2025). Teaching AI awareness in middle school classrooms: Design, implementation and evaluation of two education modules on algorithmic bias and filter bubbles. Computers and Education: Artificial Intelligence, 8, Article 100425. https://doi.org/10.1016/j.caeai.2025.100425

Islomova, M. (2026). Artificial intelligence in language learning: A case study of ChatGPT. *Advances in Science and Education*, 2(3), 32-36. https://doi.org/10.70728/edu.v02.i03.008

Isop, W. A. (2025). A conceptual ethical framework to preserve natural human presence in the use of AI systems in education. *Frontiers in Artificial Intelligence*, 7. https://doi.org/10.3389/frai.2024.1377938

Jafarian, N. R., & Kramer, A.-W. (2025). AI-assisted audio-learning improves academic achievement through motivation and reading engagement. Computers and Education: Artificial Intelligence, 8, Article 100357. https://doi.org/10.1016/j.caeai.2024.100357

Jaleniauskienė, E., Lisaitė, D., & Daniusevičiūtė-Brazaitė, L. (2023). Artificial Intelligence in Language Education: A Bibliometric Analysis. Sustainable Multilingualism, 23(1), 159-194. https://doi.org/10.2478/sm-2023-0017

Jamshaid, M. M., Akbar, Z., & Hassaan, A. (2025). Preparing human oversight talent for agentic AI workplaces: A competency framework for education and workforce systems. *Contemporary Journal of Social Science Review*, 3(4), 1561-1574. https://doi.org/10.63878/cjssr.v3i4.1732

Jansone, A., & Cīrule, Z. A. (2026). Artificial intelligence-based pedagogical agent in an e-learning environment. *Computers*, 15(7), Article 401. https://doi.org/10.3390/computers15070401

Jantakoon, T., Jantakun, T., Jantakun, K., Pongpanich, W., Pasmala, R., Wannapiroon, P., & Nilsook, P. (2025). The effectiveness of artificial intelligence in English instruction for speaking and listening skills: A meta-analysis. *Contemporary Educational Technology*, 17(4), ep596. https://doi.org/10.30935/cedtech/17310

Jara Chiriboga, S. P., Troncoso Burgos, A. L., Ruiz Avila, M. M., Cosquillo Chida, J. L., Aldas Macias, K. J., Castro Morante, Y. E., & Bernal Párraga, A. P. (2025). Artificial intelligence and personalized learning in foreign languages: An analysis of chatbots and virtual assistants in education. *Revista Científica de Salud y Desarrollo Humano*, 6(1), 882–905. https://doi.org/10.61368/r.s.d.h.v6i1.515

Jiang, L., Lee, I., & Yu, S. (2024). Conceptualizing multimodal feedback literacy for L2 writing teachers in the digital age. *International Journal of Applied Linguistics*, 34(4), 1479–1496. https://doi.org/10.1111/ijal.12578

Jin, Y., Martinez-Maldonado, R., Gašević, D., & Yan, L. (2025). GLAT: The generative AI literacy assessment test. Computers and Education: Artificial Intelligence, 9, Article 100436. https://doi.org/10.1016/j.caeai.2025.100436

Jin, Y., Yan, L., Echeverría, V., Gašević, D., & Martinez-Maldonado, R. (2025). Generative AI in higher education: A global perspective of institutional adoption policies and guidelines. Computers and Education: Artificial Intelligence, 8, Article 100348. https://doi.org/10.1016/j.caeai.2024.100348

Jinowat, N., Wiboolyasarin, K., Chomchuen, F., & Wiboolyasarin, W. (2026). Trust, resistance, and transformation: A Q-methodological study of teachers' perspectives on AI-generated feedback in second language writing. *Foreign Language Annals*, 59(1), 34–62. https://doi.org/10.1111/flan.70042

Johnson, M. S., Choi, I., & McEachin, A. (2026). Ethical considerations and challenges in AI-based educational assessment. In *The Role of AI in Assessment* (pp. 23–52). Routledge. https://doi.org/10.4324/9781003621522-2

Jones-Jang, S. M., Chung, M., & Choi, J. (2025). Fairness perceptions of AI in grading systems: Examining how discontent with the status quo and outcome favorability reduce AI reluctance. Computers and Education: Artificial Intelligence. https://doi.org/10.1016/j.caeai.2025.100419

Junaidi, J., & Wahyono, T. (2025). AI-driven competency recommendations based on attendance patterns and academic performance. Computers and Education: Artificial Intelligence, 8, Article 100423. https://doi.org/10.1016/j.caeai.2025.100423

Jung, J. Y., Tyack, L., & von Davier, M. (2025). Towards the implementation of automated scoring in international large-scale assessments: Scalability and quality control. Computers and Education: Artificial Intelligence, 8, Article 100375. https://doi.org/10.1016/j.caeai.2025.100375

Kahila, J., Vartiainen, H., Tedre, M., Arkko, E., Lin, A., Pope, N., Jormanainen, I., & Valtonen, T. (2024). Pedagogical framework for cultivating children's data agency and creative

abilities in the age of AI. *Informatics in Education*, 323–360. https://doi.org/10.15388/infedu.2024.15

Kalisa, P., & Hadi, M. Z. P. (2026). Beyond writing assistance: Generative AI, multilingual meaning-making, and authorial agency in EMI critical writing classrooms. *Research Square*. https://doi.org/10.21203/rs.3.rs-10218437/v1

Kannan, J., & Munday, P. (2018). New trends in second language learning and teaching through the lens of ICT, networked learning, and artificial intelligence. Círculo de Lingüística Aplicada a la Comunicación, 76, 13–30. https://doi.org/10.5209/CLAC.62495

Kargaran, A., Imani, A., & Yvon, F. (2023). GlotLID: Language identification for low-resource languages. In *Findings of the Association for Computational Linguistics: EMNLP 2023* (pp. 6155-6218). https://doi.org/10.18653/v1/2023.findings-emnlp.410

Karim, S. A., Hamzah, A. Q. S., Anjani, N. M., et al. (2023). Promoting EFL students' speaking performance through ELSA Speak: An artificial intelligence in English language learning. JOLLT Journal of Languages and Language Teaching. https://doi.org/10.33394/jollt.v11i4.8958

Kartal, G., & Yeşilyurt, Y. E. (2024). A bibliometric analysis of artificial intelligence in L2 teaching and applied linguistics between 1995 and 2022. ReCALL, 36(3), 359–375. https://doi.org/10.1017/S0958344024000077

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeiffer, F., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kic-Drgas, J., & Kılıçkaya, F. (2024). Using artificial intelligence (AI) to create language exam questions: A case study. *XLinguae*, 17(1), 20–33. https://doi.org/10.18355/xl.2024.17.01.02

Kim, K., & Kwon, K (2023). Exploring the AI competencies of elementary school teachers in South Korea. Computers and Education: Artificial Intelligence. https://doi.org/10.1016/j.caeai.2023.100137

Kinder, A., Briese, F. J., Dern, N., et al. (2025). Effects of adaptive feedback generated by a large language model: A case study in teacher education. Computers and Education: Artificial Intelligence, 8, Article 100349. https://doi.org/10.1016/j.caeai.2024.100349

Kitsara, I. (2022). Artificial intelligence and the digital divide: From an innovation perspective. In A. Bounfour (Ed.), Platforms and artificial intelligence. Springer. https://doi.org/10.1007/978-3-030-90192-9_12

Kizilcec, R. F., & Lee, H (2022). Algorithmic fairness in education. In W. Holmes & K. Porayska-Pomsta (Eds.), The ethics of artificial intelligence in education: Practices, challenges, and debates (pp. 174-202). Routledge.

Kılıç, E., Almasri, F., & Çelik, H. E. (2025). Are pre-service teachers ready to teach the Alpha generation? The impact of pre-service teachers' ChatGPT literacy levels on behavioral intentions toward ChatGPT-4.0. Computers and Education: Artificial Intelligence, 9, Article 100486. https://doi.org/10.1016/j.caeai.2025.100486

Knoth, N., Tolzin, A., Janson, A., & Leimeister, J. M. (2024). AI literacy and its implications for prompt engineering strategies. Computers & Education: Artificial Intelligence, 6, Article 100225. https://doi.org/10.1016/j.caeai.2024.100225

Kohnke, L., & Green-Eneix, C. (2025). Preparing pre-service and in-service teachers for the AI-driven classroom (pp. 235–252). Routledge. https://doi.org/10.4324/9781032688602-15

Kohnke, L., Moorhouse, B. L., & Zou, D (2023). Exploring generative artificial intelligence preparedness among university language instructors: A case study. Computers and Education: Artificial Intelligence. https://doi.org/10.1016/j.caeai.2023.100156

Kohnke, L., Zou, D., & Ou, A. W. (2025). Preparing future educators for AI-enhanced classrooms: Insights into AI literacy and integration. Computers and Education: Artificial Intelligence. https://doi.org/10.1016/j.caeai.2025.100398

Kohnke, L., Zou, D., & Su, F. (2025). Exploring the potential of GenAI for personalised English teaching: Learners' experiences and perceptions. Computers and Education: Artificial Intelligence, 8, Article 100371. https://doi.org/10.1016/j.caeai.2025.100371

Kong, B., & He, C. (2025). Deep learning and fuzzy algorithm in improving the effectiveness of college English translation teaching. Computers and Education: Artificial Intelligence, 8, Article 100378. https://doi.org/10.1016/j.caeai.2025.100378

