

Allocating Decision Rights in Generative AI-Assisted Arabic Writing: A Stage-Gated Human-in-the-Loop Framework for Beginner Learners of Arabic as a Foreign Language

Dr. Mahmoud Abd ElFattah Ibrahim Essa

Academy of Language Studies, Universiti Teknologi MARA (UiTM), Terengganu, Malaysia

DOI: <https://doi.org/10.47772/IJRISS.2026.100800797>

Received: 04 September 2026; Accepted: 09 September 2026; Published: 18 September 2026

ABSTRACT

The general calls for human oversight in generative AI-assisted writing fail to identify who has consequential decision rights at each step, which is particularly significant in beginner Arabic writing. This article proposes an evidence-informed, stage-gated human-in-the-loop approach to decision rights between the learner, teacher/institution, and GenAI. The framework was created by conceptual synthesis of a bounded, purposively constructed evidence base, not by engaging in a new empirical intervention. It involves five components: nine decision-right points (DR0-DR8), eight writing stages (S0-S7), five Arabic linguistic verification gates (L1-L5), conditional teacher mediation, and the Learner-Control Process Audit. The verification gates are directed at orthography and written form, morphology and agreement, syntax and sentence construction, lexical-semantic choice, meaning, coherence and learner voice. Throughout the framework, learners have control over the communicative purpose, who initiates and what type of support is provided, how suggestions are evaluated, revision, final wording and submission. Task specific conditions for AI use are controlled by teachers or institutions and teachers mediate unresolved linguistic, semantic, procedural, and ethical issues. GenAI should be treated as an assistive actor, and is not able to make revisions or final decisions on the text independently, but can offer suggestions, explanations and feedback when requested. The process audit connects the learner's intention, pre-AI drafting, AI interactions relevant to the task, accept–modify–reject decisions, revision history, teacher mediation, if necessary, and disclosure. These records are not proof of authorship or AI detection, but rather process evidence in line with learner control. The framework provides a testable conceptual specification for responsible GenAI-assisted beginner Arabic writing and will need to be further validated with A1-A2 learners in the future to make claims of feasibility and effectiveness.

Keywords: Generative artificial intelligence, Arabic as a foreign language, decision rights, learner agency, human-in-the-loop writing

INTRODUCTION

The Problem of Unallocated Decision Rights in Generative AI Assistance

Generative artificial intelligence (GenAI) has been used more and more in second language writing for planning, text generation, feedback, revision, and language correction, with evidence varying at different stages and outcomes (Crosthwaite & Sun, 2026; Moorhouse et al., 2025; Rong & Yao, 2026). Much of this work focuses on what AI offers or what learners can do when they have access to AI. Less explicit is the distribution of responsibility between learner, teacher and AI. The issue is not, therefore, whether or not to use GenAI, but who has control over purpose, who starts the process of assistance, who sets the boundaries, who checks the suggestions, who resolves the ambiguities, and who decides on the final wording.

This is an important point to consider because GenAI can serve as a linguistic resource, feedback provider, generator and interactive partner at the same time. Control is not default in the learner or in the technology, but is a function of interaction design and the decisions that are held by human participants (Godwin-Jones, 2024). Learner initiation is not enough. Huang and Wang (2025) reported learner-directed interaction, requests for

whole essays and efforts to disguise the AI-generated writing, and Pretorius (2025) mentioned dependence, homogenized language, and loss of L2 voice. Kim et al. (2026) also identified observable behaviours related to suggestions: seeking, accepting, dismissing, and modifying. Learner control should thus be looked at as a staged sequence of decisions and not just as a result of access, prompting or textual similarity.

Fragmented but Complementary Design Precedents

The current literature offers models to structure AI-assisted writing, which are scattered. Kong et al. (2024) suggested a process-oriented pedagogy, which consists of planning, prompting, verifying, integrating, and reflecting. Learner initiated local feedback, reactive assistance, limitations on AI composition, sentence level alternatives, and manual learner revision (Emerson et al., 2026; Jin et al., 2026; Kim et al., 2026) have been more constrained designs. Other research distinguished between human planning or original writing and subsequent AI-supported rewriting, but allowed students to choose which changes to make (Alyasin & Shah, 2026; Ou et al., 2024).

Another precedent is teacher mediation. Han and Li (2024) inserted teachers in the middle of ChatGPT feedback and the learner, while Emerson et al. (2026) relied on teacher intervention when a breakdown or ongoing issues occurred. The evidence taken together suggests that human-led planning, bounded assistance, accept–modify–reject (A/M/R) decisions, manual revision, teacher mediation, and process documentation are all warranted. However, it does not offer an integrated model of how these mechanisms will be linked, who will have which consequential decision rights, and what will happen to AI permissions at each stage.

The Need for Arabic-Specific Linguistic Verification

There is an additional requirement for beginner Arabic writing: generic instructions to “check AI output” are not enough. Evidence specific to Arabic shows that problems commonly occur in spelling and written form, agreement, construct-state relations, word order, function words, lexical choice, and sentence construction (Daud et al., 2022; Emiyati & Kamalia, 2026; Sopian, 2019; Zaghouani et al., 2015). Similarly, Zubaidi et al. (2025) emphasize the need for human oversight in AI-assisted Arabic writing to ensure linguistic correctness and contextual relevance. An operational example is given by Chirkunov et al. (2025) where learners review the identified problems and proposed corrections and then manually revise.

These studies differ in terms of the level of the learners and the design of the study, and do not provide an extensive A1–A2 error taxonomy or GenAI accuracy in beginner Arabic correction. They do, however, provide support for five bounded verification domains: orthography and written form; morphology and agreement; syntax and sentence construction; lexical-semantic choice; and meaning, coherence, and learner voice. These are operationalized in the framework as L1–L5 verification gates and considered as proposed components that need to be validated, with regards to their sequence, activation points and decision rules.

