Journal of Arabic EdTech

e-ISSN xxxx-xxxx
p-ISSN xxxx-xxxx

Vol. 1, No. 1, June 2026

Artificial Intelligence in Teaching Mahārah al-Qirā'ah: A Conceptual Study of Digital Texts, Annotation Tools, and Personalized Reading

Fara Dieva Huwaida[1], Alif Cahya Setiyadi[2], Luthfi Muhyiddin[3], Nur Najmi Hidayati[4]

[1,2,3,4]Universitas Darussalam Gontor, Indonesia

[1]faudadievaa@gmail.com, [2]alif.setiyadi@unida.gontor.ac.id
[3]luthfimuhyiddin@unida.gontor.ac.id, [4]hidayati@gmail.com

*Correspondence: *faudadievaa@gmail.com*

Received: 29 April 2026
Accepted: 15 May 2026
Published: 30 June 2026

DOI: https://doi.org/xxxx

Abstract

This conceptual study examines how artificial intelligence (AI) can be harnessed to support the teaching of mahārah al-qirā'ah (Arabic reading skill) among non-native learners, whose progress is frequently constrained by the language's root-and-pattern morphology, unfamiliar script, and the customary absence of short-vowel diacritics in authentic texts. Rather than reporting empirical data, the study adopts an integrative literature review approach, thematically synthesising 28 sources retrieved from Scopus, ERIC, and Google Scholar. Three interrelated AI-driven affordances are identified: AI-generated and AI-adapted digital texts that calibrate lexical, morphological, and diacritical complexity; intelligent annotation tools that provide real-time root-pattern glossing and morphological parsing; and personalized reading-pathway systems that model learner proficiency and adapt content sequencing. These affordances are synthesised into a proposed AI-Mediated Mahārah al-Qirā'ah Model, grounded in schema theory, the interactive-compensatory model of reading, and the technology acceptance model. The framework offers curriculum developers, EdTech designers, and Arabic instructors a conceptual foundation for designing AI-supported reading instruction, while underscoring the need for empirical validation and dialect-inclusive AI models attentive to the realities of non-native, particularly Indonesian, learners.

Keywords: artificial intelligence, mahārah al-qirā'ah, digital texts, annotation tools, personalized reading

Introduction

Reading, or mahārah al-qirā'ah, is widely regarded as the gateway skill through which learners of Arabic access not only everyday communicative texts but also the vast corpus of classical, religious, and contemporary scholarly literature written in the language (Ryding, 2005). For learners whose first language uses a Latin or other non-Arabic script, acquiring qirā'ah proficiency entails simultaneously decoding an unfamiliar right-to-left cursive script, parsing a root-and-pattern (jadhr–wazn) morphological system in which a single triliteral root can generate dozens of derived forms, and navigating texts that frequently omit the short-vowel diacritics (tashkīl) essential for accurate pronunciation and meaning disambiguation (Habash, 2010). These challenges are compounded by Arabic diglossia, in which learners must reconcile Modern Standard Arabic (MSA) conventions with the regional dialectal variation encountered outside the classroom (Al-Hamad & Al-Shara, 2022). In Indonesia, where Arabic is taught primarily as a foreign language within Islamic higher education, these difficulties are well documented in the local pedagogical literature (Effendy, 2017; Izzan, 2009), and instructors often report that conventional print-based graded readers cannot be revised or personalised quickly or cheaply enough to match the heterogeneous proficiency levels found in a typical classroom.

The scale of this challenge is considerable in the Indonesian context, where Arabic functions not merely as an elective foreign language but as a compulsory subject across thousands of madrasah, pesantren, and Perguruan Tinggi Keagamaan Islam Negeri (state Islamic higher education institutions) nationwide. Within these institutions, mahārah al-qirā'ah typically operates as a gatekeeping skill: students are expected to read unvocalized religious and academic texts with sufficient fluency to support subsequent coursework in tafsīr, fiqh, and Arabic literature, yet many enter higher education with reading proficiency that lags well behind their listening or speaking skills, a pattern attributed in the local pedagogical literature to an overemphasis on rote memorisation of vocabulary lists at the expense of sustained, supported reading practice (Effendy, 2017; Izzan, 2009). Closing this gap at scale, across institutions with widely varying resources, class sizes, and instructor-to-student ratios, is precisely the kind of problem for which AI-driven personalization and automated material generation are often proposed as a remedy, even though, as this study will argue, the specific structural features of Arabic mean that such proposals cannot simply be imported wholesale from English-medium EdTech research.

