

# The Role of Artificial Intelligence in Enhancing Foreign Language Learning: Toward Guided Human–AI Language Learning (GHAILL)

Gumarpi Rahis Pasaribu<sup>1\*</sup> Nanda Dwi Astri<sup>2</sup> Nurainun Hasibuan<sup>3</sup> Dara Mubshirah<sup>4</sup> John Bunce<sup>5</sup>

<sup>1</sup>STIT Al-Ittihadiyah Labuhanbatu Utara, Indonesia

<sup>2</sup>Universitas Prima Indonesia, Indonesia

<sup>3</sup>STAI Darul Arafah, Indonesia

<sup>4</sup>UIN Ar-Raniry Banda Aceh, Indonesia

<sup>5</sup>California State University, United States

<sup>1</sup> [gumarpi\_rahis@stit-al-ittihadiyahlabura.ac.id]

\* corresponding author

## ARTICLE INFO

## ABSTRACT

### Article history:

Received: 14 August 2026

Revised: 17 August 2026

Accepted: 21 August 2026

### Keywords:

artificial intelligence  
foreign language learning  
generative AI  
speaking skills  
learner agency

Artificial intelligence (AI) has become a prominent component of contemporary foreign language learning, particularly through generative systems that provide conversational rehearsal, linguistic explanation, feedback, and individualized practice. This article is a conceptual paper rather than an empirical qualitative report or research protocol. It synthesizes the literature cited in the manuscript to examine how AI-supported learning can be understood through five interconnected domains: low-risk speaking rehearsal; personalized feedback and linguistic noticing; learner agency and self-regulated learning; engagement, confidence, and willingness to communicate; and tensions involving accuracy, dependency, privacy, and communicative authenticity. The results of the conceptual synthesis distinguish perceived speaking improvement from objectively measured proficiency and clarify the conceptual relations among AI literacy, agency, self-regulation, confidence, and willingness to communicate. The integrated results and discussion propose Guided Human–AI Language Learning (GHAILL), a human-centred process model in which learners move from goal setting and independent attempt to AI-mediated rehearsal or feedback, critical evaluation, learner revision, repeated production, human communication, and reflection. GHAILL differs from general self-regulated learning and technology-assisted language-learning frameworks by specifying three safeguards for generative-AI-supported language production: learner production before or alongside AI assistance, critical evaluation before uptake, and transfer from machine-mediated practice to human communication. The article argues that AI is most educationally valuable when it expands practice without displacing learner judgment, teacher mediation, or authentic interpersonal communication.

Copyright ©. All rights reserved.

## I. Introduction

Artificial intelligence has moved from a specialist technology to an everyday learning resource in a remarkably short period. In higher education, students now interact with generative AI systems to ask questions, generate examples, clarify concepts, obtain feedback, rehearse presentations, and organize independent study [1–2]. Foreign language education is especially affected because successful language learning requires repeated exposure, interaction, timely feedback, and opportunities to produce language. In many conventional classrooms, however, learners have limited individual speaking time and teachers cannot provide immediate personalized feedback to every student. AI-supported tools potentially extend the learning environment by making language support available whenever learners choose to engage with it.



DOI:

W : <http://ijarlit.org> | E : [info@ijarlit.org](mailto:info@ijarlit.org)

The growing accessibility of generative AI also changes the nature of computer-assisted language learning. Earlier CALL tools commonly relied on fixed exercises, predetermined feedback, or restricted conversational paths. Contemporary systems can respond dynamically to user input, generate customized dialogues, adapt explanations, propose vocabulary, reformulate sentences, simulate communicative situations, and sustain extended exchanges. Systematic work on ChatGPT and AI chatbots in ESL/EFL education indicates substantial research growth around these affordances, while also revealing variation in pedagogical design and learning outcomes [3–4]. Rather than functioning only as a repository of information, AI may therefore become part of the interactional ecology through which learners prepare, practise, monitor, and revise language.

This shift is particularly relevant to speaking. Oral production is cognitively and affectively demanding because learners must retrieve lexical items, organize grammar, control pronunciation and fluency, interpret an interlocutor, and manage social pressure in real time. AI-mediated rehearsal can reduce some of this pressure by allowing repeated practice without immediate human evaluation [5–6]. Recent studies have reported promising relationships between AI-supported activities and speaking performance, enjoyment, confidence, and reduced anxiety, although the effects vary according to task design and platform affordances [7]. Such findings make speaking an important domain for investigating not only whether students use AI, but how they experience it as part of their learning process.

The pedagogical significance of this development extends beyond technological novelty. Foreign language learning is inherently practice-intensive: learners need repeated exposure to comprehensible input, opportunities to retrieve and produce language, feedback on emerging performance, and interaction that gradually becomes more spontaneous. In many university contexts, however, classroom time is limited, student numbers are relatively large, and the amount of individual speaking practice available to each learner is uneven. AI-supported environments potentially address part of this structural limitation by giving students access to language assistance and rehearsal opportunities outside scheduled lessons. From this perspective, AI is relevant not because it can generate fluent text, but because it can enlarge the time, frequency, and variety of language-learning activity available to learners.

The expansion of AI also changes the relationship between formal and informal language learning. Students increasingly move between teacher-designed classroom activities and self-initiated digital practices in which they ask AI to explain vocabulary, simplify difficult input, generate conversation topics, correct sentences, simulate role-plays, or provide alternative expressions. This movement is important because language development does not occur only during formal instruction. Informal digital learning can extend exposure and encourage learners to take greater responsibility for practice, especially when they use AI intentionally to address weaknesses that emerge during classroom learning. Recent work on generative AI-mediated informal digital language learning has therefore emphasized the importance of AI literacy, learner emotion, and willingness to communicate as interconnected dimensions of contemporary language learning [8].