Kong, S. C., & Hu, W. (2025). A study of developing administrative staff's conceptual understanding of generative artificial intelligence through professional development: Evaluation of a course using tests, surveys and thematic analysis of reflective writings. Computers and Education: Artificial Intelligence, 9, Article 100444. https://doi.org/10.1016/j.caeai.2025.100444

Kong, S.-C., & Yang, Y. (2024). A human-centered learning and teaching framework using generative artificial intelligence for self-regulated learning development through domain knowledge learning in K-12 settings. IEEE Transactions on Learning Technologies, 17, 1562-1573. https://doi.org/10.1109/TLT.2024.3392830

Kong, S. C., Zhu, J., & Yang, Y. N. (2025). Developing and validating a scale of empowerment in using artificial intelligence for problem-solving for senior secondary and university students. Computers & Education: Artificial Intelligence. https://doi.org/10.1016/j.caeai.2024.100359

Kostopoulos, G., Gkamas, V., Rigou, M., & Kotsiantis, S. (2025). Agentic AI in education: State of the art and future directions. *IEEE Access*, *13*, 177467–177491. https://doi.org/10.1109/access.2025.3620473

Kovalenko, I., & Baranivska, N. (2024). Integrating artificial intelligence in English language teaching: Exploring the potential and challenges of AI tools in enhancing language learning outcomes and personalized education. European Socio-Legal and Humanitarian Studies. https://ehs-journal.ro/wp-content/uploads/2024/05/ESLHS-1_24-1.pdf

Krashen, S. D (1982). Principles and practice in second language acquisition. Pergamon Press.

Krashen, S. D (1985). The input hypothesis: Issues and implications. Longman.

Kundu, A., & Bej, T. (2025). Transforming EFL teaching with AI: A systematic review of empirical studies. *International Journal of Artificial Intelligence in Education*, 35(4), 2281-2314. https://doi.org/10.1007/s40593-025-00470-0

Lai, C. Y., Cheung, K. Y., Chan, C. S., & Law, K. K. (2024). Integrating the adapted UTAUT model with moral obligation, trust and perceived risk to predict ChatGPT adoption for

assessment support: A survey with students. Computers & Education: Artificial Intelligence, 6, Article 100246. https://doi.org/10.1016/j.caeai.2024.100246

Lai, Z. C. C. (2025). The impact of AI-assisted blended learning on writing efficacy and resilience. International Journal of Computer-Assisted Language Learning and Teaching, 15(1). https://doi.org/10.4018/IJCALLT.377174

Lampos, V., Mintz, J., & Qu, X. (2021). An artificial intelligence approach for selecting effective teacher communication strategies in autism education. NPJ Science of Learning, 6, Article 25. https://doi.org/10.1038/s41539-021-00102-x

Lantolf, J. P., & Thorne, S. L. (2006). Sociocultural theory and the genesis of second language development. Oxford University Press.

Latif, E., & Zhai, X. (2024). Fine-tuning ChatGPT for automatic scoring. Computers & Education: Artificial Intelligence, 6, Article 100210. https://doi.org/10.1016/j.caeai.2024.100210

Le, H., Shen, Y., & Tang, L. (2025). Breaking human dominance: Investigating learners' preferences for learning feedback from generative AI and human tutors. *British Journal of Educational Technology*, 56(5), 1758-1783. https://doi.org/10.1111/bjet.13614

Lee, H., Kim, H.-J., & Wei, Y. (2025). Promoting inclusive AI and technology in K-12 education: A review of context, instructional strategies, and learning outcomes. Computers and Education: Artificial Intelligence, 9, Article 100478. https://doi.org/10.1016/j.caeai.2025.100478

Lee, S., Jeon, J., & Choe, H (2024). Enhancing pre-service teachers' global Englishes awareness with technology: A focus on AI chatbots in 3D metaverse environments. https://doi.org/10.1002/tesq.3300

Lee, S., Jeon, J., & McKinley, J. (2025). Generative AI and English language teaching: A global Englishes perspective. *Annual Review of Applied Linguistics*, 45, 85-108. https://doi.org/10.1017/s0267190525100184

Lee, U., Jung, H., Jeon, Y., Sohn, Y., Hwang, W., Moon, J., & Kim, H. (2024). Few-shot is enough: Exploring ChatGPT prompt engineering method for automatic question generation in english education. *Education and Information Technologies*, 29(9), 11483–11515. https://doi.org/10.1007/s10639-023-12249-8

Levy, M., & Stockwell, G (2006). CALL dimensions: Options and issues in computer-assisted language learning. Lawrence Erlbaum Associates.

Li, C., Xing, W., Song, Y., et al. (2025). RICE AlgebraBot: Lessons learned from designing and developing responsible conversational AI using induction, concretization, and exemplification to support algebra learning. Computers and Education: Artificial Intelligence, 8, Article 100338. https://doi.org/10.1016/j.caeai.2024.100338

Li, H., Chai, C. S., & Wang, X. (2025). Promoting pre-service music teachers' TPACK with generative AI: An intervention through designing with AI. Computers and Education: Artificial Intelligence, 9, Article 100525. https://doi.org/10.1016/j.caeai.2025.100525

Li, H., Zhang, H., & Li, M. (2025). Toward robust multilingual adaptation of LLMs for low-resource languages. *arXiv*. https://doi.org/10.48550/arxiv.2510.14466

Li, J. (2025). Navigating ethical dilemmas in AI-enhanced education. *International Journal of Knowledge Management*, 21(1), 1–21. https://doi.org/10.4018/ijkm.382384

Li, T., & Ding, Y (2025). Exploring the GenAI literacy of Chinese university students in EFL learning. https://doi.org/10.4018/IJCALLT.377175

Li, X. (2025). Development and implementation of an intelligent assisted teaching system for Chinese English based on natural language processing and reinforcement learning. Computers and Education: Artificial Intelligence, 9, Article 100466. https://doi.org/10.1016/j.caeai.2025.100466

Li, X., Lan, Y.-J., & Pi, Z. (2024). Pedagogical agent positioning in external videos improves English academic presentation proficiency in desktop virtual reality settings. *British Journal of Educational Technology*, 56(4), 1507-1529. https://doi.org/10.1111/bjet.13531 [CORRECTION ISSUED 2025-08-17: https://doi.org/10.1111/bjet.13617]

Li, Y.-X. (2023). Cognitive engagement of low proficiency learners: Evaluating the impact of ChatGPT feedback in EFL writing. *Frontiers in Educational Research*, *6*(28). https://doi.org/10.25236/fer.2023.062809

Li, Z., Wang, Z., Wang, W., & Cheng, G. (2025). Retrieval-augmented generation for educational application: A systematic survey. Computers and Education: Artificial Intelligence, 8, Article 100417. https://doi.org/10.1016/j.caeai.2025.100417

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. Patterns, 4(7), Article 100779. https://doi.org/10.1016/j.patter.2023.100779

Lim, J., & Kessler, M. (2023). Multimodal composing and second language acquisition. *Language Teaching*, 57(2), 183-202. https://doi.org/10.1017/s0261444823000125

Lim, K., Song, J., & Park, J. (2023). Neural automated writing evaluation for Korean L2 writing. *Natural Language Engineering*, 29(5), 1341–1363. https://doi.org/10.1017/s1351324922000298

Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. International Journal of Management Education, 21(2), Article 100790. https://doi.org/10.1016/j.ijme.2023.100790

Linden, K., Hachem, H.-H., & Kondyli, V. (2025). Homo promptus: Predicting the impact of generative AI on human memory and creativity. *Memory, Mind & Media*, 4. https://doi.org/10.1017/mem.2025.10012

Little, D (2002). Autonomy in language learning: Some theoretical and practical considerations. In A. Swarbrick (Ed.), Teaching modern languages (pp. 89-95). Routledge. (Original work published 1994).

Liu, G., & Zhao, X (2025). A scoping review of AI-mediated informal language learning: Mapping out the terrain and identifying future directions. ReCALL, 38(1), 111-130. https://doi.org/10.1017/S0958344025100359

Liu, M. (2026). The rise of digital tutors: A bibliometric exploration of intelligent agent-supported language learning from 2005 to 2025. *European Journal of Education*, 61(1). https://doi.org/10.1111/ejed.70529

Liu, W. (2024). The integration of ChatGPT in language education: Exploring contexts, assessing impacts, and unpacking implications. In F. Pan (Ed.), AI in language teaching, learning, and assessment (pp. 1–21). IGI Global. https://doi.org/10.4018/979-8-3693-0872-1.ch001

Liu, W. (2025). Language teacher AI literacy: Insights from collaborations with ChatGPT. *Journal of China Computer-Assisted Language Learning*, 5(2), 287-316. https://doi.org/10.1515/jccall-2024-0030

Lobalsamo, T., Segreti, D., Jamali, M. J., & Gaspari, S. (2024). The reality of artificiality: The impact of artificial intelligence on language and culture course assessments and rubrics. In F. Pan (Ed.), AI in Language Teaching, Learning, and Assessment (pp. 43–60). IGI Global.

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1–16). ACM. https://doi.org/10.1145/3313831.3376727

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), Handbook of second language acquisition (pp. 413–468). Academic Press.