Research Gap, Purpose, and Conceptual Contribution

Within the approved evidence base, no included source integrates in one framework a decision required at each writing stage, an identified human decision-right holder, stage-specific limits on GenAI intervention, A/M/R decisions, a teacher-escalation pathway, Arabic-specific linguistic verification, and a process record that documents learner control without claiming to prove authorship.

This paper therefore develops an evidence-informed, stage-gated human-in-the-loop framework for GenAI-assisted Arabic writing among beginner learners of Arabic as a foreign language. It addresses three questions:

RQ1. How can decision rights be allocated among beginner learners of Arabic, teachers, and generative AI across the stages of AI-assisted Arabic writing?

RQ2. How can stage-specific AI limits, teacher escalation, and Arabic linguistic verification gates be integrated into this allocation?

RQ3. How can a Learner-Control Process Audit document learner control over the writing process without being treated as proof of authorship?

The framework integrates a Decision-Rights Matrix (DR0–DR8), eight writing stages (S0–S7), five Arabic Linguistic Verification Gates (L1–L5), conditional teacher mediation, and a Learner-Control Process Audit. It distinguishes documented mechanisms [D] from proposed rules [P] and treats process records as evidence consistent with learner control rather than proof of authorship. The article does not claim improved writing performance, prevention of dependence, authorship preservation, AI detection, or learning retention after GenAI withdrawal. Its contribution is a transparent, testable conceptual specification for future empirical validation.

Framework Development Approach

Design and Epistemic Scope

This paper employs an evidence-informed conceptual synthesis of a bounded, purposively assembled evidence base, integrating documented GenAI-writing mechanisms with contextual Arabic-writing evidence. It is not a systematic, scoping, comprehensive, or exhaustive review. The evidence base combines reviews of GenAI-supported L2 writing (Crosthwaite & Sun, 2026; Rong & Yao, 2026), conceptual accounts of pedagogy and distributed agency (Godwin-Jones, 2024; Kong et al., 2024), and empirical studies of learner initiative, personalized assistance, peer feedback, revision, teacher mediation, and process documentation (Alyasin & Shah, 2026; Guo & Li, 2024; Guo et al., 2024; Han & Li, 2024; Huang & Wang, 2025; Jin et al., 2026; Kim et al., 2026; Ou et al., 2024). No new student, teacher, questionnaire, assessment, or institutional data were collected. The framework is therefore proposed for future validation rather than presented as an empirically evaluated intervention.

Evidence Assembly and Verification

Evidence was compiled in an iterative manner using bounded procedures: an initial file screening, structured exploratory searching of official open sources, verification of publishers and DOIs, targeted full text retrieval, and one bounded contextual round. Thirty-one (31) candidate records or files were initially screened. Five locally available PDFs were analyzed in depth, one of which was used as direct evidence, and the rest of which were not used due to document mismatch, lack of focus on writing, questionnaire status, or lack of relevance to the writing process. Questionnaires, student reports and grade data collected in the past were not utilized due to the unknown ethical status and/or relevance.

Two avenues of exploratory searching were followed. The direct pathway brought together GenAI/ChatGPT, L2 writing, and the concepts of agency, authorship, voice, decision rights, teacher feedback, and writing stages. The contextual pathway consisted of the combination of Arabic writing and the terms beginners, non-native learners, errors, orthography, learner corpora, Malaysia, and UiTM. These searches were not exhaustive and were not considered as a PRISMA-style retrieval flow.

Included sources were verified with the publisher or journal pages, proceedings repositories, DOI records, or verified author versions. Ideally, full text evidence was used; abstract-only use was restricted to explicit claims. Three publisher verified records were at abstract or metadata level (Bai & Wei, 2024; Lee, 2024; Zou et al., 2026) and no unstated details were inferred. Search strings were prepared for Scopus, Web of Science, ERIC and LLBA but were not directly executed and completeness or index coverage is not claimed.

Final Evidence Base and Analytical Separation

There are 22 direct-evidence records (D01–D22) and nine contextual records (C01–C09) in the bounded evidence matrix; an Exclusion Log is also provided for 31 excluded sources or files. These are the project evidence figures, not the size of the literature. GenAI-assisted writing, learner agency/voice, role allocation, intervention limits, teacher mediation, and process documentation were informed by direct evidence. Contextual evidence was limited to Arabic-writing risks and potential verification targets, and was not used to conclude GenAI

effectiveness. Evidence with Arabic-speaking EFL learners was not considered evidence of Arabic as the target language.

Eligibility, Extraction, and Non-Inference Rules

Direct-stream sources were eligible when they addressed GenAI-assisted writing or L2 written feedback and provided information relevant to writing stages, roles, intervention limits, agency, voice, learner control, or process documentation. Contextual sources were eligible when they documented Arabic writing by non-native or institutionally relevant learners or provided explicit evidence of orthographic, morphological, syntactic, lexical, semantic, or textual difficulties.

Extraction recorded bibliographic identity, access status, study type, context and sample, writing stage, learner role, GenAI role, teacher role, intervention boundary, actual measure, relevant finding, evidence directness, limitations, and proposed framework use. Information not explicitly reported was recorded as “not stated” rather than inferred. This rule applied especially to CEFR level, teacher intervention, automatic application of AI suggestions, final-text generation, learner-authorship outcomes, and reasons for learner decisions.

Conceptual Synthesis Procedure

The synthesis proceeded through four operations. First, recurring mechanisms were identified across the direct-evidence stream, including human-led planning, pre-AI text, learner-initiated requests, bounded AI support, accept–modify–reject behaviour, manual revision, teacher review, interaction records, and reflection. Second, mechanisms were reorganized around the decisions they addressed, producing the Decision-Rights Matrix. Third, they were arranged into stages S0–S7, from AI-use governance through planning, pre-AI drafting, bounded assistance, evaluation, revision, teacher mediation, and the final Learner-Control Process Audit. No included study evaluated this integrated sequence; it is a conceptual synthesis. Fourth, contextual Arabic-writing evidence was consolidated into five verification domains: orthography and written form; morphology and agreement; syntax and sentence construction; lexical and semantic choice; and meaning, coherence, and learner voice.