For foreign language speaking, these affordances are especially relevant. Speaking requires learners to coordinate lexical retrieval, grammatical encoding, pronunciation, fluency, listening, turn-taking, and pragmatic choices under real-time pressure. Even students who possess adequate declarative knowledge may struggle to transform that knowledge into spontaneous speech. In addition, speaking frequently carries a strong affective burden: learners may hesitate because they fear making errors, being judged by classmates, losing face, or failing to express an intended meaning. The classic willingness-to-communicate framework highlights how perceived competence, confidence, and situational factors influence whether learners actually enter into second-language communication [9]. AI-mediated rehearsal may alter this decision-making environment by allowing students to practise before they face higher-stakes human interaction.

This does not mean that interaction with AI is equivalent to communication with people [10]. Human conversation involves interpersonal intentions, negotiation of identity, emotion, cultural knowledge, gesture, timing, and unpredictable responses that cannot be reduced to linguistic correctness. Nevertheless, AI may serve as a preparatory layer between individual planning and authentic communication. Learners can generate questions, rehearse possible answers, request reformulations, ask for pronunciation guidance, and repeat an interaction without exhausting a human partner. Recent research has begun to show that carefully designed AI-mediated speaking activities can be associated with improved speaking performance, enjoyment, reduced anxiety, and stronger emotional engagement [7]. The pedagogical issue is therefore not whether AI should replace human interaction, but how AI-supported rehearsal can be connected to subsequent communication with teachers, peers, and wider communities.

Another important contribution of AI lies in the immediacy of feedback. In conventional language classrooms, teachers cannot respond to every learner error at the moment it occurs. Generative AI can provide explanations, alternatives, corrections, examples, and reformulations almost instantly. Studies of AI use in foreign language education report that learners value such accessibility for grammar, vocabulary, writing support, and individualized practice [11–12]. Yet immediate feedback should not automatically be equated with effective feedback. From a language-learning perspective, the critical process is whether learners notice a gap, understand the reason for a change, compare alternatives, and incorporate the revised form into later production.

The distinction between simply receiving a corrected answer and actively learning from correction is central to evaluating the educational value of AI.

The same tension applies to learner autonomy and self-regulation. AI can give students considerable control over timing, content, pace, difficulty, and repetition. A learner can ask for easier explanations, request more challenging questions, create a personalized vocabulary test, rehearse a presentation several times, or focus on a recurring grammatical problem. These possibilities align with broader understandings of self-regulated learning in which learners set goals, select strategies, monitor performance, and adapt subsequent actions [13]. Recent research on generative AI-supported self-regulated language learning similarly indicates that productive use depends on learners' awareness, evaluative capacity, and interactions with teachers, peers, and AI systems [14]. Thus, access to AI may create conditions for autonomy, but it does not guarantee autonomous learning.

Indeed, the convenience of generative AI introduces a paradox [15]. The same system that can support reflection can also remove the productive struggle through which learning occurs. Students may ask AI to formulate complete responses before attempting their own, accept corrections without understanding them, or depend on generated scripts during speaking tasks. Such practices can create an appearance of improved performance while masking limited internalization. Research on university students' reliance on generative AI has therefore distinguished between appropriate reliance, over-reliance, and resistance, emphasizing that learners need to calibrate trust rather than either accept or reject AI output automatically [16]. This issue is particularly important in foreign language learning because learners with lower proficiency may have less linguistic knowledge available for detecting inaccurate, unnatural, or pragmatically inappropriate output.

Critical AI literacy consequently becomes part of language-learning competence. Effective users need to know not only how to prompt an AI system but also how to evaluate what it produces, compare alternative expressions, verify uncertain information, recognize limitations, and decide when human feedback is preferable. Emerging EFL research suggests that explicit pedagogical intervention can strengthen students' critical engagement with generative AI rather than allowing them to treat fluent output as inherently reliable [17–18]. In this sense, responsible AI-assisted language learning involves a double learning process: students learn the foreign language while simultaneously learning how to manage the strengths and weaknesses of an artificial conversational partner.

The university context provides an especially useful setting for examining these processes because university learners are expected to become increasingly independent while operating within assessment systems, institutional expectations, and uneven levels of digital and linguistic competence. AI may therefore function differently across learners and tasks: for some, it can support independent speaking practice and feedback; for others, it may become a shortcut that reduces meaningful language production. This variation underscores the need for a conceptual account that distinguishes access to AI from the quality of learning activity.

Against this background, the present article treats AI-supported foreign language learning as a human–AI collaborative process rather than a technology-driven outcome. The learner remains the principal agent who sets goals, interprets feedback, decides whether to trust an output, modifies language, and transfers practice into real communication. The teacher remains central as a designer of meaningful tasks, a source of expert judgment, and a guide for critical AI literacy. This positioning allows the article to integrate enabling and constraining dimensions of AI while avoiding two simplistic assumptions: that AI will automatically improve language learning, or that AI use necessarily undermines authentic learning.

Despite rapid publication growth, several conceptual gaps remain. First, much of the early literature emphasizes attitudes toward AI, perceived usefulness, intention to use, or general acceptance. These studies are valuable for understanding adoption, but positive attitudes do not necessarily demonstrate meaningful learning. A learner may consider AI convenient while using it mainly to obtain finished answers; conversely, a learner who questions AI output, compares alternatives, and reformulates language may remain cognitively and linguistically active. The key distinction is therefore between AI use and AI-mediated learning.

Second, experimental, survey, and mixed-method research can identify patterns of improvement or perception, but these findings are often difficult to integrate unless the underlying learner processes are specified. Research on reliance and resistance already shows that trust in generative AI can take productive or problematic forms [16]. A conceptual synthesis is therefore needed to connect observable outcomes with mechanisms such as independent attempt, feedback uptake, verification, revision, repeated production, and transfer to human communication.

Third, the relationship between AI literacy, learner agency, self-regulation, confidence, and willingness to communicate requires sharper conceptual differentiation. AI can expand learner choice by increasing control over topic selection, pacing, repetition, feedback requests, and task difficulty, but it can also weaken agency when learners transfer decision-making and language production to the system. Recent research on AI-supported self-regulated language learning emphasizes the importance of evaluation, AI literacy, and interaction with teachers and peers [14]. The present article therefore clarifies these constructs before integrating them into GHAILL.