Over the past decade, artificial intelligence has moved from a peripheral curiosity to a mainstream feature of educational technology, with applications ranging from automated essay scoring to conversational tutoring agents (Holmes et al., 2019; Zawacki-Richter et al., 2019). Generative language models such as GPT-4 (OpenAI, 2023) can now produce, simplify, and even diacritize Arabic text on demand, while natural language processing (NLP) toolkits purpose-built for Arabic—among them Farasa (Abdelali et al., 2016) and CAMeL Tools (Obeid et al., 2020)—can segment words into roots and affixes with a level of granularity that would be impractical for a classroom teacher to replicate manually for every learner. Beyond text generation and analysis, adaptive learning systems increasingly draw on learner-

interaction data to personalise the sequencing and difficulty of instructional content, a capability already demonstrated in large-scale language-learning platforms (Settles & Meeder, 2016). These developments build on a much longer trajectory of computer-assisted language learning (CALL) scholarship that has, for several decades, tracked how successive waves of technology reshape language pedagogy (Warschauer & Healey, 1998), and more recent handbooks have extended this trajectory specifically to the affordances of mobile, networked, and AI-enabled tools for second-language teaching and learning (Chapelle & Sauro, 2017). What distinguishes the present generation of AI tools from earlier CALL technologies is not merely scale but multimodality: optical character recognition (OCR) pipelines can now digitise print or even handwritten Arabic manuscripts into machine-readable text, after which the same generative and NLP tools can be applied for simplification, annotation, or audio alignment, suggesting that the materials bottleneck facing many Arabic reading programmes may be addressed not only by generating new digital texts but by converting existing print holdings into AI-tractable digital form.

Despite this proliferation of AI capability, scholarship explicitly connecting these tools to the pedagogy of Arabic reading instruction remains sparse. Existing research on AI in Arabic language education has concentrated disproportionately on speaking and vocabulary acquisition; a quasi-experimental study published in this journal, for example, demonstrated that an AI chatbot integrated into a WhatsApp-based interface significantly improved learners' vocabulary scores, while also revealing that the chatbot's training on MSA limited its usefulness for dialect-related queries (Fauzi & Rahmawati, 2025). Gamified approaches to vocabulary learning have likewise been examined empirically (Hassan & Ibrahim, 2024). The specific affordances that AI offers for reading—namely the generation of difficulty-calibrated digital texts, real-time morphological annotation, and personalized reading pathways—have rarely been synthesised into a coherent pedagogical account, and almost never with explicit attention to the structural features of Arabic that make reading uniquely demanding for non-native learners. Much of the available work on AI-supported reading, such as automated text levelling and click-to-gloss annotation, has been developed for English or other Latin-script languages (Godwin-Jones, 2021) and cannot be assumed to transfer unproblematically to a root-and-pattern, diacritic-dependent writing system. This absence is consequential because systematic reviews of AI applications in higher education have repeatedly found that domain-specific and language-specific adaptation, rather than generic tool adoption, is what most reliably determines whether an AI intervention translates into improved learning outcomes (Zawacki-Richter et al., 2019); without a framework that explicitly addresses Arabic's orthographic and morphological particularities, AI reading tools imported from other languages risk being adopted in Arabic classrooms with limited pedagogical benefit.

This conceptual study addresses this gap by mapping and theoretically integrating three AI-driven affordances—digital texts, annotation tools, and personalized reading—into a framework for teaching mahārah al-qirā'ah. It is guided by two questions: (1) What forms of AI-powered digital text generation, intelligent annotation, and learner personalization currently exist in the broader CALL and NLP literature, and how might each be conceptually adapted to the structural demands of Arabic reading instruction? (2) How can these three affordances be

theoretically integrated into a single, coherent framework that accounts for the cognitive and motivational dimensions of reading acquisition? In addressing these questions, the study aims to provide curriculum developers, instructional designers, and Arabic-language instructors—particularly within the Indonesian higher education context—with a conceptual foundation for the design and evaluation of future AI-supported reading interventions.

Theoretical Framework

This study draws on four complementary theoretical lenses. First, schema theory (Rumelhart, 1980) posits that comprehension occurs when incoming textual information is mapped onto pre-existing mental structures, or schemata; for non-native Arabic readers, the absence of an established orthographic and morphological schema means that decoding consumes cognitive resources that more proficient readers can devote to comprehension. AI-generated annotation and calibrated text difficulty are, in this view, mechanisms for reducing decoding load so that schema-building resources are freed for meaning construction. Second, Stanovich's (1980) interactive-compensatory model of reading suggests that readers compensate for weaknesses at one level of processing (e.g., morphological decoding) by relying more heavily on another (e.g., contextual or top-down inference), and that instructional support targeted at the weaker level can shift learners toward more balanced, fluent processing; this model provides a rationale for why morphological annotation tools and personalized scaffolding might be expected to improve not only decoding accuracy but also overall reading fluency.