López-Costa, M., Donate-Beby, B., Cabrera-Lanzo, N., & Maina, M. F (2025). Understanding AI adoption among secondary education teachers: A PLS-SEM approach. Computers and Education: Artificial Intelligence, Article 100416. https://doi.org/10.1016/j.caeai.2025.100416

Lucas, M., Bem-haja, P., Zhang, Y., Llorente-Cejudo, C., & Palacios-Rodríguez, A (2025). A comparative analysis of pre-service teachers' readiness for AI integration. Computers and Education: Artificial Intelligence. https://doi.org/10.1016/j.caeai.2025.100396

Lys, F. (2024). Creating stories: Generative artificial intelligence tools as writing tutors. In F. Pan (Ed.), AI in language teaching, learning, and assessment (pp. 222–243). IGI Global. https://www.igi-global.com/chapter/creating-stories/339098

Lyu, W., Zhang, S., Chung, T., Sun, Y., & Zhang, Y (2025). Understanding the practices, perceptions, and (dis)trust of generative AI among instructors: A mixed-methods study in the U.S. higher education. Computers and Education: Artificial Intelligence. https://doi.org/10.1016/j.caeai.2025.100383

Ma, G., & Shi, Y. (2026). Toward a human-AI collaboration framework for intercultural competence development: Evidence from a mixed-methods study in Chinese EFL context. *Journal of China Computer-Assisted Language Learning*. Advance online publication. https://doi.org/10.1515/jccall-2025-0032

Ma, M., Ng, D. T. K., Liu, Z., & Wong, G. K. W. (2025). Fostering responsible AI literacy: A systematic review of K-12 AI ethics education. Computers and Education: Artificial Intelligence, 8, Article 100422. https://doi.org/10.1016/j.caeai.2025.100422

Mahyiyuddin, A. A., & Erlita. (2026). Bridging the gap between digital literacy and academic interest in higher education: A systematic literature review of human-centered educational technology. *Research Square*. https://doi.org/10.21203/rs.3.rs-10221208/v1

Maity, S., Deroy, A., & Sarkar, S. (2024). How ready are generative pre-trained large language models for explaining Bengali grammatical errors?. *TechRxiv*. https://doi.org/10.36227/techrxiv.171665626.60610675/v1

Maniyan, S., Ghousi, R., & Haeri, A (2024). Data mining-based decision support system for educational decision makers: Extracting rules to enhance academic efficiency. Computers & Education: Artificial Intelligence. https://doi.org/10.1016/j.caeai.2024.100242

Manorat, P., Tuarob, S., & Pongpaichet, S. (2025). Artificial intelligence in computer programming education: A systematic literature review. Computers & Education: Artificial Intelligence, 8, Article 100403. https://doi.org/10.1016/j.caeai.2025.100403

Mariyono, D. (2026). Governing agentic AI in global education: A justice-oriented framework for alignment, responsibility, and distributed autonomy [Preprint]. Research Square. https://doi.org/10.21203/rs.3.rs-8890914/v1

Mariyono, D. (2026). Artificial intelligence, power, and ethics in sustainability education: A systematic review of moral risks, governance failures, and leadership mediation. *Research Square*. https://doi.org/10.21203/rs.3.rs-8437389/v1

Markus, A., Pfister, J., Carolus, A., Hotho, A., & Wienrich, C. (2024). Empower the user: The impact of functional understanding training on usage, social perception, and self-determined interactions with intelligent voice assistants. Computers & Education: Artificial Intelligence, 6, Article 100229. https://doi.org/10.1016/j.caeai.2024.100229

Masry-Herzallah, A., & Makaldy, R (2025). Technological self-efficacy and sense of coherence: Key drivers in teachers' AI acceptance and adoption. Computers and Education: Artificial Intelligence. https://doi.org/10.1016/j.caeai.2025.100377

Mat Yusoff, S., Mohamad Marzaini, A. F., Hao, L., Zainuddin, Z., & Basal, M. H. (2025). Understanding the role of AI in Malaysian higher education curricula: An analysis of student perceptions. *Discover Computing*, 28(1). https://doi.org/10.1007/s10791-025-09567-5

Mateos-Blanco, B., Álvarez-Ramos, E., Alejaldre-Biel, L., & Parrado-Collantes, M. (2024). Vademecum of artificial intelligence tools applied to the teaching of languages. Journal of Technology and Science Education, 14(1), 77–94. https://doi.org/10.3926/jotse.2522

Mawalim, C. O., Leong, C. W., Sivan, G., Huang, H. H., & Okada, S. (2025). Beyond accuracy: Multimodal modeling of structured speaking skill indices in young adolescents. Computers and Education: Artificial Intelligence, 8, Article 100386. https://doi.org/10.1016/j.caeai.2025.100386

Mazarakis, A., Bernhard-Skala, C., & Braun, M. (2023). What is critical for human-centered AI at work? - Toward an interdisciplinary theory. *Frontiers in Artificial Intelligence*, 6. https://doi.org/10.3389/frai.2023.1257057

McDonald, N., Johri, A., Ali, A., & Hingle, A (2025). Generative artificial intelligence in higher education: Evidence from an analysis of institutional policies and guidelines. Computers in Human Behavior: Artificial Humans, 3, 100121. https://doi.org/10.1016/j.chbah.2025.100121

McKinley, J., & Rose, H. (2026). Towards inclusive feedback in multilingual writing: A Global Englishes perspective. *RELC Journal*, 57(1), 242–252. https://doi.org/10.1177/00336882261418341

META. (2022). New AI model translates 200 languages, making technology accessible to more people. https://about.fb.com/news/2022/07/new-meta-ai-model-translates-200-languages-making-technology-more-accessible/.

Meti, A. (2026). Agentic artificial intelligence in education: A review on personalized adaptive learning through autonomous tutoring systems. *International Journal of Applied Resilience and Sustainability*, *2*(2), 40–74. https://doi.org/10.70593/deepsci.0202002

Meyer, J., Jansen, T., Schiller, R., Liebenow, L. W., Steinbach, M., Horbach, A., & Fleckenstein, J. (2024). Using LLMs to bring evidence-based feedback into the classroom: AI-generated feedback increases secondary students' text revision, motivation, and positive emotions. Computers and Education: Artificial Intelligence, 6, Article 100199. https://doi.org/10.1016/j.caeai.2023.100199

Mi, Y., Rong, M., & Chen, X. (2026). Exploring the affordances and challenges of GenAI feedback in L2 writing instruction: A comparative analysis with peer feedback. *ECNU Review of Education*, 9(1). https://doi.org/10.1177/20965311241310883

Miao, F., & Cukurova, M. (2024). *AI competency framework for teachers*. UNESCO. https://doi.org/10.54675/ZJTE2084

Miao, F., Holmes, W., Huang, R., & Zhang, H. (2021a). AI and education: Guidance for policy-makers. https://unesdoc.unesco.org/ark:/48223/pf0000376709

Miralay, F. (2024). Use of artificial intelligence and augmented reality tools in art education course. Pegem Journal of Education and Instruction, 14(3), 44–50. https://doi.org/10.47750/pegegog.14.03.04

Mishra, P., & Henriksen, D. (2026). Agentic AI in education: Whose agent? Whose agency? *TechTrends*, 70(4), 799–805. https://doi.org/10.1007/s11528-026-01213-1

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, *108*(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x

Mishra, P., Warr, M., & Islam, R. (2023). TPACK in the age of ChatGPT and generative AI. Journal of Digital Learning in Teacher Education, 39(4), 235-251. https://doi.org/10.1080/21532974.2023.2247480

Mohammadi, M., Tajik, E., Martinez-Maldonado, R., Sadiq, S., Tomaszewski, W., & Khosravi, H. (2025). Artificial intelligence in multimodal learning analytics: A systematic literature review. *Computers and Education: Artificial Intelligence*, *8*, 100426. https://doi.org/10.1016/j.caeai.2025.100426

Mohebbi, A. (2025). Enabling learner independence and self-regulation in language education using AI tools: A systematic review. Cogent Education, 12(1), Article 2433814. https://doi.org/10.1080/2331186X.2024.2433814

Molenaar, I (2022). Towards hybrid human-AI learning technologies. European Journal of Education, 57(4), 632-645. https://doi.org/10.1111/ejed.12527

Morris, R., Perry, T., & Wardle, L. (2021). Formative assessment and feedback for learning in higher education: A systematic review. Review of Education, 9(3), Article e3292. https://doi.org/10.1002/rev3.3292

Mosqueira-Rey, E., Hernández-Pereira, E., Alonso-Ríos, D., Bobes-Bascarán, J., & Fernández-Leal, Á. (2023). Human-in-the-loop machine learning: A state of the art. Artificial Intelligence Review, 56, 3005–3054. https://doi.org/10.1007/s10462-022-10246-w

Mun, C. (2024). EFL learners' English writing feedback and their perception of using ChatGPT. *STEM Journal*, 25(2), 26-39. https://doi.org/10.16875/stem.2024.25.2.26

Mustofa, R. H., Kuncoro, T. G., Atmono, D., Hermawan, H. D., & Sukirman (2025). Extending the technology acceptance model: The role of subjective norms, ethics, and trust in AI tool adoption among students. Computers and Education: Artificial Intelligence. https://doi.org/10.1016/j.caeai.2025.100379

Naeem, M., Siraj, M., & Farooq, S. (2023). Teacher and administrator perceptions of using artificial intelligence in education. *Journal of Education & Social Sciences*, 11(2), 71–88. https://doi.org/10.20547/jess1122311205