Accordingly, this article develops a conceptual account of AI-supported foreign language learning with particular attention to speaking-related practice. Rather than treating AI use as a direct cause of linguistic improvement, the paper asks how particular learner processes may convert AI affordances into meaningful language-learning activity. It therefore distinguishes technological access from learning quality and treats claims about speaking improvement cautiously unless they are supported by independent performance measures. The article also clarifies the relationship among AI literacy, learner agency, self-regulated learning, confidence, and willingness to communicate before integrating these constructs into the proposed GHAILL model.

RQ1. What pedagogically relevant affordances and constraints of artificial intelligence are identified in current foreign language learning literature?

RQ2. Through what mechanisms can AI-supported practice plausibly contribute to speaking preparation, feedback uptake, confidence, engagement, and willingness to communicate?

RQ3. How should AI literacy, learner agency, self-regulated learning, confidence, and willingness to communicate be conceptually differentiated and related?

RQ4. What distinctive contribution does the Guided Human–AI Language Learning (GHAILL) model make to existing accounts of self-regulated and technology-assisted language learning?

## II. Method

### A. *Research Design and Scope*

This article adopts a conceptual synthesis rather than an empirical qualitative design. The purpose is to integrate arguments and findings from the peer-reviewed literature cited in the manuscript into a coherent process account of AI-supported foreign language learning. The paper does not report newly collected participant data, interviews, test scores, or classroom observations; consequently, no claim is made that the present article independently demonstrates improvement in speaking proficiency.

The analysis is human-centred and process-oriented. AI is treated as a set of affordances—including rapid feedback, dialogue simulation, reformulation, personalization, and on-demand practice—whose educational consequences depend on learner activity, task design, evaluative judgment, and the connection between machine-mediated practice and human communication [3–4, 10–16]. This orientation avoids a simple technology-effects assumption in which access to AI is treated as equivalent to learning.

A central analytical boundary concerns the difference between perceived improvement and objectively measured development. Research based on learner perceptions can illuminate confidence, strategy use, engagement, perceived progress, or attitudes, but these outcomes should not be interpreted as equivalent to standardized gains in pronunciation, fluency, accuracy, or overall proficiency. Where the cited literature includes experimental or mixed-method evidence, those findings are treated separately from perception-based claims [5–7, 11].

The conceptual focus is foreign language learning in university and EFL-related contexts because this is the dominant setting represented in the literature reviewed here. The paper nevertheless formulates GHAILL at a sufficiently general level to be tested in other languages, proficiency levels, age groups, and instructional environments. Such extension remains an empirical question rather than an assumption of the model.

### B. *Literature Sources and Selection*

The source base consists of the scholarly works already cited in this manuscript, including systematic reviews, meta-analytic or mixed-method studies, empirical investigations of AI-assisted language learning, work on self-regulated learning and willingness to communicate, and foundational accounts of noticing, self-efficacy, and scaffolding [3–25]. The purpose is integrative rather than exhaustive: the article does not claim the search coverage, screening procedure, or reproducibility standards of a systematic review.

Sources were used according to their evidential role. Review and meta-synthesis studies were used to identify recurring affordances and challenges across AI-supported language learning [3–4, 10, 18]. Empirical studies informed claims about speaking practice, feedback, motivation, reliance, and AI literacy [5–8, 11–17, 20, 22]. Foundational theoretical sources were used to define self-regulation, willingness to communicate, noticing, self-efficacy, and scaffolding [9, 13, 19, 21, 25].

Because AI platforms, model capabilities, and institutional policies change rapidly, the synthesis emphasizes mechanisms that are more stable than any single product feature. These mechanisms include learner attempt, feedback uptake, verification, revision, repeated production, self-monitoring, and transfer to human interaction. Product-specific claims are therefore avoided unless they are central to the cited study.

Table 1. Conceptual distinctions used in the article

| Construct                  | Working definition                                                             | Role in the process                                                            | Key anchors    |
|----------------------------|--------------------------------------------------------------------------------|--------------------------------------------------------------------------------|----------------|
| AI literacy                | Capacity to understand, evaluate, verify, and selectively use AI output        | Enables informed judgment and calibrated reliance                              | [8, 14, 16–18] |
| Learner agency             | Intentional control over learning choices and the use or refusal of AI support | Determines who remains the decision-maker                                      | [14, 16]       |
| Self-regulated learning    | Cyclical goal setting, strategy use, monitoring, and reflection                | Operationalizes agency across learning episodes                                | [13–14]        |
| Confidence                 | Affective appraisal of perceived competence                                    | May increase through successful rehearsal but is not proof of proficiency gain | [21–22]        |
| Willingness to communicate | Situated readiness to initiate or sustain communication with other people      | Represents transfer from preparation to interpersonal language use             | [8–9]          |

### C. Analytical Procedure

The literature was conceptually organized around recurrent mechanisms rather than around individual AI platforms. The organizational procedure borrows the general logic of iterative categorization and theme refinement from thematic analysis scholarship [23–24], but it is applied here to published concepts and findings rather than to participant transcripts. Five domains were retained because they capture both enabling and constraining dimensions of AI-supported language learning: speaking rehearsal; feedback and noticing; agency and self-regulation; engagement, confidence, and willingness to communicate; and accuracy, dependency, privacy, and authenticity. These domains are treated as analytically distinct but dynamically connected.

Within each domain, the analysis asks two questions. First, what learner activity is required for the AI affordance to become educationally meaningful? Second, what form of misuse or bypass could transform the same affordance into superficial learning or dependence? For example, immediate correction can support noticing when learners compare and reuse forms, but it can reduce learning when it simply replaces independent formulation.

The cross-domain synthesis then traces a process from learner intention to subsequent human communication. This process became the basis of GHAILL: learner goal -> independent attempt -> AI-mediated rehearsal or feedback -> critical evaluation -> learner revision -> repeated production -> human communication -> reflection. The sequence is proposed as a conceptual heuristic, not as a validated causal model.