Evidence-Status Calibration

Two statuses were given to framework claims. “[D]” indicates a mechanism, role, boundary, or practice that was explicitly stated in at least one approved source; “[P]” indicates a proposed rule that was developed by integrating, transferring, or operationalizing evidence. The “[D]” does not mean causal effectiveness, validation for beginner Arabic learners or universal applicability. Similarly, the original drafts, interaction histories, revision records and reflections are documented practices, but their combination into the Learner-Control Process Audit is proposed. The resulting record can be used to show evidence of the process, not evidence of authorship, as per learner control. To ensure that proposed rules are not offered as empirical findings, the source basis and treatment of central framework claims is included in Appendix A.

The conceptual and empirical foundations of the study.

GenAI-assisted L2 writing provides direct evidence for Learner Agency, Learner Intervention Boundaries, and Teacher Mediation, and contextual evidence in Arabic writing identifies targets for linguistic verification. No direct study has tested the integrated mechanisms with beginner Arabic learners, and the Arabic studies do not set up GenAI governance, so these two evidence streams are complementary but not interchangeable. The framework thus combines documented mechanisms and proposed rules, but does not assert demonstrated effectiveness.

Learner Agency, Voice and Final Decision Control.

Here, learner agency is understood to be ongoing control of communicative purpose, the need to ask for help, the extent of help, assessment of the help and final text selection. Research has revealed mechanisms by which such control can still be seen. Alyasin and Shah (2026) found a structured sequence of planning and drafting

first, then using ChatGPT for refining, with filtering, adapting, and rejecting suggestions being reported, but agency was not measured separately. Kim et al. (2026) made the behaviours of making learner requests and accept–modify–reject observable, and their planning scaffold was linked to increased independent-authoring patterns and decreased similarity. But semantic similarity was not directly related to agency. In one advanced-writer case study, Jacob et al. (2025) demonstrated verification of references, incorporation of evidence, recovery of voice, and a decrease in the use of ChatGPT, but the study cannot be generalised to beginners.

Learner initiation alone is therefore insufficient as evidence of control. Huang and Wang (2025) found predominantly learner-led interaction but also documented requests for complete essays, selective verification, and ambiguous attribution. Pretorius (2025) reported perceived threats to voice associated with overreliance, homogenisation, and diminished human contribution, but did not measure textual voice objectively. Together, the evidence supports repeated human judgement without establishing preserved authorship.

The framework therefore proposes final text-level authority remain with the learner through explicit acceptance, modification, or rejection. Planning records, a pre-AI draft, prompts, version comparisons, and learner explanations may provide process evidence consistent with learner control, but do not prove authorship. The Learner-Control Process Audit records whether meaningful learner decisions remain visible; it is not an authorship-authentication device.

AI Intervention, Staged and Bounded

The use of GenAI capability is not a determinant of when or how much intervention is allowed; intervention design impacts the amount of learner work remaining. The research on GenAI-supported L2 written-feedback has focused on specific feedback and revision applications, leaving the scope of the assistance in the entire writing process under-explored (Crosthwaite and Sun, 2026). Kong et al. (2024) spread out the responsibility of the learner throughout the phases, but their model is still a design proposal and not an actual allocation of decision rights.

Partial precedents are provided by empirical studies. Emerson et al. (2026) placed learner-initiated feedback following an initial contribution, restricted AI to three local problems, and mandated manual revision with teacher oversight. The reported gains were in terms of knowledge of grammar, not in terms of the quality of writing or authorship. Kim et al. (2026) limited assistance to alternatives to the sentence requested by the learner and ensured that the decision of uptake was visible. Jin et al. (2026) found that a proactive questioning strategy reorganised the reasoning pattern without making a significant direct impact on writing scores, compared to the reactive questioning strategy, while the agent was not allowed to write the final paragraph of the learner. The unsupported task was not a delayed-retention test. Alyasin and Shah (2026) and Ou et al. (2024) also kept human writing as the initial writing and preserved interaction records or reflection.

These designs include: timing, learner initiation, textual granularity, number of suggestions, and mode of application for constraining AI. They do not set a general standard for a human first draft or any one intervention limit. Delayed AI entry, a learner draft prior to AI, and S0–S7 are proposed as safeguards for beginner Arabic writing. They are not intended to assert previous validation in this target population, but to ensure convenience does not silently transfer decision control.

Teacher Mediation and Escalation

The most defensible teacher involvement is to safeguard learning and interpretive judgement, not supplant learner decisions. Yan (2024) found that immediate gains in the assisted draft were different from the performance in the new writing task: individual processing yielded the best immediate gains, while teacher-supported processing yielded the best performance in the new writing task. This is a short-term EFL outcome and does not imply retention or transfer to Arabic, but rather demonstrates that product improvement and subsequent learning do not have to be correlated.

Han and Li (2024) offer a mediation sequence. ChatGPT produced initial feedback, teachers reviewed and modified it (many times from direct feedback to coded indirect feedback) and learners corrected the problems in

the revision stage. About 64% of the revisions made due to errors were successful, though the non-comparative design does not allow for separating the AI effect from the teacher effect. When learners became stuck or issues continued, teacher monitoring and intervention were also utilized by Emerson et al. (2026). Ou et al. (2024) reported a teacher-led setting where learners initially wrote and then recorded, assessed accuracy and usefulness, and reflected on ownership, voice, and ethics, but no writing gains were demonstrated.

The framework calls for teacher control of use conditions, assessment compliance, and escalation, and text-level uptake is with the learner. Escalation is suggested where AI guidance is linguistically ambiguous, is inconsistent with the intended meaning, goes beyond the scope of the guidance, or involves ethical or policy issues. These triggers are proposed rules and not a validated protocol. The teacher's responsibility is to diagnose, explain, model verification, or redirect, not to make the AI output an unquestioned answer.