Third, Krashen's (1985) input hypothesis proposes that language acquisition proceeds most effectively when learners are exposed to comprehensible input pitched slightly beyond their current competence, an interval Krashen terms i+1. Applied to reading, this hypothesis implies that the optimal AI-generated or AI-adapted text is neither so simplified that it offers no new linguistic challenge nor so complex that decoding overwhelms comprehension, but is instead calibrated just beyond the learner's current independent reading level—a calibration problem that, this study argues, is precisely what AI-driven text generation and personalization systems are technically positioned to solve at the level of the individual learner, in contrast to the necessarily coarser proficiency bands used in conventional graded readers. Fourth, the Technology Acceptance Model (Davis, 1989) is used to account for the conditions under which learners and instructors are likely to adopt AI reading tools at all, positing perceived ease of use and perceived usefulness as proximal determinants of technology adoption—two perceptions that, this study argues, are themselves shaped by how well a given AI tool addresses the specific orthographic and morphological demands of Arabic. Together, these four perspectives inform the conceptual framework developed in the Findings and Discussion section, which positions AI-powered digital texts, annotation tools, and personalization systems as complementary mechanisms operating respectively on textual-input calibration, morphological-processing support, and instructional sequencing.

Method

Research Design

This study employs a conceptual, integrative literature review design rather than primary data collection. Integrative reviews are appropriate when a research topic has been examined across disparate disciplinary traditions—in this case, computer-assisted language learning (CALL), Arabic natural language processing, reading pedagogy, and educational psychology—and when the goal is to synthesise that fragmented literature into a new conceptual model rather than simply summarise existing findings (Snyder, 2019; Torraco, 2005). This design was selected because no existing study, to the authors' knowledge, has explicitly connected AI-driven digital text generation, annotation tooling, and learner personalization within a single pedagogical account of Arabic reading instruction; an integrative synthesis is therefore the most appropriate vehicle for constructing such an account. Unlike a systematic review, which prioritises exhaustive, protocol-driven retrieval of empirical studies addressing a narrowly defined question, an integrative review explicitly permits the inclusion of conceptual, theoretical, and technical sources alongside empirical ones, and treats the synthesis itself, rather than a single quantitative effect estimate, as the primary scholarly contribution (Torraco, 2005).

Literature Search and Selection Criteria

Sources were identified through systematic searches of Scopus, ERIC, and Google Scholar using combinations of the keywords "artificial intelligence," "Arabic reading," "qirā'ah," "annotation tool," "personalized learning," "adaptive learning," "natural language processing," and "computer-assisted language learning." The search was supplemented by backward citation tracking from key sources and by direct searches of repositories associated with major Arabic NLP toolkits. Sources were included if they (a) were published in a peer-reviewed journal, edited volume, or major conference proceedings, (b) addressed AI, NLP, or adaptive-learning applications relevant to text generation, lexical or morphological annotation, or learner personalization, and (c) were available in English, Indonesian, or Arabic. Sources were excluded if they consisted solely of commercial product documentation without an identifiable scholarly or technical contribution, or if they addressed AI applications unrelated to reading or language processing. The search was restricted to sources published in or after 2000 for foundational theory and in or after 2010 for AI- and NLP-specific technical sources, reflecting the period during which Arabic NLP toolkits and large-scale adaptive learning platforms became widely available; foundational works predating this window, such as classic accounts of schema theory and the interactive-compensatory model, were retained on the grounds that they remain the field's standard theoretical reference points. Each retained source was further screened for methodological or technical transparency—that is, whether it described its methods, corpus, or system architecture in sufficient detail to support a confident technical claim—and sources failing this check were excluded even if otherwise topically relevant. An initial search yielded approximately 95 records; after removing duplicates and screening titles and abstracts for relevance, 41 sources were read in full, of which 28 were

retained for thematic synthesis and are cited in this article, supplemented by foundational theoretical works identified through citation tracking.

Analytical Procedure

Retained sources were analysed following the six-phase thematic analysis procedure of Braun and Clarke (2006), adapted here for conceptual rather than empirical data: (1) familiarisation through close reading, (2) generation of initial descriptive codes (e.g., "automatic diacritization," "morphological glossing," "adaptive sequencing"), (3) collation of codes into candidate themes, (4) review of themes against the full set of sources for coherence and distinctiveness, (5) refinement and naming of three final themes corresponding to digital texts, annotation tools, and personalized reading, and (6) integration of the three themes, together with the theoretical perspectives outlined above, into a single conceptual model through iterative concept mapping. To strengthen the trustworthiness of the synthesis, both authors independently generated initial codes for a subset of sources before comparing and reconciling their coding schemes through discussion, a procedure analogous to inter-coder reliability checks in empirical thematic analysis, even though the units being coded here were conceptual and technical claims rather than interview transcripts. Both authors are Arabic-language educators based in Indonesian Islamic higher education, a positionality that informed which gaps in the literature were judged most pedagogically consequential and that is acknowledged here as a feature of, rather than a threat to, the conceptual lens adopted in this review. This analytical procedure is reported in the Findings and Discussion section alongside the substantive synthesis it produced.