The model is therefore evaluated in this article by conceptual coherence and explanatory usefulness rather than by participant coding frequencies or statistical effects. Its empirical validity remains open to future qualitative, experimental, longitudinal, and classroom-based testing.

### D. Conceptual Boundaries and Ethical Position

The article does not use human participants, identifiable personal data, newly collected interviews, or private AI interaction logs. Institutional ethical approval for human-subject data collection is therefore not applicable to the present conceptual paper. This statement should not be generalized to future empirical tests of GHAILL, which would require informed consent, confidentiality protections, secure data handling, and relevant institutional review procedures.

Responsible AI use nevertheless remains part of the conceptual argument. Language-learning activities may involve personal information, assessment-related work, or sensitive interaction histories. Learners and institutions should therefore consider privacy, data governance, attribution, academic integrity, and the possibility that model output may be inaccurate or pragmatically inappropriate [15–18].

A second boundary concerns causal language. The synthesis can identify plausible mechanisms and relationships supported across prior studies, but it cannot establish that GHAILL causes proficiency gains. Likewise, evidence of confidence or willingness to communicate should not be collapsed into evidence of linguistic development. These outcomes occupy different levels of the learning process.

A third boundary concerns technological heterogeneity. Text-based chatbots, voice-enabled systems, automated speech evaluation tools, and general-purpose generative AI differ in affordances and limitations [5–7, 10]. GHAILL therefore specifies learner actions rather than assuming that one platform or interface is pedagogically representative of all AI systems.

Finally, the analysis treats human communication as a necessary reference point for communicative language learning. AI can simulate many interactional features, but it does not fully reproduce interpersonal stakes, identity negotiation, cultural expectations, embodied cues, or genuine social consequences [10]. The model consequently positions AI as complementary scaffolding rather than as a substitute for teachers, peers, or other human interlocutors.

These boundaries directly address the risk of overstating what AI-supported language learning can demonstrate. The article's claims are therefore conceptual and conditional: AI can create additional opportunities for meaningful practice, but the educational value of those opportunities depends on what learners do before, during, and after AI assistance.

Table 2. Conceptual synthesis framework

| Domain                                 | Central mechanism                                                             | Boundary / risk                                                                                     |
|----------------------------------------|-------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------|
| Speaking rehearsal                     | Repeated low-risk production before socially exposed speaking                 | Script dependence or no transfer to human interaction                                               |
| Feedback and noticing                  | Comparison, explanation, revision, and reuse of feedback                      | One-step correction and copying                                                                     |
| Agency and self-regulation             | Goal setting, selective support, monitoring, and withdrawal of support        | Delegation of linguistic decisions to AI                                                            |
| Engagement, confidence, and WTC        | Lower entry barrier to practice and increased readiness for interpersonal use | Confidence confined to the AI environment                                                           |
| Accuracy, dependency, and authenticity | Verification, calibrated trust, and complementary use                         | Hallucination, over-reliance, privacy risk, and simulated interaction replacing human communication |

### III. Results and Discussion

The literature can be synthesized into five interrelated conceptual domains: (1) AI as a low-risk space for speaking rehearsal; (2) personalized feedback and linguistic noticing; (3) learner agency and self-regulated practice; (4) engagement, confidence, and willingness to communicate; and (5) tensions involving accuracy, dependency, privacy, and communicative authenticity. These domains are not independent lists of benefits and risks. Together they describe a conditional pathway in which the value of AI depends on learner production, critical evaluation, revision, and eventual transfer to human communication.

Table 3. Conceptual domains and literature anchors

| Conceptual domain                                | Analytic focus                                                                 | Literature anchors |
|--------------------------------------------------|--------------------------------------------------------------------------------|--------------------|
| Domain 1: Low-risk speaking rehearsal            | Repetition, preparation, reduced evaluative pressure, transfer                 | [5–7, 10]          |
| Domain 2: Feedback and noticing                  | Immediate support, comparison, explanation, revision, reuse                    | [3, 11–12, 19–20]  |
| Domain 3: Agency and self-regulation             | Goal setting, selective assistance, monitoring, AI literacy                    | [13–14, 16]        |
| Domain 4: Engagement, confidence, and WTC        | Practice initiation, persistence, perceived competence, interpersonal transfer | [8–9, 21–22]       |
| Domain 5: Accuracy, dependency, and authenticity | Verification, calibrated reliance, privacy, limits of simulation               | [10, 15–18]        |

#### A. AI as a Low-Risk Space for Speaking Rehearsal

A recurring proposition in the literature is that AI can function as a rehearsal environment before learners perform language in socially exposed situations. Conventional classrooms cannot always provide sustained individual speaking time, whereas AI-supported interaction can allow learners to repeat a task, modify an answer, request alternative formulations, and restart a conversation without consuming classmates' speaking opportunities [5–7]. The value of this affordance lies in additional production, not in the mere presence of a chatbot.

Rehearsal can support both task-specific preparation—such as oral presentations, interviews, classroom discussions, or role-plays—and more open-ended conversational practice. Learners can pause, repeat, request simpler wording, compare registers, and rehearse the same response several times. These options may be especially useful for learners who need more planning time before spontaneous production.

The low-risk quality of AI-mediated practice is also affective. Foreign language speaking can involve fear of error, hesitation, pronunciation difficulty, and negative evaluation. AI may reduce the immediate social cost of unsuccessful attempts and thereby encourage experimentation. However, reduced pressure is educationally useful only when it leads to more active language production rather than to passive dependence on generated scripts.

This distinction separates productive rehearsal from cognitive offloading. Productive rehearsal requires retrieval, formulation, repetition, and reformulation by the learner. In contrast, reliance on complete AI-generated scripts can create fluent performance without equivalent internalization. The relevant pedagogical question is therefore whether AI expands the learner's own production or substitutes for it.