Arabic Linguistic Vulnerabilities as Verification Targets

The Arabic-writing evidence identifies what requires checking. It neither demonstrates GenAI effectiveness nor establishes a comprehensive taxonomy of beginner errors. Sopian (2019), in a first-year UiTM sample, identified difficulties involving spelling, gender agreement, verb use, and *idāfa*, but reported neither error frequencies nor coding reliability. Zaghouani et al. (2015), using corpus-based annotation across mixed proficiency levels, documented punctuation, definiteness, lexical choice, hamza, conjunctions and prepositions, deletion and addition, agreement, case marking, and word order; the aggregated levels prevent attribution of these categories specifically to beginners.

More focused studies sharpen targets without resolving the proficiency gap. Brosh (2015) coded 457 spelling errors among second- and third-year university learners, with wrong-letter selection and omission of long-vowel letters among the largest categories; participants were not A1–A2 and inter-coder reliability was not reported. Kadaï et al. (2025) documented present-tense formation errors in a Malaysian Level 1 Arabic course, including difficulties with morphological marking and written verb forms. “Level 1” is treated only as partly equivalent to beginner status, not CEFR A1–A2. Khamaiseh et al. (2024) provide narrower evidence on verb patterns and contextual morphological selection, but their third- and fourth-year sample cannot validate beginner rules. Ismail and Che Mat (2026) identify broader Malaysian difficulties in grammar, sentence construction, idea connection, and paragraph development, without item-level error results.

The evidence supports verification coverage rather than prevalence rankings. The framework therefore groups the documented vulnerabilities into five domains: L1 orthography and written form; L2 morphology and agreement; L3 syntax and sentence construction; L4 lexical and semantic choice; and L5 meaning, coherence, and learner voice. The learner-voice component of L5 derives from the direct writing evidence in Section 3.1, not from Arabic error studies. The five-gate organisation is proposed rather than an exhaustive Arabic error classification. Its function is to trigger learner checking and, when necessary, teacher mediation; it does not authorise automatic correction, establish that every category is beginner-specific, or imply that the integrated gate sequence has already been tested.

The Proposed Stage-Gated Human-In-The-Loop Framework

Framework Design Principles

The proposed framework is to split the helping from the authority. GenAI can make suggestions, explanations, or feedback about what it was asked to do, but does not have a decision making right over the learner's communicative purpose, use of a suggestion, final wording or submission [P]. Learner control is thus not identified by the presence or absence of AI but rather by the observable decisions that are made. The framework keeps directly documented practices like learner-initiated requests, accept–modify–reject (A/M/R) behaviour, manual incorporation and teacher adaptation of feedback and proposes their consolidation in a single stage-gated architecture [P].

The architecture is based on five principles: each suggestion is learner-evaluated and an A/M/R decision is made [P]; assistance is learner-initiated and problem-bounded [P]; a pre-AI draft is saved and used as a basis for

comparison with the AI-generated suggestions [P]; the right-holder for each consequential decision is identified [P]; and teacher mediation is conditional [P]. Human-led planning, pre-AI drafting, reactive/local support, A/M/R behaviour, teacher adaptation of feedback are documented practices [D] but are proposed for integration into these rules [P].

Decision-Rights Allocation

The framework is divided into decision rights (DR0-DR8) and writing stages (S0-S7): The DR codes answer the question of who decides what, while the S codes answer the question of when work proceeds. DR0 controls the conditions of use of AI, DR1-DR2 control communicative purpose and planning, DR3 controls when and how much assistance is allowed, DR4-DR5 control evaluation and A/M/R, DR6 control conditional teacher mediation, and DR7-DR8 control final wording, disclosure, and submission.

GenAI is a functional role, without a primary right-holder designation. DR0 and escalated resolution at DR6 is controlled by the teacher/institution; meaning, requests, evaluation, uptake, revision, final wording, and submission is controlled by the learner. Teacher intervention alters source of support without ceding typical text-level control [P].

Table 1. Condensed Decision-Rights Matrix

Code	Consequential decision	Primary right-holder	GenAI boundary	Teacher role	Status
DR0	Establish task-specific AI-use conditions	Teacher/institution	No task use before conditions are communicated	Sets guidance on prompting, verification, privacy, copyright, and ethics	Allocation [P]; guidance practices [D]
DR1	Determine communicative purpose and intended meaning	Learner	May assist but not redefine purpose independently	Clarifies task without selecting learner position	[P] allocation built on [D] learner contribution
DR2	Prepare initial plan	Learner	Ideational support only after intention is recorded	Scaffolds task without supplying plan	[P] requirement; human-led planning [D]
DR3	Decide when, why, and within what scope to seek assistance	Learner	Responds to learner-defined, bounded request	Monitors or clarifies permitted scope	Behaviour [D]; general boundary [P]
DR4–DR5	Evaluate and accept, modify, or reject suggestions	Learner	Suggests but cannot determine or automatically apply uptake	Clarifies unresolved uncertainty after escalation	A/M/R [D]; framework rule [P]
DR6	Resolve persistent or ambiguous feedback	Teacher for escalated cases	Provides input, not a final ruling	Verifies, adapts, or converts feedback into a cue	Mediation [D]; escalation-only allocation [P]
DR7–DR8	Determine final wording, document AI use, and submit	Learner	No independent finalisation or submission authority	Checks procedural compliance	Practices [D]; formal allocation [P]

The Eight Writing Stages: S0–S7

The eight stages are not a continuous generation pipeline, but rather procedural gates. Passing a stage does not imply linguistic mastery, authorship or effectiveness [P]. S0 sets the conditions for using AI; S1 documents intention and planning; S2 stores the learner’s pre-AI draft, but does not prohibit the use of AI for ideational support later.

At S3 the learner is able to identify a problem and ask for limited help. The returned suggestion is evaluated by the learner at S4 with the corresponding linguistic gates and an A/M/R decision is made. The learner does the revision instead of letting GenAI do it automatically at S5. The learner may go back to S3 from S4 or S5 if another resolvable problem occurs, but GenAI does not do this. If there is any doubt or if it is a question of policy or ethics, it could be referred to S6. Teacher checks/remodels feedback and sends the learner back to S5 to do revision. S7 completes the process with final linguistic review, disclosure, record consistency and the learner's decision to submit [P].