Nakai, S., & Shintani, N. (2025). From 'nativelike' to practical: Comparing human and AI reformulation as written feedback. *RELC Journal*, 57(1), 48-68. https://doi.org/10.1177/00336882251397735

Nakano, M. (2026). Bridging the AI ethics gap. *International Journal of Software Innovation*, 14(1), 1-14. https://doi.org/10.4018/ijsi.401497

Namaziandost, E., & Çakmak, F. (2025). Impact of AI-generated storytelling vs. gamified learning on vocabulary retention and engagement in CALL environments. Computers and Education: Artificial Intelligence, 9, Article 100505. https://doi.org/10.1016/j.caeai.2025.100505

Namoun, A., Ibrahim, I. A., Mustafa, E., Alrehaili, A., Tufail, A., Shuja, J., Bilal, K., & Alanazi, M. H. (2024). Generative artificial intelligence in education: An umbrella review of applications and challenges. *Research Square*. https://doi.org/10.21203/rs.3.rs-4892155/v2

Nazaretsky, T., Mejia-Domenzain, P., Swamy, V., Frej, J., & Käser, T (2025). The critical role of trust in adopting AI-powered educational technology for learning: An instrument for measuring student perceptions. Computers and Education: Artificial Intelligence. https://doi.org/10.1016/j.caeai.2025.100368

Ng, D. T. K., & Lo, C. K. (2025). Opportunities, challenges and school strategies for integrating generative AI in education. Computers and Education: Artificial Intelligence, 8, Article 100373. https://doi.org/10.1016/j.caeai.2025.100373

Ng, D. T. K., Leung, J. K. L., Chu, K. W. S., & Qiao, M. S. (2021). AI literacy: Definition, teaching, evaluation and ethical issues. Proceedings of the Association for Information Science and Technology, 58(1), 504–509. https://doi.org/10.1002/pra2.487

Ng, M. L., Ghaithi, A. A., & Behforouz, B. (2025). The effect of Gemini on EFL learners' writing skills and motivation. *Research Square*. https://doi.org/10.21203/rs.3.rs-6428474/v1

Ngo, K. T. (2024). The use of ChatGPT for vocabulary acquisition: A literature review. International Journal of AI in Language Education, 1(2), 1–17. https://doi.org/10.54855/ijaile.24121

Nie, E., Liang, S., & Schmid, H. (2023). Cross-lingual retrieval augmented prompt for low-resource languages. In *Findings of the Association for Computational Linguistics: ACL 2023* (pp. 8320-8340). https://doi.org/10.18653/v1/2023.findings-acl.528

Niloy, A. C., Islam, M., Akter, S., & Moira, K. S (2025). Can generative AI be an effective co-teacher? An experiment. Computers and Education: Artificial Intelligence. https://doi.org/10.1016/j.caeai.2025.100418

Niu, H. (2026). The effects of immersive virtual reality on language learning: Causal effects on CEFR-aligned grammatical and lexical performance and within-VR comparative effects on communication and intercultural competence. *Journal of Educational Computing Research*, 64(4), 883-923. https://doi.org/10.1177/07356331261419686

Nkongolo, M. (2025). Development of a multilingual lexicon based on sentiment analysis for low-resource languages. *ResearchSquare*. https://doi.org/10.21203/rs.3.rs-6864363/v1

NLLB Team (2024). Scaling neural machine translation to 200 languages. https://doi.org/10.1038/s41586-024-07335-x

Nofal, A. B., Ali, H., Hadi, M., Ahmad, A., Qayyum, A., Johri, A., Al-Fuqaha, A., & Qadir, J. (2025). AI-enhanced interview simulation in the metaverse: Transforming professional skills training through VR and generative conversational AI. Computers and Education: Artificial Intelligence, 8, 100347. https://doi.org/10.1016/j.caeai.2024.100347

Nouzri, S., El Fatimi, M., Othmane, M., et al. (2024). Beyond chatbots: Enhancing Luxembourgish language learning through multi-agent systems and large language model. In Advances in Practical Applications of Agents, Multi-Agent Systems (pp. 385–401). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-77367-9_29

Nunes, A., Cordeiro, C., Limpo, T., et al. (2022). Effectiveness of automated writing evaluation systems in school settings: A systematic review of studies from 2000 to 2020. Journal of Computer Assisted Learning, 38(2), 599-620. https://doi.org/10.1111/jcal.12635

Nyaaba, M., & Zhai, X. (2025). Developing a theory-grounded AI tool for the generation of culturally responsive lesson plans. Computers and Education: Artificial Intelligence, 9, Article 100474. https://doi.org/10.1016/j.caeai.2025.100474

Nyaaba, M., Kyeremeh, P., & Nabang, M. (2026). GenAITEd Ghana as a context-aware and curriculum-aligned conversational AI agent for teacher education [Preprint]. Research Square. https://doi.org/10.21203/rs.3.rs-8564669/v1

Oda, M., & Saito, R. (2025). Development and evaluation of a course integrating machine translation and generative AI in higher education English learning. *JALTCALL Trends*, 1(1), 2151. https://doi.org/10.29140/jct.v1n1.2151

Ode, E., Nana, R., & Boro, I. O. (2025). A cross-country analysis of self-determination and continuance use intention of AI tools in business education: Does instructor support matter?. Computers and Education: Artificial Intelligence, 8, Article 100402. https://doi.org/10.1016/j.caeai.2025.100402

OECD (2019a). OECD future of education and skills 2030: OECD learning compass 2030 (A series of concept notes). https://www.oecd.org/content/dam/oecd/en/about/projects/edu/education-2040/1-1-learning-compass/OECD_Learning_Compass_2030_Concept_Note_Series.pdf

OECD (2019b). Recommendation of the Council on Artificial Intelligence (OECD/LEGAL/0449) [OECD AI Principles]. https://oecd.ai/en/ai-principles

OECD. (2021). OECD digital education outlook 2021. OECD Publishing. https://doi.org/10.1787/589b283f-en

OECD. (2023). OECD digital education outlook 2023: Towards an effective digital education ecosystem. OECD Publishing. https://doi.org/10.1787/c74f03de-en

Okonkwo, C. W., & Ade-Ibijola, A (2021). Chatbots applications in education: A systematic review. Computers and Education: Artificial Intelligence, 2, 100033. https://doi.org/10.1016/j.caeai.2021.100033

Oliveira, M., Zednik, C., Bombaerts, G., Sadowski, B., & Conijn, R. (2025). Assessing students' DRIVE: A framework to evaluate learning through interactions with generative AI. Computers and Education: Artificial Intelligence, 9, Article 100497. https://doi.org/10.1016/j.caeai.2025.100497

Pack, A., Barrett, A., & Escalante, J. (2024). Large language models and automated essay scoring of English language learner writing: Insights into validity and reliability. Computers & Education: Artificial Intelligence, 6, Article 100234. https://doi.org/10.1016/j.caeai.2024.100234

Pallivathukal, R. G., Soe, H. H. K., Donald, P. M., Samson, R. S., & Ismail, A. R. H. (2024). ChatGPT for academic purposes: Survey among undergraduate healthcare students in Malaysia. *Cureus*. https://doi.org/10.7759/cureus.53032

Pan, F. (Ed.). (2024). AI in language teaching, learning, and assessment. IGI Global. https://www.igi-global.com/book/language-teaching-learning-assessment/325066

Park, J. S., O'Brien, J., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. In Proceedings of the 36th Annual

ACM Symposium on User Interface Software and Technology (UIST '23) (pp. 1–22). ACM. https://doi.org/10.1145/3586183.3606763

Parsons, D., MacCallum, K., & Burilin, E. (2024). Supporting tertiary students' critical AI literacy: A practical tool for educators. *He Rourou*, 1(1), 10. https://doi.org/10.54474/herourou.v1i1.10577

Paul, D., Gohain, D., Thomas, T. G., & Harrini S, B. J. (2026). From theory to practice: A human-AI synergy framework for ethical, inclusive and pedagogically-driven language education. *Frontiers in Education*, 11, Article 1841386. https://doi.org/10.3389/feduc.2026.1841386

Pei, B., Lu, J., & Jing, X (2025). Empowering preservice teachers' AI literacy: Current understanding, influential factors, and strategies for improvement. Computers and Education: Artificial Intelligence. https://doi.org/10.1016/j.caeai.2025.100406

Pérez-Paredes, P., Curry, N., & Ordoñana Guillamón, C. (2025). Critical AI literacy for applied linguistics and language education students. *Journal of China Computer-Assisted Language Learning*, 5(2), 175-214. https://doi.org/10.1515/jccall-2025-0005

Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024a). The Artificial Intelligence Assessment Scale (AIAS): A framework for ethical integration of generative AI in educational assessment. Journal of University Teaching and Learning Practice, 21(6). https://doi.org/10.53761/q3azde36

Perkins, M., Roe, J., Vu, B. H., et al. (2024b). Simple techniques to bypass GenAI text detectors: Implications for inclusive education. International Journal of Educational Technology in Higher Education, 21, Article 53. https://doi.org/10.1186/s41239-024-00487-w

Pesovski, I., Jolakoski, P., Trajkovik, V., Kubincova, Z., & Herzog, M. A. (2025). Predicting student achievement through peer network analysis for timely personalization via generative AI. Computers and Education: Artificial Intelligence, 8, 100430. https://doi.org/10.1016/j.caeai.2025.100430