Scope and Disciplinary Boundaries

Because mahārah al-qirā'ah is one of four canonical language skills taught in most Arabic curricula alongside listening (istimā'), speaking (kalām), and writing (kitābah), it is necessary to delineate the boundaries of this review explicitly. Sources addressing AI applications to speaking practice, such as pronunciation-feedback systems and conversational chatbots, or to writing assessment, such as automated grammatical error correction, were considered out of scope unless they also bore directly on reading instruction, and were noted only where they illuminated a methodological or theoretical point applicable across skills, as in the case of the chatbot study by Fauzi and Rahmawati (2025) discussed above. Similarly, sources addressing AI in Arabic education broadly, without specific attention to text generation, annotation, or personalization, were excluded even when topically adjacent, since the purpose of this review is to characterise these three affordances in particular rather than to catalogue the full landscape of Arabic EdTech research. This deliberate narrowing is intended to keep the resulting conceptual framework focused and internally coherent, at the acknowledged cost of leaving the integration of reading-focused AI tools with tools for the other three language skills as a question for future conceptual or empirical work.

Findings and Discussion

AI-Powered Digital Texts and Adaptive Reading Materials

Generative AI models, most prominently large language models such as GPT-4 (OpenAI, 2023), can now produce Arabic reading passages calibrated to a specified vocabulary range, sentence complexity, or thematic domain, effectively automating a task that previously required a human materials writer to hand-craft graded readers for each proficiency band. This capability directly addresses a long-standing constraint identified in the Indonesian Arabic-pedagogy literature: the scarcity of levelled reading materials suited to learners at different stages of script and morphological acquisition (Effendy, 2017; Izzan, 2009). A second, Arabic-specific application concerns automatic diacritization: because authentic printed Arabic conventionally omits short vowels, non-native readers face an additional decoding burden largely absent in most Latin-script languages. Neural diacritization systems (Madhfar & Qamar, 2021) can restore tashkīl to unvocalized text with high accuracy, allowing instructors to present learners with fully vocalized versions of authentic texts—Qur'anic, journalistic, or literary—without manually inserting diacritics by hand.

Considered jointly, AI-generated text simplification and automatic diacritization represent two complementary mechanisms for adapting digital texts to a given learner's developmental stage: the former adjusts lexical and syntactic complexity, while the latter adjusts the phonological transparency of the script itself. From the standpoint of schema theory (Rumelhart, 1980), both mechanisms function by lowering the decoding demands of a text so that a larger share of a learner's cognitive resources can be devoted to building propositional and inferential meaning rather than to letter-by-letter and root-by-root decoding. This reasoning echoes the broader argument, advanced in the learning-technology literature, that digital text adaptation is most pedagogically valuable not as a substitute for authentic material but as a scaffold that is gradually withdrawn as the learner's schema for the target script and morphology matures (Mayer, 2021). Practically, this suggests that Arabic reading curricula could draw on AI text generation and diacritization to construct a continuum of texts—from heavily scaffolded, fully vocalized, lexically simplified passages to authentic, unvocalized source texts—through which learners progress as their independent decoding ability develops.

Beyond text simplification and diacritization, AI-powered digital texts can be extended through multimodal pairing of script with sound and image. Text-to-speech (TTS) systems capable of producing natural-sounding Arabic narration allow a digital text to be read aloud synchronously with the visible script, a pairing that, consistent with Mayer's (2021) multimedia learning principles, can reduce the cognitive load associated with simultaneously decoding script and inferring correct pronunciation, particularly for learners still building automaticity in letter and word recognition. For specifically Qur'anic or classical material, recitation-alignment techniques similar in spirit to those used in the Quranic Arabic Corpus project (Dukes & Habash, 2010) can synchronise audio recitation with word-level highlighting of the corresponding script, offering a further scaffold for learners whose listening proficiency outpaces their reading proficiency, a profile commonly reported among Indonesian learners who have memorised substantial Qur'anic material orally before acquiring independent reading

skill. AI-generated illustrations or icons keyed to specific vocabulary items represent a further, if more exploratory, extension of this multimodal logic, offering a visual gloss alongside or instead of a textual one. None of these multimodal extensions has, to the authors' knowledge, been systematically evaluated for Arabic reading instruction specifically, but their convergence around a common design principle—reducing the unimodal burden of script decoding by distributing meaning-making across multiple channels—suggests a coherent direction for future tool development rather than a disconnected list of features.