Table 2. Condensed Operational Matrix of S0–S7

Stage	Core action	AI boundary	Verification / escalation	Process record
S0 AI-Use Governance	Conditions established and acknowledged	No task use before guidance [P]	Policy, privacy, copyright, ethics	Intended-use statement
S1 Intention and Planning	Learner defines purpose and initial plan	Later ideational support allowed only after intention is recorded [P]	L4–L5 may check relevance and intended meaning	Saved intention/plan
S2 Pre-AI Learner Draft	Learner produces and saves initial text	GenAI inactive until draft is saved [P]	Preliminary L1–L5 scan; teacher scaffolds only if needed	Saved pre-AI draft
S3 Bounded Request	Learner identifies problem and scope	Reactive/local assistance; no complete replacement text [P]	Scope drift or unverifiable content may escalate	Prompt/output record
S4 Evaluation and A/M/R	Learner evaluates suggestion	No automatic adoption	Apply relevant L1–L5; escalate persistent ambiguity	A/M/R decision and rationale
S5 Learner-Executed Revision	Learner performs revision	Manual incorporation; learner may return to S3	Recheck relevant gates; escalate unresolved alternatives	Revision history
S6 Teacher Mediation	Teacher verifies disputed feedback; learner revises	AI has no final ruling	Teacher may provide an indirect cue or explanation	Teacher cue + learner revision
S7 Final Process Audit	Learner finalises and submits	No unrecorded final generation [P]	Final L1–L5 review and compliance check	Complete process record and disclosure

Arabic Linguistic Verification Gates: L1–L5

The L1–L5 gates are the transformation of documented areas of Arabic-writing difficulty [D] into operational verification questions [P]. They set out what has to be verified before accepting AI-supported wording; they do

not grant GenAI the ability to diagnose or correct the learner's wording. L1 is orthography and written form; L2 is morphology and agreement; L3 is syntax and sentence construction; L4 is lexical-semantic choice; and L5 is intended meaning, coherence and learner voice.

The gates are not an exhaustive list of A1-A2 errors, a proficiency scale or an automated correction system. Their activation is stage-sensitive. L4 and L5 are implicit at S1 to maintain relevance and intended meaning, and are explicit during assistance, evaluation, revision, mediation and final review. If a gate is not passed, the AI-supported suggestion may be modified, rejected, locally revised, or escalated to the teacher [P].

Table 3. Condensed Arabic Linguistic Verification Gates

Gate	Verification focus	Operational question	Main activation
L1	Orthography and written form	Is the written form acceptable without changing the intended word?	S3–S7
L2	Morphology and agreement	Does the form express the required grammatical features and agree with surrounding words?	S3–S7
L3	Syntax and sentence construction	Is the sentence grammatically structured and are its internal relations clear?	S3–S7
L4	Lexical-semantic choice	Does the expression convey the learner's intended meaning in context?	S1, S3–S7
L5	Meaning, coherence, and learner voice	Does the revision preserve intended meaning, coherence, and learner-appropriate voice?	S1–S7

The Learner-Control Process Audit

The Learner-Control Process Audit connects the learner's decisions across S1–S7. It assembles the recorded intention or plan, pre-AI draft, learner prompts and AI outputs, A/M/R decisions, revision history, any teacher cue and subsequent learner revision, and final disclosure. These record types are separately documented practices [D]; requiring and interpreting their combined sequence as an audit is proposed [P].

The audit asks whether the record coheres with the framework's allocation of control across intention, bounded requests, uptake decisions, revision, and submission. Its conclusion remains procedural. The record may provide process evidence consistent with learner authorship, but it is not proof of authorship. No single artifact, similarity score, estimated percentage of AI contribution, or edit count independently establishes who controlled meaning and decisions. Missing records, unexplained large-scale divergence, or undisclosed complete generation trigger a compliance question or teacher review [P], not an automatic finding of misconduct or non-authorship.

Integrated Framework Logic

The integrated framework coordinates the S0–S7 process pathway, DR0–DR8 decision authority, L1–L5 verification gates, and the Learner-Control Process Audit that links records across stages. The principal route is S0 → S1 → S2 → S3 → S4 → S5 → S7. Two controlled departures preserve responsiveness: S4 or S5 may return to S3 when the learner initiates further bounded assistance, and S3–S5 may escalate to S6 when an issue remains unresolved; teacher mediation then returns the learner to S5 to execute the revision.

A transition is permissible only when the relevant human actor has made the required decision, applicable verification has been completed, and the necessary process record is available [P]. Neither revision loops nor escalation transfer final control to GenAI, and neither depends on a fixed cycle count or universal escalation threshold. The framework therefore integrates bounded assistance, Arabic-specific verification, conditional teacher mediation, and process documentation around learner-held decisions. As a complete architecture, it remains proposed [P] and requires future empirical validation with beginner learners of Arabic.

DISCUSSION

From General Human Oversight to Stage-Specific Decision Rights

The key conceptual step of the framework is the explicit assignment of consequential decision rights instead of a general requirement for “human oversight.” Human presence is not sufficient to create meaningful control: a learner can initiate an interaction, but accept a lot of generated content, and a teacher can be present, but not be the one to decide when mediation is needed. For instance, Huang and Wang (2025) reported cases of learner-directed interaction, requests for full essays, and efforts to camouflage AI-generated writing. Agency, then, does not necessarily follow from who starts an interaction or who is “around” in it.

The proposed framework, on the other hand, views control as a series of decisions regarding communicative purpose, who initiates and what type of assistance is provided, how suggestions are evaluated and incorporated, how uncertainty is resolved, revision, and final wording. This is an extension of distributed-agency accounts, which place control not necessarily in the hands of either actor (Godwin-Jones, 2024). It also moves beyond process models in which the learner is a planner, verifier, integrator, and reflector, but in which no formal authority is given to any one of these choices (Kong et al., 2024). The practices of learner initiation, A/M/R behaviour, manual revision and teacher mediation are documented practices [D] and the proposed synthesis of the framework is the conversion of these practices into stage-specific decision rights [P]. The difference is thus a qualitative one: oversight is the question of whether a human is involved, and decision-right allocation is the question of who can decide what, when, and with what constraints on AI action.