Pham, N. N. H. (2025). The use of ChatGPT in EFL students as a learning assistant in writing skills: A literature review. International Journal of AI in Language Education, 2(1), 38–54. https://doi.org/10.54855/ijaile.25213

Pikhart, M., & Klímová, B. (2025). A qualitative study on ethical issues related to the use of AI-driven technologies in foreign language learning. *Scientific Reports*, 15(1). https://doi.org/10.1038/s41598-025-13741-6

Pitychoutis, K. M., & Al Rawahi, A. (2024). Smart teaching: The synergy of multiple intelligences and artificial intelligence in English as a foreign language instruction. *Forum for Linguistic Studies*, *6*(6), 249–260. https://doi.org/10.30564/fls.v6i6.7297

Plakans, L., & Lee, K. (2025). Fairness, justice, and criticality: Reviewing second language writing assessment. *Language Teaching*, 59(1), 1-28. https://doi.org/10.1017/s0261444825100876

Popkov, A. A., & Barrett, T. S. (2024). AI vs academia: Experimental study on AI text detectors' accuracy in behavioral health academic writing. Accountability in Research, 32(7), 1072–1088. https://doi.org/10.1080/08989621.2024.2331757

Prinsloo, P (2020). Data frontiers and frontiers of power in (higher) education: A view of/from the Global South. Teaching in Higher Education, 25(4), 366-383. https://doi.org/10.1080/13562517.2020.1723537

Puentedura, R. R. (2010). *SAMR: A brief introduction*. Hippasus. hippasus.com

Qiao, H., & Zhao, A (2023). Artificial intelligence-based language learning: Illuminating the impact on speaking skills and self-regulation in Chinese EFL contexts. Frontiers in Psychology, Article 1255594. https://doi.org/10.3389/fpsyg.2023.1255594

Qiao, S., Gu, M., & Lu, C. (2025). Artificial intelligence for language learning: A systematic review of its design, theoretical foundations, implementation, and impact. *International Journal of Applied Linguistics*. Advance online publication. https://doi.org/10.1111/ijal.70034

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2022). Robust speech recognition via large-scale weak supervision. arXiv. https://doi.org/10.48550/arXiv.2212.04356

Radtke, A., & Rummel, N. (2025). Generative AI in academic writing: Does information on authorship impact learners' revision behavior? Computers and Education: Artificial Intelligence, 8, Article 100350. https://doi.org/10.1016/j.caeai.2024.100350

Rahimi, A. R. (2025). Developing and validating the scale of language teachers' design thinking competency in artificial intelligence language teaching (LTDTAILT). Computers and Education: Artificial Intelligence, 8, Article 100420. https://doi.org/10.1016/j.caeai.2025.100420

Ramesh, D., & Sanampudi, S. K. (2021). An automated essay scoring systems: A systematic literature review. *Artificial Intelligence Review*, 55(3), 2495-2527. https://doi.org/10.1007/s10462-021-10068-2

Ramos-Benitez, M. J., García-Osorio, M. E., & Delgado, Y. (2026). Critical alliance of AI in education: A pedagogical framework for safeguarding cognitive skills. *International Medical Education*, 5(1), Article 22. https://doi.org/10.3390/ime5010022

Ranalli, J. (2021). L2 student engagement with automated feedback on writing: Potential for learning and issues of trust. Journal of Second Language Writing, 52, Article 100816. https://doi.org/10.1016/j.jslw.2021.100816

Rane, J., Kaya, Ö., Mallick, S. K., & Rane, N. L. (2024). Artificial intelligence in education: A SWOT analysis of ChatGPT and its implications for practice and research. In *Generative Artificial Intelligence in Agriculture, Education, and Business*. https://doi.org/10.70593/978-81-981271-7-4_4

Ranieri, M., Biagini, G., & Cuomo, S. (2025). AI literacy in higher education. *International Journal of Digital Literacy and Digital Competence*, 16(1), 1-25. https://doi.org/10.4018/ijdldc.388469

Redecker, C. (2017). European framework for the digital competence of educators: DigCompEdu. https://publications.jrc.ec.europa.eu/repository/handle/JRC107466

Rodgers, C. R., & Raider-Roth, M. B. (2006). Presence in teaching. Teachers and Teaching: Theory and Practice, 12(3), 265–287. https://doi.org/10.1080/13450600500467548

Rodrigues, L., Xavier, C., Torrezão da Costa, N., Gašević, D., & Mello, R. F. (2025). Is GPT-4 fair? An empirical analysis in automatic short answer grading. Computers and Education: Artificial Intelligence, 8, Article 100428. https://doi.org/10.1016/j.caeai.2025.100428

Rogers, E. M. (2003). Diffusion of Innovations (5th ed.). Free Press. NY.

Romero-Rodríguez, J.-M., Ramírez-Montoya, M.-S., Buenestado-Fernández, M., & Lara-Lara, F. (2023). Use of ChatGPT at university as a tool for complex thinking: Students' perceived usefulness. *Journal of New Approaches in Educational Research*, 12(2), 323–339. https://doi.org/10.7821/naer.2023.7.1458

Ronksley-Pavia, M., Nguyễn, T. N. L., Wheeley, E., et al. (2025). A scoping literature review of generative artificial intelligence for supporting neurodivergent school students. Computers and Education: Artificial Intelligence, 9, Article 100437. https://doi.org/10.1016/j.caeai.2025.100437

Roy, R., Babakerkhell, M. D., Mukherjee, S., Pal, D., & Funilkul, S. (2022). Evaluating the intention for the adoption of artificial intelligence-based robots in the university to educate the students. *IEEE Access*, 10, 125666–125678. https://doi.org/10.1109/access.2022.3225555

Rui, J. (2025). Evaluating the longitudinal effects of AI-enhanced collaborative dialogue modes on computational thinking and language proficiency in EFL learners: A mixed-methods approach. *Journal of Educational Computing Research*, 64(3), 539-570. https://doi.org/10.1177/07356331251399452

Saeedi, Z., & Soltani, M. (2025). Developing an AI chatbot for language pragmatics instruction: From algorithms to dynamic assessment in an EFL context. *Computer Assisted Language Learning*, 1–22. https://doi.org/10.1080/09588221.2025.2532014

Sangkala, I., & Mardonovna, N. S (2024). Artificial intelligence as a personalized tutor in language learning: A systematic review. Klasikal: Journal of Education, 565-576. https://doi.org/10.52208/klasikal.v6i2.1193

Sardi, J., Darmansyah, Candra, O., Yuliana, D. F., Habibullah, Yanto, D. T. P., & Eliza, F. (2025). How generative AI influences students' self-regulated learning and critical thinking skills? A systematic review. *International Journal of Engineering Pedagogy (iJEP)*, 15(1), 94–108. https://doi.org/10.3991/ijep.v15i1.53379

Sargsyan, L. (2025). Integrating agentic AI in higher education: Balancing opportunities, challenges, and ethical imperatives. *Foreign Languages in Higher Education*, 29(1), 87–100. https://doi.org/10.46991/flhe.2025.29.1.087

Sayici, S., & Aydin, S. (2025). Using artificial intelligence in foreign language teaching: Teachers' perspectives. Proceedings of the World Conference on Education and Teaching, 4(1), 1–10. https://doi.org/10.33422/etconf.v4i1.1016

Scaria, N., Kennedy, S. J. J., & Subramani, D. (2026). Learning in blocks: A multi agent debate assisted personalized adaptive learning framework for language learning. arXiv. https://doi.org/10.48550/arXiv.2604.22770

Schmidt, R. (1990). The role of consciousness in second language learning. Applied Linguistics, 11(2), 129–158. https://doi.org/10.1093/applin/11.2.129

Schulze, M. (2025). ICALL and AI: Seven lessons from seventy years. In Y. Wang, A. Alm, & G. Dizon (Eds.), Insights into AI and language teaching and learning (pp. 11-31). Castledown, 11-31. https://doi.org/10.29140/9781763711600-02

Segbenya, M., Bervell, B., Frimpong-Manso, E., et al (2023). Artificial intelligence in higher education: Modelling the antecedents of artificial intelligence usage and effects on 21st century employability skills among postgraduate students in Ghana. Computers and Education: Artificial Intelligence, 5, 100188. https://doi.org/10.1016/j.caeai.2023.100188

Seung, Y., & Basham, J. D. (2026). Cognitive offloading in the age of generative AI: What does it mean for students with learning disabilities? *Learning Disability Quarterly*, 49(3), 121-133. https://doi.org/10.1177/07319487261439132

Shafiee Rad, H. (2024). Revolutionizing L2 speaking proficiency, willingness to communicate, and perceptions through artificial intelligence: A case of the Speeko application. Innovation

in Language Learning and Teaching, 18(4), 364–379. https://doi.org/10.1080/17501229.2024.2309539

Shi, H., & Aryadoust, V (2024). A systematic review of AI-based automated written feedback research. ReCALL, 36(2), 187-209. https://doi.org/10.1017/S0958344023000265

Shi, L., & Derakhshan, A. (2026). Exploring the role of AI in EFL students' autonomy, relatedness, and competence through the lens of self-determination theory (SDT): A latent growth curve modeling (LGCM) analysis. *International Journal of Applied Linguistics*. https://doi.org/10.1111/ijal.70272

Şimşek, A. C., Anders, G., Göth, J., Specht, L., & Huff, M. (2025). Is ChatGPT a good study companion? The role of AI-generated summaries and reflective prompts in learning from educational videos. Computers and Education: Artificial Intelligence, 9, Article 100512. https://doi.org/10.1016/j.caeai.2025.100512