Intelligent Annotation Tools and Lexical-Morphological Support

A second affordance concerns tools that annotate text in situ, providing learners with on-demand lexical or morphological information without requiring them to consult a separate dictionary. The Quranic Arabic Corpus (Dukes & Habash, 2010) exemplifies this affordance for Classical Arabic: each word in the Qur'anic text is morphologically tagged and linked to its root, part of speech, and grammatical relations, allowing a learner to click on any word and immediately see its morphological analysis. Although developed for a specific corpus, the underlying annotation architecture illustrates a model that could, in principle, be extended to any Arabic text through general-purpose morphological analysers such as Farasa (Abdelali et al., 2016) or CAMeL Tools (Obeid et al., 2020), both of which can segment running Arabic text into roots, stems, and affixes programmatically. For non-native learners, such root-pattern glossing is pedagogically significant because Arabic's derivational morphology means that recognising a word's triliteral root can unlock the meaning of dozens of morphologically related forms; an annotation tool that surfaces this root explicitly therefore does more than provide a single-word gloss—it scaffolds the learner's broader morphological awareness in a way that a conventional bilingual dictionary look-up cannot.

Generic reading-annotation platforms, which allow learners of various languages to click a word for an instant translation while reading, demonstrate that the underlying interaction pattern—click-to-annotate while reading continuous text—is well established and broadly transferable; what remains underdeveloped, the present synthesis suggests, is the Arabic-specific layer that would surface root and pattern information alongside, or instead of, a bare lexical gloss. From the perspective of the interactive-compensatory model (Stanovich, 1980), morphological annotation tools function primarily at the level of decoding support: a learner who struggles to parse an unfamiliar derived form can rely on the annotation as a compensatory aid while continuing to process the surrounding sentence for contextual meaning, rather than abandoning the reading task altogether. Over repeated exposures, the theory predicts, reliance on the annotation should diminish as the learner internalises the root-pattern relationships it makes visible—an empirically testable claim that the present conceptual synthesis surfaces as a priority for future research. A further consideration, consistent with the Technology Acceptance Model (Davis, 1989), is that annotation tools are most likely to be adopted when they are perceived as both useful, by providing morphologically rich and accurate information, and easy to use, by being integrated directly into the reading interface rather than requiring a separate look-up step; tools that interrupt the reading flow risk being perceived as effortful despite their informational value.

A further dimension of intelligent annotation concerns integration with established lexicographic resources. Classic bilingual dictionaries such as Hans Wehr's Arabic-English lexicon have long served as the reference standard against which learners and instructors verify a word's root and meaning, and an AI annotation layer that surfaces a dictionary-grade entry directly within the reading interface, rather than requiring a separate physical or digital dictionary lookup, would substantially lower the procedural cost of consulting authoritative lexical information mid-sentence. Annotation tools can also, in principle, be extended from purely lexical or morphological support toward error-sensitive feedback: by comparing a learner's oral read-aloud output, transcribed via automatic speech recognition, against the source text, a system could flag specific letters, short vowels, or root-pattern confusions that recur across a learner's reading sessions, effectively turning the annotation layer into a diagnostic instrument rather than a purely supportive one. Such error-sensitive annotation would connect directly to the personalization layer discussed in the following section, since diagnosed error patterns are precisely the kind of fine-grained signal that an adaptive system would need in order to target subsequent material effectively. At present, this integration between lexicographic annotation, oral error detection, and personalization remains largely theoretical for Arabic specifically, even though each individual component—dictionary lookup, speech recognition, and adaptive sequencing—has mature precedents in the broader CALL and NLP literature reviewed here.

Personalized Reading Pathways and Adaptive Learning

The third affordance concerns systems that model an individual learner's proficiency and adapt the sequencing, difficulty, or repetition schedule of reading material accordingly. Large-scale language-learning platforms have demonstrated the technical feasibility of this approach: Settles and Meeder's (2016) trainable spaced-repetition model, for instance, uses a learner's response history to predict the probability that a previously encountered lexical item will be recalled correctly, and schedules review accordingly—a mechanism directly transferable to Arabic vocabulary and morphological pattern recognition embedded within reading texts. More broadly, adaptive learning systems draw on interaction data, such as time on task, error patterns, and self-corrections, to construct an evolving learner model that determines which texts, at what difficulty level, a given learner should encounter next (Zawacki-Richter et al., 2019). Applied to mahārah al-qirā'ah, such personalization could, in principle, track not only general proficiency but morphologically specific gaps—for example, identifying that a learner consistently struggles with Form X derived verbs or with distinguishing visually similar letters—and prioritise subsequent reading material that provides additional exposure to precisely those features. This degree of granularity would be difficult for a single instructor to achieve manually across a class of any appreciable size, particularly in the large lecture-style Arabic courses common in Indonesian higher education.