The endpoint of rehearsal should be transfer to human communication. Successful interaction with a predictable system is not equivalent to negotiating meaning with people, whose responses involve social intentions, identity, emotion, cultural expectations, and unpredictability [10]. AI is therefore best conceptualized as a bridge between private preparation and public language use rather than as the final communicative environment.

### ***B. Personalized Feedback and Linguistic Noticing***

A second domain concerns the immediacy and personalization of AI-generated feedback. Learners can request explanations, alternatives, grammar correction, vocabulary support, reformulation, and pronunciation-related guidance at the point of difficulty [3, 11–12]. This immediacy can reduce the delay between noticing a problem and receiving assistance, extending feedback beyond scheduled classroom time.

Immediate feedback, however, is not synonymous with learning. The key distinction is between feedback as a product and feedback as a process. When learners copy a revised sentence without comparing it with their own production, cognitive engagement is limited. When they inspect differences, ask why a form is preferable, test alternatives, or reuse the corrected feature, feedback becomes a resource for noticing and learning [19–20].

Generative systems also support iterative feedback because learners can reshape the response through follow-up prompts. They can ask whether an expression sounds natural, whether it fits an academic register, how tense or modality changes meaning, or what alternatives are available. This capacity makes the quality of prompting and evaluation central to the learning process rather than peripheral technical skills.

The strongest conceptual learning cycle is therefore input → noticing → revision → reuse. AI contributes at the feedback stage, but the learner must appropriate the feedback through evaluation and subsequent production. The decisive question is not whether AI can correct a sentence, but whether the learner becomes more capable of producing and judging the language independently.

### ***C. Learner Agency and Self-Regulated Practice***

A third domain concerns learner agency. AI potentially expands control over timing, topic, pacing, repetition, and task difficulty. Learners can initiate practice without waiting for teacher direction and can focus on immediate communicative needs. Yet access to choice should not be equated with agency, because a learner may choose to use AI while simultaneously delegating the central linguistic work to the system.

Agency is better defined here as intentional control over the learning process: setting a goal, deciding what assistance is needed, regulating the amount of support, evaluating the adequacy of the response, and knowing when to continue without AI. This definition distinguishes technological choice from learning control and directly addresses the possibility that convenience can coexist with reduced learner responsibility.

Self-regulated learning describes how that agency is enacted across time. Planning, strategy selection, monitoring, and reflection can all be supported by AI [13–14]. Learners may use AI to diagnose recurring weaknesses, generate practice questions, vary task difficulty, or review errors. These functions are educationally meaningful when they help learners organize their own learning rather than merely complete isolated assignments.

AI literacy is the enabling competence that makes informed agency more likely. Evaluation, verification, prompt refinement, awareness of uncertainty, and selective reliance help learners remain decision-makers when interacting with fluent machine output [8, 14, 16–18]. AI literacy, agency, and self-regulation are therefore related but not interchangeable: literacy supports judgment, agency concerns control, and self-regulation describes the cyclical management of learning.

**D. Engagement, Confidence, and Willingness to Communicate**

A fourth domain concerns engagement. AI can lower the practical threshold for initiating practice because learners can begin a short conversation, request a quiz, generate a speaking topic, or ask for targeted feedback immediately. This accessibility may increase frequency and persistence, but interaction time alone is an insufficient indicator of learning. Engagement is more meaningfully reflected in active production, sustained effort, reflection, and transfer.

Confidence is conceptually distinct from engagement and proficiency. Repeated low-pressure rehearsal may strengthen perceived competence by giving learners successful practice experiences [21–22]. Yet confidence based on reading or memorizing generated language may not reflect independent linguistic ability. The value of confidence therefore depends on whether learners can perform without the same level of AI support.

Willingness to communicate (WTC) provides a further transfer criterion. WTC refers to readiness to enter interpersonal communication rather than comfort with a machine interface [9]. AI-supported rehearsal is most educationally significant when it helps learners participate more readily with classmates, teachers, or other human interlocutors. Confidence inside AI interaction should not automatically be treated as generalized communicative confidence.

**E. Tensions of Accuracy, Dependency, and Authenticity**

The fifth domain captures the tensions accompanying AI convenience. Generative systems can produce fluent language that is incomplete, inaccurate, factually unreliable, or pragmatically inappropriate. This is especially problematic for learners with limited proficiency because polished output can be difficult to evaluate. Verification through teachers, dictionaries, corpora, or other authoritative resources should therefore be treated as a core language-learning and AI-literacy practice [16–18].

Dependency is best understood through the sequence of learner action. An attempt → AI support → evaluation → revision sequence preserves the learner's generative role, whereas an AI generation → copy → performance sequence transfers substantial cognitive and linguistic work to the system. The technology may be identical, but the learning process is not.

Privacy and data governance constitute additional concerns because language-learning interactions may include personal details, academic work, or sensitive content. Institutional guidance should therefore address what information learners should not enter into public systems, how generated content should be attributed, and when AI use conflicts with assessment rules [15–18].

Authenticity forms another boundary. AI can simulate role-plays and maintain dialogue, but human interaction includes interpersonal intentions, identity, humor, cultural expectations, gesture, timing, interruption, repair, and genuine emotional consequences [10]. A communicative language curriculum should therefore avoid creating a closed ecosystem in which machine interaction replaces peer and teacher communication.

Across the five domains, the same affordance can support or weaken learning. Immediate feedback can promote noticing or encourage copying; personalization can strengthen agency or deepen dependence; and low-risk rehearsal can prepare learners for communication or become an alternative to it. This conditionality is the central conceptual insight of the synthesis.

The mechanism linking these domains is learner activity between AI output and subsequent language use. AI becomes pedagogically productive when learners remain cognitively and linguistically active—attempting, evaluating, revising, reproducing, and communicating—rather than when the system simply generates an acceptable final product.

This synthesis is represented in GHAILL as the following cycle: learner goal → independent attempt → AI-mediated rehearsal or feedback → critical evaluation → learner revision → repeated production → human communication → reflection. The sequence is not a rigid classroom procedure; it is a conceptual statement that AI should extend, rather than replace, the cognitive and communicative work of the learner.