Maintaining Learner Control while not Banning ALL AI Support

The framework also challenges the idea that maintaining learner control means excluding all AI contributions. Complete prohibition eliminates the assistance function, but does not solve the problem of how learners are supposed to assess the automated assistance if it is available. The opposite problem is that technical capability becomes de facto permission for the system to replace learner work when it is used without restrictions.

The difference is thus between capability and permission. While GenAI can technically produce ideas, sentences, revisions, or whole texts, an instructional design specifies which operations can be used, who can ask for which operations, and who makes the decision of whether what GenAI produces is included in the learner's text. This is a middle ground that has been found in previous studies. Prior designs (Emerson et al., 2026; Jin et al., 2026; Kim et al., 2026) have included learner-initiated local feedback, reactive assistance, restrictions on composing final text, sentence-level alternatives, and manual incorporation. Other research distinguished between human planning or initial drafting and subsequent AI-supported revision, while allowing learners to decide which revisions to adopt (Alyasin & Shah, 2026; Ou et al., 2024).

These mechanisms are examples of feasible bounded assistance but do not necessarily represent the optimal form of bounded assistance. The contribution of this framework is to bring them into the permission structure where the AI support is available, but the AI cannot redefine purpose, initiate revision cycles, adopt its own suggestions, or finalize the learner's text [P]. Learner control is therefore safeguarded without being forced to make the stronger and unsupported claim that each and every ideational or linguistic contribution from AI inevitably results in the loss of agency.

Why Arabic Linguistic Verification Gates Matter

Generic instructions to “check AI output” are too vague for Arabic writing. AI suggestions may still contain problems in orthography, agreement, sentence relations, lexical choice, or intended meaning. Arabic studies document difficulties in spelling and hamza, agreement, definiteness and construct-state relations, word order, function words, lexical selection, and sentence construction (Daud et al., 2022; Sopian, 2019; Zaghouni et al., 2015). Zubaidi et al. (2025) similarly stress human verification of linguistic accuracy and contextual fit in AI-supported Arabic writing.

The value of L1–L5 is therefore not that the gates create a complete error taxonomy, but that they transform documented vulnerability domains [D] into decision-relevant verification questions [P]. They separate written form, morphology and agreement, syntax, lexical-semantic fit, and the preservation of meaning, coherence, and learner voice. Verification is thereby linked directly to A/M/R: an AI suggestion remains a candidate until the learner has examined the relevant uncertainty. This logic is compatible with Arabic-writing systems that distinguish identified problems from proposed corrections and preserve manual learner revision (Chirkunov et al., 2025), as well as minimum-change approaches that avoid replacing an acceptable learner formulation unnecessarily (Zaghouani et al., 2015).

The five-gate structure nevertheless remains conceptual [P]. It does not represent an exhaustive taxonomy of A1–A2 errors, demonstrate that GenAI detects those errors reliably, or establish effectiveness with beginner Arabic learners. Its contribution is to specify what checking should attend to before a suggestion becomes part of the learner's text.

Process Evidence versus Authorship Detection

The Learner-Control Process Audit poses a different question than AI-authorship detection. Detection is an attempt to classify the end product or estimate the amount of AI generation. The audit does not review whether the learner has access to the decision rights specified by the framework, but rather whether the process record provided is consistent with the learner having access to the specified decision rights. Documented writing designs (Alyasin & Shah, 2026; Ou et al., 2024) have been preserved to include original drafts, interaction histories, revised texts, and reflections, but such studies do not demonstrate that such artifacts establish final-text authorship.

There are some shortcuts that are not included in this distinction. Semantic similarity between an AI suggestion and the final text may suggest reliance or engagement, as in Kim et al. (2026); however, semantic similarity alone does not prove who was in control of meaning, evaluation, or revision. Similarly, an AI contribution percentage, similarity score, or number of accepted changes are not adequate proof of authorship. Product level indicators and process level decisions address different questions.

The conclusion of the audit must therefore be qualified: it is process evidence that is consistent with learner authorship, but not proof of authorship. The absence of records, significant deviations in scale, or undisclosed complete AI generation may trigger a procedural inquiry or teacher review [P] but do not automatically indicate misconduct or non-authorship. The audit is thus an accountability mechanism, not an AI detector, authorship score, or evidentiary shortcut. Its role is to provide learner control that can be inspected without losing the distinction between a coherent process record and definitive attribution.

Pedagogical Implications

Implications for Writing-Task Design

The framework implies that GenAI use should be designed into the writing task rather than left as an unstructured activity. A task brief may require a concise statement of the learner's intended message and a saved pre-AI draft before assisted revision begins [P]. Assistance should then be linked to learner-identified problems, with bounded requests, visible accept–modify–reject decisions, and learner-executed revision. Process records should be proportionate to task stakes and complexity, retaining only what is needed to make the writing process reviewable rather than logging every minor interaction.

Implications for Teacher Practice

Teachers should clarify task-specific AI conditions before writing begins, including permitted functions, disclosure expectations, verification responsibilities, and relevant privacy, copyright, and ethical constraints [P]. Modelling should focus on how to formulate a bounded request, inspect a response through the relevant Arabic verification gates, and preserve intended meaning when accepting or modifying suggestions. During writing, mediation should remain conditional: teachers may diagnose unresolved linguistic or semantic problems, verify

feedback, or convert direct correction into an indirect cue, while the learner still performs the revision. Continuous surveillance or routine rewriting would undermine the learner control the framework is designed to preserve.