Theoretically, personalization can be understood as an institutional-scale application of the interactive-compensatory model (Stanovich, 1980): rather than leaving compensation entirely to the individual learner's own strategic choices, an adaptive system can deliberately route additional decoding support, or additional context-rich material, to precisely the

processing level at which a given learner is weakest. At the same time, the literature on AI in higher education cautions that the promise of personalization frequently outpaces its realised classroom impact, and that technology adoption alone does not guarantee improved learning outcomes absent careful pedagogical integration (Holmes et al., 2019; Reich, 2020); this caution applies with particular force to Arabic reading instruction, where no personalization algorithm has yet been validated against the specific morphological and orthographic variables identified above.

The granularity of personalization that AI systems make technically feasible also raises questions about which dimensions of a learner's profile should, as a matter of pedagogical judgement rather than technical capability alone, be tracked and acted upon. A learner model restricted to surface-level metrics such as words-per-minute reading speed risks reducing mahārah al-qirā'ah to a narrowly behavioural target, whereas a model that also incorporates morphological error patterns, self-reported confidence, and comprehension-check performance offers a more pedagogically defensible basis for adapting subsequent material, consistent with the multi-dimensional view of reading competence implied by the interactive-compensatory model itself (Stanovich, 1980). A related design question concerns the data infrastructure required to sustain such modelling: continuous interaction logging, the substrate on which most adaptive systems depend, presupposes consistent device access and connectivity that cannot be assumed across all Indonesian higher education contexts, particularly outside major urban centres. Where such infrastructure is unevenly available, personalization risks inadvertently widening rather than narrowing the proficiency gaps it is intended to address, since only learners with reliable access would accumulate the interaction history needed for the system to personalise effectively. This equity consideration is revisited in the Implications and Limitations section below, but is introduced here because it bears directly on whether personalization, as a theoretical affordance, can be expected to function as intended once embedded within the resource constraints of a specific institutional context.

Toward an Integrated Framework: The AI-Mediated Mahārah al-Qirā'ah Model

Synthesising the three affordances above, this study proposes an AI-Mediated Mahārah al-Qirā'ah Model (Figure 1) in which AI-powered digital texts, intelligent annotation tools, and personalized reading pathways operate as three interlocking layers feeding into a single instructional cycle. Digital text adaptation determines the input the learner encounters; annotation tools mediate the learner's moment-to-moment processing of that input; and personalization systems determine which input, at what difficulty, the learner encounters next, based on accumulated evidence of the first two layers' effects. All three layers are theorised to operate jointly on the learner's developing schema for Arabic script and morphology (Rumelhart, 1980), with the cumulative goal of shifting the learner from compensatory, decoding-heavy processing toward more fluent, meaning-focused reading consistent with the interactive-compensatory model (Stanovich, 1980). Table 1 summarises representative technologies corresponding to each layer of the model, illustrating that components of the proposed framework already exist in fragmented form across the CALL and NLP literature, even though they have not previously been integrated for Arabic reading pedagogy specifically.

Table 1. AI Technology Categories Supporting Mahārah al-Qirā'ah

Layer	Function	Representative Technologies / Sources
AI-Powered Digital Texts	Automatic text simplification and leveling; neural diacritization (tashkīl restoration)	GPT-4 (OpenAI, 2023); neural diacritization systems (Madhfar & Qamar, 2021)
Intelligent Annotation Tools	Morphological segmentation; root-pattern glossing; corpus-based morphological tagging	Farasa (Abdelali et al., 2016); CAMeL Tools (Obeid et al., 2020); Quranic Arabic Corpus (Dukes & Habash, 2010)
Personalized Reading Pathways	Learner modeling; spaced-repetition scheduling; adaptive content sequencing	Trainable spaced-repetition model (Settles & Meeder, 2016); adaptive learning analytics (Zawacki-Richter et al., 2019)

Note. Table summarises representative technologies identified through the integrative literature review; inclusion does not imply formal evaluation of these tools for Arabic reading instruction specifically.