**F. Guided Human–AI Language Learning (GHAILL): Integrated Interpretation**

The fourth research question concerns the distinctive contribution of Guided Human–AI Language Learning (GHAILL). GHAILL conceptualizes productive AI-assisted language learning as an eight-stage cycle: (1) learner goal, (2) independent attempt, (3) AI-mediated rehearsal or feedback, (4) critical evaluation, (5) learner revision, (6) repeated production, (7) human communication, and (8) reflection. The cycle then informs the next goal. The learner remains the principal agent, while AI functions as contingent scaffolding.

GHAILL is not intended to replace general theories of self-regulated learning. Zimmerman's account [13] explains broad cycles of planning, performance, monitoring, and reflection across learning contexts. GHAILL adds a more specific sequencing logic for generative-AI-assisted language production: the learner should attempt language before or alongside AI assistance, critically evaluate machine output before adopting it, and reconnect

AI-mediated practice to human communication. These three checkpoints respond to risks that are particularly salient when a generative system can produce a fluent answer on the learner's behalf.

GHAILL also differs from broad technology-assisted language-learning frameworks because it does not classify tools or assume that technological exposure is the explanatory unit. The model focuses on what the learner does with the tool and on where pedagogical breakdown can occur. If independent attempt is omitted, retrieval and formulation may be outsourced; if critical evaluation is omitted, inaccurate output may be accepted; if revision and repeated production are omitted, feedback becomes editing; and if human communication is omitted, practice may remain within a technologically comfortable but socially limited environment.

This process orientation explains why simple comparisons between AI users and non-users can be misleading. The category of AI use contains qualitatively different practices. Two learners may use the same system for the same task, yet one may engage in iterative rehearsal and critical revision while the other copies generated language. Their exposure is similar, but their learning activity is fundamentally different. GHAILL therefore proposes interaction quality and sequencing as more informative variables for future research.

RQ1 is answered by the identification of five interconnected domains that combine affordances and constraints. AI can extend rehearsal, feedback, personalization, and independent practice, while accuracy, dependency, privacy, and authenticity remain persistent boundaries.

RQ2 is answered by the rehearsal-to-transfer and feedback-uptake mechanisms. AI is most plausibly linked to speaking preparation when learners produce language repeatedly, evaluate feedback, revise their output, and carry the resulting language or confidence into human interaction. These mechanisms should not be confused with direct proof of proficiency gain.

RQ3 is answered by treating AI literacy as an enabling competence, agency as intentional control, self-regulation as cyclical management, confidence as perceived competence, and WTC as readiness for interpersonal communication. This conceptual differentiation prevents overlap while preserving their functional relationships within the learning process.

RQ4 is answered through GHAILL's three AI-specific safeguards: independent learner production, critical evaluation of machine output, and transfer to human communication. The model therefore extends existing learning theories by specifying where generative AI should enter, where learner judgment must intervene, and where technology-supported practice should reconnect with authentic language use.

### **G. Theoretical Contributions**

The first theoretical contribution is a shift from a technology-effects logic to a learner-mediated process account. The central question becomes not simply whether AI improves learning, but which sequences of learner action make improvement more plausible. AI affordances are treated as opportunities rather than outcomes, and the explanatory unit is the learner's pattern of attempt, assistance, evaluation, revision, production, and transfer.

This process orientation helps explain inconsistent findings across AI-assisted language studies. Differences in task design, proficiency, prompting behaviour, teacher guidance, platform affordances, and evaluation practices alter what learners actually do with AI. Future studies can therefore test not only frequency of AI use but also interaction quality, support sequencing, verification behaviour, and transfer to independent performance.

The second contribution is a differentiated account of learner agency. Access and autonomy are not synonymous. Agency is conceptualized as calibrated control: learners establish goals, decide when assistance is necessary, judge response quality, determine whether further practice is needed, and know when to stop using the tool. Under this definition, unlimited access to AI may coexist with weak agency if the system performs most of the cognitive and linguistic work.

AI literacy becomes theoretically central because verification, prompt refinement, awareness of uncertainty, and selective reliance are mechanisms through which learners preserve control [8, 14, 16–18]. Self-regulation [13–14] describes how that control is enacted over time. This arrangement clarifies the constructs rather than collapsing them into a broad notion of digital autonomy.

The third contribution concerns speaking development. AI-mediated speaking practice is conceptualized as a rehearsal-to-transfer process rather than as an autonomous form of communicative competence. Conversational systems can extend practice time and reduce the social cost of mistakes, but fluent interaction with a machine does not itself demonstrate the ability to negotiate meaning with people [10].

Accordingly, the strongest evidence of speaking enhancement in future tests of GHAILL should be independent human performance. Relevant outcomes may include more sustained turns, greater willingness to initiate interaction, improved recovery from breakdown, or verified changes in fluency, pronunciation, and accuracy. This recommendation preserves the distinction between perceived confidence and objectively measured language development.

The fourth contribution is GHAILL itself. Its novelty lies not in proposing that learners should self-regulate or that technology can scaffold learning—both ideas predate generative AI—but in specifying three checkpoints that address generative AI's capacity to produce complete language on demand: independent attempt, critical evaluation, and human-communication transfer. These checkpoints define when AI assists the learning process without becoming the origin and endpoint of performance.

GHAILL also identifies explicit breakdown points that can be tested empirically. Removing the independent attempt may reduce retrieval and formulation; removing critical evaluation may encourage uncritical uptake; removing learner revision and repeated production may convert feedback into simple editing; and removing human communication may confine practice to simulated interaction. This structure makes the model suitable for future qualitative, experimental, and longitudinal validation.

#### ***H. Pedagogical and Institutional Implications***

The conceptual synthesis has direct implications for language teachers. Rather than treating AI as either prohibited assistance or unrestricted convenience, educators can structure learning around a sequence such as individual attempt -> AI-supported rehearsal or feedback -> learner evaluation and revision -> peer interaction -> classroom performance -> reflection. This sequence operationalizes the GHAILL principle that AI should support, not replace, learner production.