Implications for Assessment and Institutional Policy

Assessment should distinguish the quality of the submitted Arabic text from evidence about how the text was produced [P]. Product quality may be judged using normal writing criteria, while process review considers compliance with task conditions, disclosure, relevant records, and learner decisions. Institutional guidance should align permitted assistance, disclosure, record retention, privacy, copyright, and equitable access to tools. Missing records, textual divergence, or suspected AI use may justify clarification, but they should not automatically be treated as proof of misconduct or non-authorship; consequential decisions should follow established institutional procedures and consider process evidence in context [P].

Limitations And Future Empirical Validation

The framework is a conceptual synthesis, informed by evidence, and is not an empirically tested intervention. None of the studies included in this research investigated DR0-DR8, S0-S7, L1-L5, conditional teacher mediation, and the Learner-Control Process Audit as a whole system. Thus, its role is specification and synthesis, not demonstrated feasibility, effectiveness, or validity for beginning Arabic writing.

Generalization is also restricted by the evidential base. The majority of direct evidence of GenAI-writing comes from EFL and university contexts, and the Arabic-writing studies mostly serve to validate linguistic verification targets, not to allocate rights of decision-making or governance of GenAI. There are a number of central requirements that are proposed rules [P] including a draft before AI, stage specific permissions, escalation triggers and the combined process audit. The audit itself has not yet been proved reliable or valid and is not proof of authorship. Further, the evidence base was bounded and purposively constructed, not systematic, scoping, exhaustive, or database-complete.

Future validation should therefore be done in stages: content validity by experts; usability and operational-clarity testing; an ethically approved A1-A2 pilot; process validation with reliability testing; comparative evaluation with unstructured GenAI use; and replication across institutions, learner groups, and writing genres. These stages should maintain the feasibility, reliability, validity, learning outcomes and transferability as separate conclusions.

CONCLUSION

To address RQ1, the framework assigns decision-making power to the learner, teacher/institution, and GenAI. The teacher or institution sets the conditions for using AI in the task, and the learner is responsible for the communicative purpose, the initiation and scope of the AI support, the evaluation and incorporation of the suggestions, revision, final wording, disclosure and submission. Thus, GenAI plays an assistive, not an authoritative, role.

As for RQ2, the framework is designed to give stage-specific permissions to support AI rather than unlimited access or outright ban. Assistance is limited to what the learner asks for and the decision being made. If there is still some doubt, whether linguistic, semantic, procedural or ethical, it can be passed on to the teacher, who confirms or intervenes in the feedback and then hands over the responsibility for revision to the learner. The L1–L5 gates link this governance logic to Arabic writing by relating it to verification at the level of written form, morphology and agreement, syntax, lexical-semantic choice, and meaning, coherence, and learner voice.

With respect to RQ3, the Learner-Control Process Audit includes the learner's intention, pre-AI draft, AI interactions relevant to the task, A/M/R decisions, revision history, teacher mediation when necessary and final disclosure. It offers process evidence consistent with learner control, rather than proof of authorship or AI detection.

The novelty of the article lies in the fact that the DR0–DR8, S0–S7, L1–L5, conditional teacher mediation, and the process audit are integrated into one single testable conceptual architecture for GenAI-assisted beginner Arabic writing. The framework is still evidence-based and proposed, and the claims of feasibility, effectiveness, authorship verification and validity for A1–A2 learners must be validated by future empirical evidence.

REFERENCES

1. Alyasin, A., & Shah, M. A. (2026). Human–AI collaboration in academic writing: Exploring university students' agency through a sociocultural lens. *Ampersand*, 16, 100256. <https://doi.org/10.1016/j.amper.2026.100256>
2. Bai, L., & Wei, Y. (2024). Exploring EFL learners' integration and perceptions of ChatGPT's text revisions: A three-stage writing task study. *IEEE Transactions on Learning Technologies*, 17, 2215–2226. <https://doi.org/10.1109/TLT.2024.3491864>
3. Brosh, H. (2015). Arabic spelling: Errors, perceptions, and strategies. *Foreign Language Annals*, 48(4), 584–603. <https://doi.org/10.1111/flan.12158>
4. Chirkunov, K., Alhafni, B., Qwaider, C., Habash, N., & Briscoe, T. (2025). ARWI: Arabic Write and Improve. In V. Padmakumar, K. Gero, T. Wambsganss, S. Sterman, T.-H. Huang, D. Zhou, & J. Chung (Eds.), *Proceedings of the Fourth Workshop on Intelligent and Interactive Writing Assistants (In2Writing 2025)* (pp. 11–18). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2025.in2writing-1.2>
5. Crosthwaite, P., & Sun, S. (2026). Generative AI and L2 written feedback studies: A scoping review. *RELC Journal*, 57(1), 207–219. <https://doi.org/10.1177/00336882251386530>
6. Daud, A. F. C., Mustapha, N. F., Hajimaming @ Toklubok, P., Abdul Jabar, M. A., & Mohamad, A. H. (2022). Problems in Arabic writing skills among Malay students in Malaysia: A review. *International Journal of Academic Research in Business and Social Sciences*, 12(1), 2028–2039. <https://doi.org/10.6007/IJARBS/v12-i1/12038>
7. Emerson, N., Holland, T., Mitchell, T. D., Shum, D., & Oga-Baldwin, W. L. Q. (2026). Embedding constrained, learner-initiated GenAI feedback in collaborative L2 writing: Targeted grammar gains with a two-week follow-up. *Technology in Language Teaching & Learning*, 8, 104008. <https://doi.org/10.29140/tltl.2026.104008>
8. Emiyati, R., & Kamalia. (2026). Arabic orthography instruction and writing accuracy among madrasah students: Insights for AI-assisted language learning. *Alibbaa': Jurnal Pendidikan Bahasa Arab*, 1(Special Issue), 210–225. <https://doi.org/10.19105/ajpba.vi.24571>
9. Godwin-Jones, R. (2024). Distributed agency in second language learning and teaching through generative AI. *Language Learning & Technology*, 28(2), 5–30. <https://doi.org/10.64152/10125/73570>
10. Guo, K., & Li, D. (2024). Understanding EFL students' use of self-made AI chatbots as personalized writing assistance tools: A mixed methods study. *System*, 124, 103362. <https://doi.org/10.1016/j.system.2024.103362>
11. Guo, K., Pan, M., Li, Y., & Lai, C. (2024). Effects of an AI-supported approach to peer feedback on university EFL students' feedback quality and writing ability. *The Internet and Higher Education*, 63, 100962. <https://doi.org/10.1016/j.iheduc.2024.100962>
12. Han, J., & Li, M. (2024). Exploring ChatGPT-supported teacher feedback in the EFL context. *System*, 126, 103502. <https://doi.org/10.1016/j.system.2024.103502>
13. Huang, Y., & Wang, D. (2025). Can ChatGPT serve as a writing collaborator? Insights from Chinese EFL learners. *System*, 133, 103775. <https://doi.org/10.1016/j.system.2025.103775>
14. Ismail, M. R., & Che Mat, A. (2026). Bridging the gap between perceptions and reality in Arabic writing proficiency among Malaysian higher education students: A triangulation approach. *Ijaz Arabi Journal of Arabic Learning*, 9(1), 482–489. <https://doi.org/10.18860/ijazarabi.v9i1.31513>
15. Jacob, S. R., Tate, T., & Warschauer, M. (2025). Emergent AI-assisted discourse: A case study of a second language writer authoring with ChatGPT. *Journal of China Computer-Assisted Language Learning*, 5(1), 1–22. <https://doi.org/10.1515/jccall-2024-0011>
16. Jin, Y., Yang, K., Martinez-Maldonado, R., Gašević, D., & Yan, L. (2026). The agency gap in AI-supported writing: How reactive and proactive agent designs shape multimodal reasoning. *Computers & Education: Artificial Intelligence*, 11, 100655. <https://doi.org/10.1016/j.caeai.2026.100655>