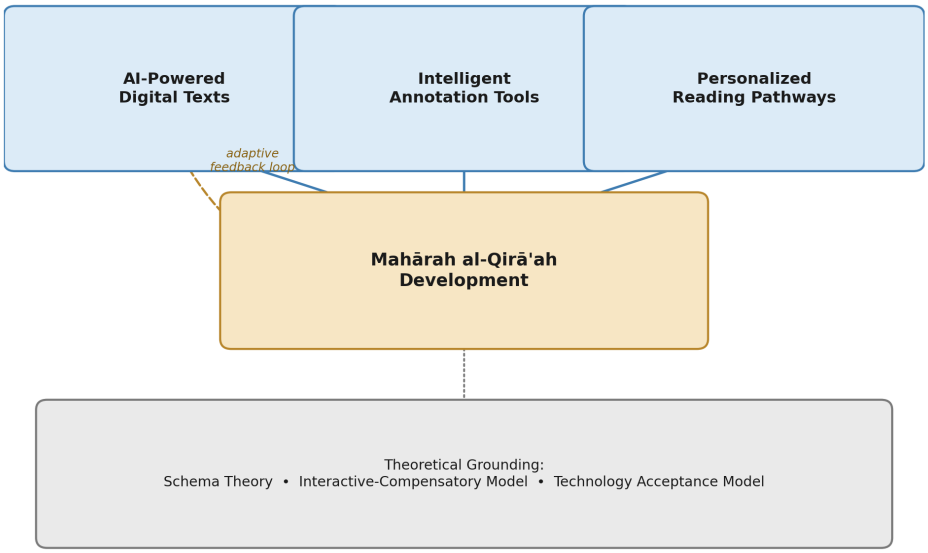

Figure 1. The AI-Mediated Mahārah al-Qirā'ah Model

Illustrative Pedagogical Scenarios Across Proficiency Levels

To make the proposed model more concrete, three brief illustrative scenarios are sketched below, corresponding to beginner, intermediate, and advanced stages of mahārah al-qirā'ah development. These scenarios are conceptual illustrations rather than reports of implemented practice; they are intended to demonstrate how the three layers of the model might jointly recalibrate as a single learner, or a single classroom of learners at different levels, progresses.

At the beginner stage, a learner with minimal prior exposure to the Arabic script might be presented with an AI-generated, fully diacritized passage restricted to a small, high-

frequency vocabulary set (Layer 1). While reading, every word would be paired with an always-visible rather than click-triggered morphological gloss, since at this stage the procedural cost of even a single look-up action may exceed its benefit, consistent with the heavy compensatory reliance on external support that the interactive-compensatory model predicts for low-proficiency readers (Layer 2; Stanovich, 1980). A personalization system would track which letters or short-vowel patterns the learner most frequently mis-reads aloud or mis-selects in comprehension checks, and would weight subsequent passages toward additional exposure to precisely those patterns (Layer 3).

At the intermediate stage, a learner who has internalised basic script recognition but still relies heavily on diacritics might be presented with passages in which diacritization is gradually withdrawn from high-frequency, already-mastered vocabulary while remaining present on newly introduced or low-frequency items (Layer 1), operationalising Krashen's (1985) i+1 principle at the level of individual words rather than whole texts. Annotation would shift from an always-visible gloss to a click-to-reveal format that surfaces root and pattern information on demand (Layer 2), and the personalization system would begin tracking root-recognition accuracy specifically, recommending subsequent texts built around roots the learner has not yet encountered in multiple derived forms, so as to consolidate the root-pattern generalisation that underlies fluent Arabic reading (Layer 3).

At the advanced stage, a learner approaching independent reading fluency might engage primarily with authentic, unvocalized texts, such as short newspaper articles or excerpts from classical commentary, with diacritization and annotation available but rarely invoked except for genuinely rare or ambiguous roots (Layers 1 and 2). Here, personalization shifts in character from remedial support toward thematic extension: rather than targeting weaknesses, the system recommends authentic texts within the learner's academic track, for instance journalistic Arabic for learners oriented toward contemporary affairs or classical commentary for learners oriented toward tafsīr studies, in order to build domain-specific reading fluency and vocabulary depth (Layer 3).

Across all three scenarios, the same underlying technical architecture, comprising AI-adapted digital texts, intelligent annotation, and personalized sequencing, recalibrates as the learner's schema for Arabic script and morphology matures. This illustrates how a single infrastructure could, in principle, support differentiated instruction across a single classroom's full range of proficiency without requiring an instructor to manually prepare parallel versions of every text by hand.

Implications for Curriculum Development and Practice

For curriculum developers and EdTech designers, the proposed model suggests a concrete design priority: integrating text generation, diacritization, morphological annotation, and adaptive sequencing within a single coherent tool rather than as disconnected utilities scattered across separate applications, since the pedagogical value of each layer, as argued above, depends partly on its coordination with the other two. For instructors, the framework positions AI as augmenting rather than replacing the teacher's role in modelling strategic

reading and providing the sociocultural context (Vygotsky, 1978) that current AI systems cannot supply on their own; an AI reading tool can scaffold decoding, but classroom discussion of a text's cultural, religious, or rhetorical significance remains a distinctly human pedagogical contribution. This implies a corresponding professional development need: instructors adopting AI reading tools would benefit from training not only in operating the tools themselves but in interpreting the diagnostic information such tools generate—for instance, distinguishing a learner's genuine morphological misunderstanding from a transient lapse in attention, a distinction the system itself cannot reliably make. At the institutional and policy level, the framework also implies that decisions about adopting AI reading tools should be informed by an explicit assessment of dialect coverage and infrastructure readiness rather than by tool availability alone, given the structural limitations discussed below.