Speaking instruction can benefit from task-specific AI use. Students can ask an AI system to generate interview questions, simulate a service encounter, challenge an argument, or provide follow-up questions after a short oral response. The important requirement is that the learner must speak or formulate the response rather than simply read a generated script. Teachers can also require students to identify one AI suggestion they rejected and explain the linguistic or pragmatic reason. Such tasks make critical AI literacy visible and assessable.

Institutions should complement classroom practice with clear guidance on privacy, academic integrity, attribution, and verification. AI literacy workshops should teach students how to protect personal data, recognize hallucinated or unreliable information, compare sources, use dictionaries and corpora, and understand when expert human feedback is necessary. Policies should differentiate legitimate learning support from outsourcing assessed work. Blanket rules are unlikely to capture the diversity of AI-mediated language tasks.

Teacher development is equally important. Educators need opportunities to experiment with AI, evaluate language output, design prompts, and understand how platform features influence learning. The goal is not to turn teachers into AI technicians but to strengthen pedagogical judgment in an AI-rich environment. Teachers are uniquely positioned to connect machine-mediated practice with curriculum goals, disciplinary expectations, cultural context, and interpersonal communication.

A practical implication of the model is that AI-supported tasks should be sequenced around learner production rather than around answer generation. A useful principle is attempt first, assistance second, revision third, and communication afterward. For example, students may first record a short response without AI, then request feedback on clarity or grammatical accuracy, revise the response in their own words, and finally perform the task with a peer. This sequence makes the learner's contribution visible and reduces the likelihood that a polished AI-generated script will substitute for language development.

Teacher mediation remains essential because students differ in proficiency, AI literacy, and capacity to evaluate fluent machine output. Teachers can model how to ask for explanations rather than answers, compare AI suggestions with dictionaries or corpora, identify pragmatic problems, and explain why one formulation is more appropriate than another. Assessment can also be redesigned to reward evidence of process. Short reflections, revision rationales, oral follow-up questions, or comparison of an initial and revised response can reveal whether students understand the language they submit and whether AI support has contributed to learning rather than simple task completion.

At institutional level, responsible integration requires guidance that is more nuanced than blanket permission or prohibition. Students need explicit instruction on privacy, data protection, academic integrity, attribution, hallucinated information, verification, and the limits of AI-mediated communication. Policies should also recognize legitimate pedagogical uses such as rehearsal, formative feedback, vocabulary exploration, and individualized practice. Clear distinctions between acceptable support and unacceptable outsourcing connect ethical AI use to learning purpose rather than to the mere presence of a particular tool.

#### ***I. Limitations and Future Research***

This article has several limitations. First, it is a conceptual synthesis rather than a systematic review or empirical study; therefore, the literature base is selective and the proposed relationships should not be interpreted as pooled effects. Second, the studies cited vary in participant selection, target language, proficiency, institutional context, and research design, which limits direct generalization across learner populations. Third, AI platforms differ substantially in interface, voice capability, feedback quality, privacy conditions, and model behaviour, so findings associated with one system may not transfer to another. Fourth, the literature includes many self-report and perception-based outcomes, making it essential to distinguish perceived speaking

improvement from objectively measured development. Fifth, generative AI evolves rapidly, which can make platform-specific conclusions unstable. Finally, GHAILL has not yet been empirically validated as an integrated model. Future research should test its stages and breakdown points through longitudinal speaking assessments, classroom observations, interaction logs, qualitative interviews, and experimental comparisons across proficiency levels, languages, and levels of teacher guidance.

#### IV. Conclusion

Artificial intelligence is reshaping foreign language learning by expanding access to personalized feedback, conversational rehearsal, and independent practice. The conceptual synthesis developed in this article shows that these affordances do not constitute learning outcomes by themselves. Their educational value depends on learner production, critical evaluation, revision, self-regulation, and transfer to human communication. By distinguishing AI literacy, agency, self-regulation, confidence, and willingness to communicate, the paper provides a clearer account of how these constructs can interact without being conflated. GHAILL extends this account through an eight-stage human-centred cycle and three generative-AI-specific safeguards: independent attempt, critical evaluation, and human-communication transfer. The model should be treated as a testable conceptual framework rather than as an empirically validated causal theory. Its central proposition is that AI can enhance language learning most plausibly when it expands opportunities for practice while learners retain judgment, teachers retain pedagogical responsibility, and authentic interpersonal communication remains the ultimate reference point.