17. Kadai, K., Baharudin, H., Wicaksono, T., & Fitriani, N. (2025). Analysis of errors in the formation of fi'il mudari' by the University of Malaysia Sarawak students. *Ijaz Arabi Journal of Arabic Learning*, 8(2), 613–629. <https://doi.org/10.18860/ijazarabi.v8i2.30329>
18. Khamaiseh, M. A., Osman, N., & Elawadi, E. M. S. (2024). Morphological errors in the ten wazan for the writing skill of non-native Arabic language learners. *Ijaz Arabi Journal of Arabic Learning*, 7(1), 1–14. <https://doi.org/10.18860/ijazarabi.v7i1.26073>
19. Kim, S., So, H.-J., & Park, K. (2026). Supporting learner agency in collaborative writing with generative AI. *British Journal of Educational Technology*, 57(4), 984–1008. <https://doi.org/10.1111/bjet.70015>
20. Kong, S.-C., Lee, J. C.-K., & Tsang, O. (2024). A pedagogical design for self-regulated learning in academic writing using text-based generative artificial intelligence tools: 6-P pedagogy of Plan, Prompt, Preview, Produce, Peer-review, and Portfolio-tracking. *Research and Practice in Technology Enhanced Learning*, 19, 030. <https://doi.org/10.58459/rptel.2024.19030>
21. Lee, Y.-J. (2024). Can my writing be polished further? When ChatGPT meets human touch. *ELT Journal*, 78(4), 401–413. <https://doi.org/10.1093/elt/ccae039>
22. Moorhouse, B. L., Wan, Y., Wu, C., Wu, M., & Ho, T. Y. (2025). Generative AI tools and empowerment in L2 academic writing. *System*, 133, 103779. <https://doi.org/10.1016/j.system.2025.103779>
23. Ou, A. W., Khuder, B., Franzetti, S., & Negretti, R. (2024). Conceptualising and cultivating critical GAI literacy in doctoral academic writing. *Journal of Second Language Writing*, 66, 101156. <https://doi.org/10.1016/j.jslw.2024.101156>
24. Pretorius, M. (2025). 'I want the text to still be me': Foregrounding ChatGPT's impact on L2 voice. *English Text Construction*, 18(1), 38–67. <https://doi.org/10.1075/etc.25001.pre>
25. Rong, M., & Yao, Y. (2026). From assistance to advancement: A systematic review of how GenAI supports L2 writing learning from the lens of activity theory. *ECNU Review of Education*, 9(2). <https://doi.org/10.1177/20965311261437802>
26. Sopian, A. (2019). Kemahiran menulis bahasa Arab: Analisa kelemahan pelajar UiTM Melaka [Arabic writing skills: UiTM Melaka student weakness analysis]. *International Journal of Humanities, Philosophy and Language*, 2(7), 58–65. <https://doi.org/10.35631/ijhpl.27005>
27. Yan, D. (2024). Comparing individual vs. collaborative processing of ChatGPT-generated feedback: Effects on L2 writing task improvement and learning. *Language Learning & Technology*, 28(1), 1–19. <https://doi.org/10.64152/10125/73597>
28. Zaghouni, W., Habash, N., Bouamor, H., Rozovskaya, A., Mohit, B., Heider, A., & Oflazer, K. (2015). Correction annotation for non-native Arabic texts: Guidelines and corpus. In A. Meyers, I. Rehbein, & H. Zinsmeister (Eds.), *Proceedings of the 9th Linguistic Annotation Workshop* (pp. 129–139). Association for Computational Linguistics. <https://doi.org/10.3115/v1/W15-1614>
29. Zou, S., Guo, K., Wang, J., & Liu, Y. (2026). Investigating students' uptake of teacher- and ChatGPT-generated feedback in EFL writing: A comparison study. *Computer Assisted Language Learning*, 39(4), 1062–1091. <https://doi.org/10.1080/09588221.2024.2447279>
30. Zubaidi, A., Munip, A., Widodo, S. A., & Zerrouki, T. (2025). Enhancing Arabic writing skills using Chat GPT-based AI learning models: A tridimensional human–AI collaboration framework. *Indonesian Journal of Applied Linguistics*, 15(1), 87–101. <https://doi.org/10.17509/ijal.v15i1.75378>