Slyusar, V. (2025). A multi-agent environment with artificial intelligence for English language learning and teaching (pp. 135-182). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3373-3561-2.ch006

Son, J.-B., Ružić, N. K., & Philpott, A. (2025). Artificial intelligence technologies and applications for language learning and teaching. *Journal of China Computer-Assisted Language Learning*, 5(1), 94–112. https://doi.org/10.1515/jccall-2023-0015

Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. Frontiers in Psychology, 14, Article 1260843. https://doi.org/10.3389/fpsyg.2023.1260843

Song, Y., Weisberg, L. R., Zhang, S., Tian, X., Boyer, K. E., & Israel, M (2024). A framework for inclusive AI learning design for diverse learners. Computers & Education: Artificial Intelligence. https://doi.org/10.1016/j.caeai.2024.100212

Soomro, R. B. K., Sahito, Z., Soomro, A. B., Bhayo, N. H., & Soomro, A. H. G. (2026). Architecting the future: A cross-national analysis of industry-academia co-developed AI microcredentials and their impact on pedagogical innovation. *Research Square*. https://doi.org/10.21203/rs.3.rs-8469128/v1

Soto, C. (2026). Automated reading fluency assessment: A multilayer educational framework for scalable literacy evaluation. *Reading Psychology*, 1-12. https://doi.org/10.1080/02702711.2026.2692349

Sriganesh, N., & Eamer, A. (2025). Chatting with Al Capone: Leveraging virtual conversational agents in emotional landscapes. *Australian Journal of Applied Linguistics*, *8*(2), Article 102513. https://doi.org/10.29140/ajal.v8n2.102513

Stockwell, G. (2022). Mobile assisted language learning: Concepts, contexts and challenges. Cambridge University Press. https://doi.org/10.1017/9781108652087

Suhan, M., & Wolf, M. K (2025). A comparative study of the human, automated scoring model, and GPT-4 ratings of young EFL students' writing. Language Testing, 43(1), 66-78. https://doi.org/10.1177/02655322251346860

Swain, M. (1995). Three functions of output in second language learning. In G. Cook, & B. Seidlhofer (Eds.), Principle and practice in applied linguistics: Studies in honour of H. G. Widdowson (pp. 125-144). Oxford: Oxford University Press.

Sweller, J (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science. https://doi.org/10.1016/0364-0213(88)90023-7

Tai, T.-Y (2024). Exploring the effects of intelligent personal assistants on bilingual education teachers' L2 speaking proficiency. Interactive Learning Environments, 33(1), 440-451. https://doi.org/10.1080/10494820.2024.2350646

Talehati. (2025). Bridging the digital divide: An analysis of AI and VR integration for equitable English language education. *European Journal of Social Science Education and Research, 12*(4). https://doi.org/10.26417/31v0hq49

Tan, L. Y., Hu, S., Yeo, D. J., & Cheong, K. H. (2025). Artificial intelligence-enabled adaptive learning platforms: A review. Computers and Education: Artificial Intelligence, 9, Article 100429. https://doi.org/10.1016/j.caeai.2025.100429

Tan, X., Cheng, G., & Ling, M. H. (2025a). Enhancing teachers' AI competency: A professional development intervention study based on intelligent-TPACK framework. Computers and Education: Artificial Intelligence, 9, Article 100521. https://doi.org/10.1016/j.caeai.2025.100521

Tan, X., Cheng, G., & Ling, M. H. (2025b). Investigating the mediating role of TPACK on teachers' AI competency and their teaching performance in higher education. Computers and Education: Artificial Intelligence, 9, Article 100461. https://doi.org/10.1016/j.caeai.2025.100461

Tan, X. J., Cheng, G., & Ling, M. H. (2025). Artificial intelligence in teaching and teacher professional development: A systematic review. Computers and Education: Artificial Intelligence, 8, Article 100355. https://doi.org/10.1016/j.caeai.2024.100355

Tang, J., Li, P., & Luo, R. (2026). Teacher feedback VS AI-assisted peer feedback in L2 writing: A quasi-experimental study in a Chinese university. *PLOS One*, 21(6), e0345976. https://doi.org/10.1371/journal.pone.0345976

Tang, Y., Du, X., Hung, J.-L., Wang, M., & Ding, T. (2026). Enhancing collaborative problem solving through AI personal assistant interactions: Insights from diverse prior knowledge groups. *IEEE Transactions on Learning Technologies*, 19, 766–777. https://doi.org/10.1109/tlt.2026.3713908

Tao, S., Lan, M., Wang, M., & Li, H. (2026). Potential risks of generative artificial intelligence integration into K-12 education: A scoping review. *Computers and Education: Artificial Intelligence*, 10, 100561. https://doi.org/10.1016/j.caeai.2026.100561

Tashtoush, M. A., Wardat, Y., Ali, R. A., & Saleh, S. (2024). Artificial intelligence in education: Mathematics teachers' perspectives, practices and challenges. *Iraqi Journal for Computer Science and Mathematics*, 5(1). https://doi.org/10.52866/ijcsm.2024.05.01.004

The White House. (2023). Executive order on the safe, secure, and trustworthy development and use of artificial intelligence. https://www.whitehouse.gov. https://bidenwhitehouse.archives.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/

Thi, N. K., & Nikolov, M. (2023). Effects of teacher, automated, and combined feedback on syntactic complexity in EFL students' writing. *Asian-Pacific Journal of Second and Foreign Language Education*, 8(1). https://doi.org/10.1186/s40862-022-00182-1

Thomas, M. L., Yildirim-Erbasli, S. N., & Hariharan, S. (2026). Exploring undergraduate students' perceptions of AI vs. human scoring and feedback. *The Internet and Higher Education*, 68, 101052. https://doi.org/10.1016/j.iheduc.2025.101052

Tongchai, A., & Malakul, S. (2025). Thai teachers' perceptions of integrating generative AI in K-12 education: Opportunities and challenges. *TechTrends*, 70(1), 93-105. https://doi.org/10.1007/s11528-025-01128-3

Torres, P. J., & Kahveci, Y. E. (2025). Effectiveness of artificial intelligence (AI) in language teaching. *Computers and Education: Artificial Intelligence*, 9, 100522. https://doi.org/10.1016/j.caeai.2025.100522

Tran, H. N., Le, T. T. N., & Tran, V. B. U (2025). AI tools in learning academic writing: Benefits and challenges for MA students in the English language studies at the Industrial University of Ho Chi Minh City. International Journal of AI in Language Education. https://doi.org/10.54855/ijaile.25215

Tran, T. T. P., Dang, T. N., & Nguyen, V. L. P (2025). Master students' perceptions of how ChatGPT influenced critical thinking in academic writing at The Industrial University of Ho Chi Minh City. International Journal of AI in Language Education. https://doi.org/10.54855/ijaile.25222

Tse, S. (2026). Exploring student perceptions of generative AI therapists and their benefits and challenges in schools. *International Journal of Psychology*, *61*(3). https://doi.org/10.1002/ijop.70207

Tu, Y.-F., Liu, G.-P., Hwang, G.-J., Chen, X.-W., & Guo, X.-G. (2026). AI self-efficacy and knowledge graph-integrated generative AI feedback in higher education. *The Internet and Higher Education*, 69, 101079. https://doi.org/10.1016/j.iheduc.2026.101079

Türk, N., Batuk, B., Kaya, A., & Yıldırım, O. (2025). What makes university students accept generative artificial intelligence? A moderated mediation model. *BMC Psychology*, 13(1). https://doi.org/10.1186/s40359-025-03559-2

Turós, M. (2025). What percentage of secondary school students do their homework with the help of artificial intelligence? - A survey of attitudes towards artificial intelligence. Computers and Education: Artificial Intelligence, 8, Article 100394. https://doi.org/10.1016/j.caeai.2025.100394

UNESCO. (2019). Beijing consensus on artificial intelligence and education. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000368303

UNESCO. (2021b). Recommendation on the ethics of artificial intelligence. UNESCO. https://unesdoc.unesco.org. https://unesdoc.unesco.org/ark:/48223/pf0000381137

UNESCO (2023). Guidance for generative AI in education and research. https://unesdoc.unesco.org/ark:/48223/pf0000386693

Uslu, B., & Bozkurt, A. (2026). Regulations of AI technologies and their impact on higher education: Global perspectives, institutional cases and emerging challenges. *Higher Education Quarterly*, *80*(2). https://doi.org/10.1111/hequ.70108

Valeri, F., Nilsson, P., & Cederqvist, A.-M (2025). Exploring students' experience of ChatGPT in STEM education. Computers & Education: Artificial Intelligence. https://doi.org/10.1016/j.caeai.2024.100360

van Niekerk, J., Delport, P. M. J., & Sutherland, I. (2025). Addressing the use of generative AI in academic writing. Computers and Education: Artificial Intelligence, 8, Article 100342. https://doi.org/10.1016/j.caeai.2024.100342

Varsamidou, A (2024). Integration of artificial intelligence in foreign language teaching: Exploring views and beliefs of Greek foreign language teachers. International Journal of Language and Literary Studies, 6(3), 299-320. https://doi.org/10.36892/ijlls.v6i3.1818

Viet, Q. P., Van Nguyen, D., Xiem, N. T., & Nguyen, D. N. (2026). Survey on academic self-efficacy, academic stress, critical thinking, performance expectations, and students' AI dependency. *Data in Brief*, 67, 112952. https://doi.org/10.1016/j.dib.2026.112952

Voskoboinik, E., von Zansen, A., Phan, N. C., Getman, Y., Grosz, T., & Kurimo, M (2025). Enhancing second language speaking assessment: Integrating large language models for Finnish and Finland Swedish proficiency scoring. Language Testing, 42(4), 508-538. https://doi.org/10.1177/02655322251351648

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press.