#### References

- [1] Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, 54(2), 537–550. <https://doi.org/10.1177/00336882231162868>
- [2] Moorhouse, B. L., Wong, K. M., & Li, L. (2023). Teaching with technology in the post-pandemic digital age: Technological normalisation and AI-induced disruptions. *RELC Journal*, 54(2), 311–320. <https://doi.org/10.1177/00336882231176929>
- [3] Lo, C. K., Yu, P. L. H., Xu, S., Ng, D. T. K., & Jong, M. S. Y. (2024). Exploring the application of ChatGPT in ESL/EFL education and related research issues: A systematic review of empirical studies. *Smart Learning Environments*, 11, 50. <https://doi.org/10.1186/s40561-024-00342-5>
- [4] Koç, F. Ş., & Savaş, P. (2025). The use of artificially intelligent chatbots in English language learning: A systematic meta-synthesis study of articles published between 2010 and 2024. *ReCALL*, 37(1), 4–21. <https://doi.org/10.1017/S0958344024000168>
- [5] Zou, B., Du, Y., Wang, Z., Chen, J., & Zhang, W. (2023). An investigation into artificial intelligence speech evaluation programs with automatic feedback for developing EFL learners' speaking skills. *SAGE Open*, 13(3), 1–9. <https://doi.org/10.1177/21582440231193818>
- [6] Wu, X., & Li, R. (2026). Unraveling effects of AI chatbots on EFL learners' language skill development: A meta-analysis. *The Asia-Pacific Education Researcher*, 35, 703–714. <https://doi.org/10.1007/s40299-024-00853-2>
- [7] Yan, H., & Singh, M. K. S. (2026). The impact of AI-mediated instruction on speaking proficiency, enjoyment, anxiety, and emotional engagement: A mixed-methods approach. *Humanities and Social Sciences Communications*, 13, 568. <https://doi.org/10.1057/s41599-026-06705-2>
- [8] Zou, M., Reinders, H., & Amjad, F. (2026). Understanding the potential role of GenAI-mediated informal digital learning of English (GenAI-IDLE) in the Global South: AI literacy, emotions, and willingness to communicate as outcomes. *ReCALL*, 38(1), 131–149. <https://doi.org/10.1017/S0958344025100360>
- [9] MacIntyre, P. D., Clément, R., Dörnyei, Z., & Noels, K. A. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *The Modern Language Journal*, 82(4), 545–562.
- [10] Jeon, J., & Lee, S. (2024). Can learners benefit from chatbots instead of humans? A systematic review of human-chatbot comparison research in language education. *Education and Information Technologies*, 29, 23329–23360. <https://doi.org/10.1007/s10639-024-12725-9>
- [11] Karataş, F., Abedi, F. Y., Gunyel, F. O., Karadeniz, D., & Kuzgun, Y. (2024). Incorporating AI in foreign language education: An investigation into ChatGPT's effect on foreign language learners. *Education and Information Technologies*, 29, 19343–19366. <https://doi.org/10.1007/s10639-024-12574-6>
- [12] Slamet, J. (2024). Potential of ChatGPT as a digital language learning assistant: EFL teachers' and students' perceptions. *Discover Artificial Intelligence*, 4, 46. <https://doi.org/10.1007/s44163-024-00143-2>
- [13] Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). Academic Press.
- [14] Wang, X., & Zhang, L. J. (2026). Fostering generative AI-supported self-regulated learning (GenAI-SRL) in informal digital language learning through literacy and interactions: A two-stage PLS-SEM-ANN approach. *ReCALL*, 38(3), 382–398. <https://doi.org/10.1017/S0958344026100524>
- [15] Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10, 15. <https://doi.org/10.1186/s40561-023-00237-x>

- [16] Wang, F., Li, N., Cheung, A. C. K., & Wong, G. K. W. (2025). In GenAI we trust: An investigation of university students' reliance on and resistance to generative AI in language learning. *International Journal of Educational Technology in Higher Education*, 22, 59. <https://doi.org/10.1186/s41239-025-00547-9>
- [17] Cong-Lem, N., Nguyen, T. T., Nguyen, K. N. H., et al. (2026). Critical thinking in Vietnamese EFL students' use of generative AI tools for academic learning: Qualitative insights from a brief intervention. *English Teaching & Learning*, 50, 295–315. <https://doi.org/10.1007/s42321-025-00217-z>
- [18] Liu, J., Sihes, A. J. B., & Lu, Y. (2025). How do generative artificial intelligence tools and large language models influence language learners' critical thinking in EFL education? A systematic review. *Smart Learning Environments*, 12, 48. <https://doi.org/10.1186/s40561-025-00406-0>
- [19] Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158.
- [20] Rahimi, A. R., & Mosalli, Z. (2025). Language learners' surface, deep, and organizing approaches to ChatGPT-assisted language learning: What contextual, individual, and ChatGPT-related factors contribute? *Smart Learning Environments*, 12, 16. <https://doi.org/10.1186/s40561-025-00368-3>
- [21] Bandura, A. (1997). Self-efficacy: The exercise of control. W. H. Freeman.
- [22] Huang, J., & Mizumoto, A. (2025). The effects of generative AI usage in EFL classrooms on the L2 motivational self system. *Education and Information Technologies*, 30, 6435–6454. <https://doi.org/10.1007/s10639-024-13071-6>
- [23] Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
- [24] Braun, V., & Clarke, V. (2021). Thematic analysis: A practical guide. SAGE.
- [25] Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- [26] Pasaribu, G. R., Arfianty, R., & Mubshirah, D. (2024). Integrasi kecerdasan buatan (artificial intelligence) pada pembelajaran bahasa. *Educandumedia: Jurnal Ilmu Pendidikan dan Kependidikan*, 3(2), 21–38. <https://doi.org/10.61721/educandumedia.v3i2.511>
- [27] Anh, N. H., Baharloo, A., & Pasaribu, G. R. (2025). Integrating AI-based tools to enhance speaking fluency in EFL classrooms in Labuhanbatu Utara. *JELIM: Journal of Education, Language, Social and Management*, 3(1), 7–16. <https://doi.org/10.65787/jelim.v3i1.588>
- [28] Zhang, Q., Saleem, A., & Pasaribu, G. R. (2025). The role of deep learning in improving English language acquisition. *PEBSAS: Jurnal Pendidikan Bahasa dan Sastra*, 3(2), 39–48. <https://doi.org/10.61721/pebsas.v3i2.601>
- [29] Zuwanda, M. E., & Pasaribu, G. R. (2025). The application of deep learning-based artificial intelligence in enhancing English listening and speaking skills. *ONTOLOGI: Jurnal Pembelajaran dan Ilmiah Pendidikan*, 3(2), 33–41. <https://doi.org/10.65787/ontologi.v3i2.642>
- [30] Pasaribu, G. R., & Zang, Y. (2024). Leveraging phonetics to elevate pronunciation in second language acquisition in STIT Al-Ittihadiyah North Labuhanbatu. *SIMPLE: International Journal of English Education*, 2(2), 7–15. <https://doi.org/10.24176/simple.v2i2.13632>
- [31] Pasaribu, G. R., Daulay, S. H., & Saragih, Z. (2023). The implementation of ICT in teaching English by the teacher of MTS Swasta Al-Amin. *English Language and Education Spectrum*, 3(2), 47–60. <https://doi.org/10.53416/electrum.v3i2.146>