Wang, C., & Zou, B. (2025). D-ID Studio: Empowering language teaching with AI avatars. TESOL Journal, 16, Article e70034. https://doi.org/10.1002/tesj.70034

Wang, F., Cheung, A., & Neitzel, A. (2024). Does chatting with chatbots improve language learning performance? A meta-analysis of chatbot-assisted language learning. *Review of Educational Research*, 95(4), 623-660. https://doi.org/10.3102/00346543241255621

Wang, G.-Y., Hatori, Y., Satō, Y., et al. (2025). Predicting learners' engagement and help-seeking behaviors in an e-learning environment by using facial and head pose features. Computers and Education: Artificial Intelligence, 8, Article 100387. https://doi.org/10.1016/j.caeai.2025.100387

Wang, J. (2024). Integrating ChatGPT into a culture-focused flipped curriculum for Chinese as a foreign language: Experiments and implications. In F. Pan (Ed.), AI in Language Teaching, Learning, and Assessment (pp. 173–194). IGI Global. https://www.igi-global.com/chapter/integrating-chatgpt-into-a-culture-focused-flipped-curriculum-for-chinese-as-a-foreign-language/339096

Wang, Y., Huang, J., Du, L., Guo, Y., Liu, Y., & Wang, R. (2025). Evaluating large language models as raters in large-scale writing assessments: A psychometric framework for reliability and validity. Computers & Education: Artificial Intelligence, 9, Article 100481. https://doi.org/10.1016/j.caeai.2025.100481

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. https://doi.org/10.1080/17501229.2025.2509759

Warschauer, M., & Healey, D (1998). Computers and language learning: An overview. Language Teaching, 31(2), 57-71. https://doi.org/10.1017/S0261444800012970

Wei, L. (2023). Artificial intelligence in language instruction: Impact on English learning achievement, L2 motivation, and self-regulated learning. *Frontiers in Psychology*, 14. https://doi.org/10.3389/fpsyg.2023.1261955

Weng, X., & Chiu, T. K. F (2023). Instructional design and learning outcomes of intelligent computer assisted language learning: Systematic review in the field. Computers and Education: Artificial Intelligence, 4, 100117. https://doi.org/10.1016/j.caeai.2022.100117

Wetzler, E. L., Cassidy, K. S., Jones, M. J., Frazier, C. R., Korbut, N. A., Sims, C. M., Bowen, S. S., & Wood, M. (2024). Grading the graders: Comparing generative AI and human

assessment in essay evaluation. *Teaching of Psychology*, 52(3), 298–304. https://doi.org/10.1177/00986283241282696

Wiboolyasarin, W., Wiboolyasarin, K., Suwanwihok, K., Jinowat, N., & Muenjanchoey, R. (2024). Synergizing collaborative writing and AI feedback: An investigation into enhancing L2 writing proficiency in wiki-based environments. Computers & Education: Artificial Intelligence, 6, Article 100228. https://doi.org/10.1016/j.caeai.2024.100228

Wiese, L., Patil, I., Schiff, D., & Magana, A. J. (2025). AI ethics education: A systematic literature review. Computers and Education: Artificial Intelligence, 8, Article 100405. https://doi.org/10.1016/j.caeai.2025.100405

Willert, N., & Würz, P. K (2025). Assisting quality assurance of examination tasks: Using a GPT model and Bayesian testing for formative assessment. Computers and Education: Artificial Intelligence. https://doi.org/10.1016/j.caeai.2024.100343

Williams, P. (2025). Human-AI learning: Architecture of a human-AgenticAI learning system. *Information*, 16(12), Article 1101. https://doi.org/10.3390/info16121101

Williamson, B. (2024). The social life of AI in education. International Journal of Artificial Intelligence in Education, 34(1), 97–104. https://doi.org/10.1007/s40593-023-00342-5

Wong, C. W. (2026). The cognitive exoskeleton and the E.F.F.E.C.T. framework: Scaffolding affective shifts and self-regulated learning in L2 writing-An exploratory micro-genetic pilot study. *OSF Preprints*. https://doi.org/10.35542/osf.io/k9nm6_v1

Woollaston, S., Flanagan, B., & Wijerathne, I. (2026). Agentic AI and pedagogical best practice: The tension between automation and learning. arXiv. https://doi.org/10.48550/arXiv.2606.04543

Wu, M., Hao, H., Bi, S., et al. (2026). Scaling laws for educational AI agents. arXiv. https://doi.org/10.48550/arXiv.2603.11709

Wu, Q., Bansal, G., Zhang, J., et al. (2023). AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation. arXiv, arXiv:2308.08155. https://doi.org/10.48550/arXiv.2308.08155

Wu, X.-Y. (2024). Artificial Intelligence in L2 learning: A meta-analysis of contextual, instructional, and social-emotional moderators. System, 126, Article 103498. https://doi.org/10.1016/j.system.2024.103498

Xi, X (2010). Automated scoring and feedback systems: Where are we and where are we heading?. Language Testing, 27(3), 291-300. https://doi.org/10.1177/0265532210364643

Xi, Z., Chen, W., Guo, X., et al. (2023). The rise and potential of large language model based agents: A survey. arXiv. https://doi.org/10.48550/arXiv.2309.07864

Xiao, F., Li, Z., Lin, J., Zou, X., Yang, D., Zou, W., & Xiong, J. (2025). Leveraging an LLM-enhanced bilingual conversational agent for EFL children's dialogic reading: Insights from children, parents, and educators. Computers and Education: Artificial Intelligence, 9, Article 100484. https://doi.org/10.1016/j.caeai.2025.100484

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). https://doi.org/10.1177/20965311241310881

Xiao, F., Zou, E. W., & Lin, J. (2025). Parent-led vs. AI-guided dialogic reading: Evidence from a randomized controlled trial in children's e-book context. *British Journal of Educational Technology*, 56(5), 1784-1813. https://doi.org/10.1111/bjet.13615

Xing, W., Kim, T., & Song, Y. (2026). Unveiling interaction patterns between students and generative AI teachable agent: Focusing on students' agency and AI agents' authority. *British Journal of Educational Technology*, 57(4), 896-923. https://doi.org/10.1111/bjet.70038

Xu, B., & Ismail, H. H. (2025). Evaluating the effectiveness of using AI in spoken English: A literature review. International Journal of Academic Research in Progressive Education and Development, 14(3), 1819–1835. https://doi.org/10.6007/IJARPED/v14-i3/26366

Xu, L. Z., Velez, G., & Robinson, A. (2026). When confidence shifts toward the tool: AI appropriation, reliance calibration, and school guidance in high school academic work. *Research Square*. https://doi.org/10.21203/rs.3.rs-10233077/v1

Xue, L. (2024). Urgent, but how? Developing English foreign language teachers' digital literacy in a professional learning community focusing on large language models. *European Journal of Education*, 60(1). https://doi.org/10.1111/ejed.12899

Xun, H., & Wang, X. (2026). Mapping the triadic interaction network in AI-enhanced L2 classrooms: A social network analysis of learning outcomes and the moderating role of AI literacy. *System*, 140, 104050. https://doi.org/10.1016/j.system.2026.104050

Yan, L., Echeverría, V., & Jin, Y. (2024). Evidence-based multimodal learning analytics for feedback and reflection in collaborative learning. *British Journal of Educational Technology*, 55(5), 1900-1925. https://doi.org/10.1111/bjet.13498

Yan, L., Greiff, S., & Lodge, J. M. (2025). Distinguishing performance gains from learning when using generative AI. *Nature Reviews Psychology*, 4(7), 435-436. https://doi.org/10.1038/s44159-025-00467-5

Yan, W., Nakajima, T., & Sawada, R. (2025). Beyond tool use: Tracking the evolution of generative AI literacy among university students through a process-oriented investigation. Computers and Education: Artificial Intelligence, 9, Article 100465. https://doi.org/10.1016/j.caeai.2025.100465

Yan, Y., & Chau, T. (2025). A systematic review of AI ethics in education. *Journal of Global Information Management*, 33(1), 1-50. https://doi.org/10.4018/jgim.386381

Yang, C., Chen, J., Hou, S., Wider, W., & Yi, W. L. (2025). Charting the developmental landscape of artificial intelligence in language education using bibliometric methods. Discover Artificial Intelligence. https://doi.org/10.1007/s44163-025-00732-9

Yang, K., Raković, M., Liang, Z., Yan, L., Zeng, Z., Fan, Y., Gašević, D., & Chen, G. (2025). Modifying AI, enhancing essays: How active engagement with generative AI boosts writing quality. In Proceedings of the 15th International Learning Analytics and Knowledge Conference (pp. 568–578). https://doi.org/10.1145/3706468.3706544

Yang, S., Liu, Y., & Wu, T.-C. (2024). ChatGPT, a new "Ghostwriter": A teacher-and-students poetic autoethnography from an EMI academic writing class. *Digital Applied Linguistics*, 1, 2244. https://doi.org/10.29140/dal.v1.2244

Yang, S. J., Ogata, H., Matsui, T., & Chen, N.-S. (2021). Human-centered artificial intelligence in education: Seeing the invisible through the visible. Computers and Education: Artificial Intelligence, 2, Article 100008. https://doi.org/10.1016/j.caeai.2021.100008

Yang, W., Li, H., & Lee, J. C. K. (2025). Tailoring AI agents for early learning: The Creative Project Approach. Computers and Education: Artificial Intelligence, 9, Article 100473. https://doi.org/10.1016/j.caeai.2025.100473

Yilmaz, H., Maxutov, S., Baitekov, A., & Balta, N. (2023). Student attitudes towards chat GPT: A technology acceptance model survey. *International Educational Review*, 1(1), 57–83. https://doi.org/10.58693/ier.114

Yu, H., Guo, Y., Yang, H., Zhang, W., & Dong, Y. (2024). Can ChatGPT revolutionize language learning? Unveiling the power of AI in multilingual education through user insights and pedagogical impact. *European Journal of Education*, *60*(1), Article e12749. https://doi.org/10.1111/ejed.12749

Yu, S., Zhang, E. D., & Liu, C. (2024). Research into practice: Digital multimodal composition in second language writing. *Language Teaching*, 59(2), 197-213. https://doi.org/10.1017/s0261444824000375

Zai, F., & Zhou, X. (2026). The impact of AI-assisted learning on the agency of foreign language learners: A meta-analysis. *Behavioral Sciences*, 16(3), Article 379. https://doi.org/10.3390/bs16030379

Zaman, S., Hussain, M. S., & Tabassam, M. (2024). Use of artificial intelligence in education: English language teachers' identity negotiation in higher education. Journal of Asian Development Studies, 13(3), 861–869. https://poverty.com.pk/index.php/Journal/article/view/763

Zeevy-Solovey, O. (2024). Comparing peer, ChatGPT, and teacher corrective feedback in EFL writing: Students' perceptions and preferences. *Technology in Language Teaching & Learning*, 6(3), 1482. https://doi.org/10.29140/tltl.v6n3.1482

Zhai, X., Krajcik, J., & Pellegrino, J. W. (2021). On the validity of machine learning-based Next Generation Science Assessments: A validity inferential network. Journal of Science Education and Technology, 30(2), 298–312. https://doi.org/10.1007/s10956-020-09879-9

Zhang, C., Meng, Y., & Ma, X (2024). Artificial intelligence in EFL speaking: Impact on enjoyment, anxiety, and willingness to communicate. System, 121. https://doi.org/10.1016/j.system.2024.103259

Zhang, D., Wu, J. G., & Fu, Z. (2024). From shy to fly: Facilitating EFL learners' willingness to communicate with an AI chatbot and an intelligent tutoring system. System, 127, Article 103501. https://doi.org/10.1016/j.system.2024.103501

Zhang, J., & Rhee, G. S. (2025). AI ethics knowledge among Chinese teachers: An empirical assessment. *Research Square*. https://doi.org/10.21203/rs.3.rs-7042877/v1

Zhang, J., Hussin, M., & Abd Majid, M. Z. (2025). Artificial intelligence in language teaching in the context of higher vocational education: A bibliometric mapping of research trends (2000–2024). Forum for Linguistic Studies, 7(9), 850–864. https://doi.org/10.30564/fls.v7i9.10958

Zhang, L., Carter, R. A., Jr., & Lim, S. N. (2025). Designing intelligent agents for students with disabilities: Promoting inclusion and equity through the lens of cultural-historical activity theory. *Journal of Special Education Technology*, 41(2), 163–176. https://doi.org/10.1177/01626434251349405

Zhang, N., & He, X. Z (2024). Corpus-based research on English–Chinese translation teaching combining vocabulary learning and practice. Journal of Electrical Systems, 20(3s). https://doi.org/10.52783/jes.3187

Zhang, Q., Nie, H., Fan, J., & Liu, H. (2025). Exploring the dynamics of artificial intelligence literacy on English as a foreign language learners' willingness to communicate: The critical

mediating roles of artificial intelligence learning self-efficacy and classroom anxiety. *Behavioral Sciences*, 15(4), 523. https://doi.org/10.3390/bs15040523

Zhang, R., Zou, D., Xie, H., & Wang, F. L. (2025). Engagements with GPT responses and learner prompts in ChatGPT-based learning of English argumentative writing logic and their impacts. Computers and Education: Artificial Intelligence, 9, Article 100489. https://doi.org/10.1016/j.caeai.2025.100489

Zhang, Y. (2024). A lesson study on a MOOC-based and AI-powered flipped teaching and assessment of EFL writing model: Teachers' and students' growth. International Journal for Lesson & Learning Studies, 13(1), 28–40. https://doi.org/10.1108/IJLLS-07-2023-0085

Zhang, Y. (2025). Exploring the role of AI-mediated L2 education in Chinese EFL teachers' classroom rapport with their students. *European Journal of Education*, *60*(3). https://doi.org/10.1111/ejed.70142

Zheng, Y., & Stewart, N. (2024). Improving EFL students' cultural awareness: Reframing moral dilemmatic stories with ChatGPT. Computers and Education: Artificial Intelligence, 6, Article 100223. https://doi.org/10.1016/j.caeai.2024.100223

Zheng, Y., Chen, D., & Li, Y. (2026). From resistance to expressive drive': How a domain-specific LLM enhances L2 writing outcomes through guided feedback. *Innovation in Language Learning and Teaching*, 1–21. https://doi.org/10.1080/17501229.2026.2687097

Zheng, Y., Tu, F., Shu, F., Shang, C., Chen, L., & Meng, J. (2026). How AI agents transform reflective practices: A three-semester comparative study in socially shared regulation of learning. *Computer Standards & Interfaces*, 97, 104094. https://doi.org/10.1016/j.csi.2025.104094

Zhu, Y., & Li, H. (2025). Enhancing college English education in China with AI: A teacher–AI–student triad model. English Language Teaching, 18(7), 15–30. https://doi.org/10.5539/elt.v18n7p15

Zhuang, Y., Zhao, R., Xie, Z., & Yu, P. L. H. (2025). Enhancing language learning through generative AI feedback on picture-cued writing tasks. Computers and Education: Artificial Intelligence, 9, Article 100450. https://doi.org/10.1016/j.caeai.2025.100450

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. Theory Into Practice, 41(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2

Ziqi, C., Xinhua, Z., Qi, L., & Wei, W. (2024). L2 students' barriers in engaging with form and content-focused AI-generated feedback in revising their compositions. *Computer Assisted Language Learning*, 39(3), 715–735. https://doi.org/10.1080/09588221.2024.2422478

Zou, B., Guan, X., Shao, Y., & Chen, P. (2023). Supporting speaking practice by social network-based interaction in artificial intelligence (AI)-assisted language learning. *Sustainability*, 15(4), Article 2872. https://doi.org/10.3390/su15042872

About the Series

Contemporary Studies on Digital Humanities, edited by Ali Tilbe, Yusuf Topaloğlu and Gökhan Dinar, is an international, peer-reviewed book series devoted to the study of human expression, knowledge and culture in an age when the tools of reading, writing, teaching and interpreting are increasingly computational. Published by Maurer Press (Frankfurt am Main) the series brings together scholars from literary studies, linguistics, language education, translation studies, history, philosophy, cultural and media studies, the arts and the information sciences. It welcomes research on computational approaches to texts and cultures, digital archives and heritage, corpus-based inquiry, artificial intelligence in scholarship, teaching and creative practice, digital pedagogy and assessment, and the ethical and epistemological questions these developments raise. Approaching technology neither as a miracle nor as a threat, but as a new condition of humanistic work, the series seeks to build a plural and rigorous dialogue on the digital transformation of the humanities, with particular attention to research from Türkiye and its neighbouring academic communities alongside contributions from across the globe. It publishes monographs, edited volumes and thematic collections.

About the Editors

Ali Tilbe (Prof. Dr.), Professor of French Language Education and Head of Department at İstanbul University-Cerrahpaşa, Hasan Ali Yücel Faculty of Education. His research focuses on contemporary and Francophone literature, autofiction, sociocriticism and the teaching of French as a foreign language.

Yusuf Topaloğlu (Assoc. Prof. Dr.), Associate Professor of French Language Education at İstanbul University-Cerrahpaşa, Hasan Ali Yücel Faculty of Education. His research focuses on French linguistics and syntax, written production in French as a foreign language and artificial intelligence in language education.

Gökhan Dinar (Asst. Prof. Dr.), Assistant Professor of French Language Education at İstanbul University-Cerrahpaşa, Hasan Ali Yücel Faculty of Education, and the department's Erasmus coordinator. His research focuses on postcolonial Francophone literature, sociocriticism, artificial intelligence in language education and translation studies.

About the Author of the Edited Volume

Tuncer Can is a member of the English Language Teaching Department at Istanbul University-Cerrahpaşa, Türkiye. His research spans technology-mediated language teacher education, three-dimensional virtual learning environments, immersive and game-based language learning, differentiated instruction with mobile technologies, and, most recently, the integration of artificial intelligence in foreign language education. He has coordinated and contributed to international projects on virtual worlds and teacher training, and his work has appeared in journals including the *Turkish Online Journal of Distance Education*, *Journal of Research on Technology in Education*, and *Culture and Education*. "