Within the Indonesian higher education context specifically, any implementation must also reckon with equity considerations, since reliable internet connectivity and device access cannot be assumed uniformly across institutions or student populations, a concern already raised in relation to personalization above and one that applies with comparable force to AI-generated digital texts and annotation tools, both of which typically depend on real-time access to cloud-based models or APIs rather than functioning fully offline.

Limitations and Directions for Future Research

A further, structural limitation concerns dialect coverage: the Arabic NLP tools reviewed here, like the chatbot examined in a previous AI-related study published in this journal, are trained predominantly on Modern Standard Arabic, and their handling of dialectal or Classical input remains limited (Fauzi & Rahmawati, 2025); this recurrence across independently developed tools suggests that dialectal limitation is a structural feature of current Arabic AI tooling generally, rooted in the composition of available training corpora, rather than a flaw specific to any single application. Until dialect-inclusive Arabic NLP models become more widely available, AI reading tools built on the affordances reviewed here will be best suited to MSA-medium academic and religious texts, which, fortunately, constitute the majority of texts targeted by mahārah al-qirā'ah instruction in Indonesian higher education, even if they cannot fully address learners' eventual need to read dialect-inflected informal text.

As a conceptual synthesis, this study's own limitations should also be acknowledged. The proposed framework has not been empirically tested, and its claims about reduced decoding load, compensatory scaffolding, and fluency gains, including the illustrative scenarios offered above, remain theoretically derived rather than measured. The review draws predominantly on published, largely English-language NLP and CALL literature, which may underrepresent practitioner knowledge embedded in Arabic-medium or Indonesian-medium sources that were not indexed in the databases searched, and the search, while systematic, was not exhaustive in the manner of a full systematic review. Future empirical work, whether design-based research implementing a prototype tool aligned with the proposed model, or quasi-experimental validation analogous to prior studies in this journal, is needed to test whether the mechanisms proposed here translate into measurable gains in reading fluency and

comprehension, and whether the specific design parameters illustrated in the beginner, intermediate, and advanced scenarios above hold up under classroom conditions.

Conclusion

This conceptual study set out to map and theoretically integrate three AI-driven affordances—digital texts, annotation tools, and personalized reading pathways—into a coherent account of how artificial intelligence might support the teaching of mahārah al-qirā'ah to non-native Arabic learners. Synthesising 28 sources spanning Arabic natural language processing, computer-assisted language learning, and reading pedagogy through an integrative literature review, the study advances an AI-Mediated Mahārah al-Qirā'ah Model in which AI-generated and diacritized digital texts, root-pattern annotation tools, and adaptive personalization systems jointly target the decoding and comprehension demands that make Arabic reading uniquely challenging for learners unfamiliar with its script and morphology. The framework is offered as a conceptual contribution rather than an empirically validated intervention, and its principal limitation is precisely this absence of empirical testing; the cognitive and motivational mechanisms proposed here—reduced decoding load, compensatory scaffolding, and technology-acceptance-mediated adoption—require verification through design-based research or quasi-experimental studies analogous to prior empirical work in this journal. Curriculum developers and EdTech designers are encouraged to use the proposed model as a starting point for prototyping integrated AI reading tools, while instructors are reminded that such tools are best positioned as scaffolds within, rather than substitutes for, classroom-based reading instruction. Future research should prioritise dialect-inclusive AI models, longitudinal assessment of reading fluency gains, and attention to equitable access within the Indonesian higher education context. More broadly, by translating an admittedly fragmented technical literature into a single pedagogically grounded model, this study aims to narrow the distance between what AI-driven CALL and NLP research has already made technically possible and what mahārah al-qirā'ah instruction, as currently practised in much of Indonesian higher education, has so far been able to put into practice.

Acknowledgements

The authors thank colleagues at their respective institutions for their feedback on earlier drafts of this conceptual framework. This research received no specific external funding.

References

Abdelali, A., Darwish, K., Durrani, N., & Mubarak, H. (2016). Farasa: A fast and furious segmenter for Arabic. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations* (pp. 11–16). Association for Computational Linguistics.

Al-Hamad, M. Q., & Al-Shara, I. (2022). Integrating technology in Arabic language teaching: Challenges and opportunities. *International Journal of Arabic-English Studies, 22*(1), 45–68. https://doi.org/10.33806/ijaes2000.22.1.3

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Chapelle, C. A., & Sauro, S. (Eds.). (2017). *The handbook of technology and second language teaching and learning.* Wiley-Blackwell.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008

Dukes, K., & Habash, N. (2010). Morphological annotation of Quranic Arabic. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)* (pp. 2530–2536). European Language Resources Association.

Effendy, A. F. (2017). *Metodologi pengajaran bahasa Arab* (Rev. ed.). Misykat.

Fauzi, A., & Rahmawati, S. (2025). AI-powered chatbots in Arabic vocabulary learning: A quasi-experimental study in Indonesian higher education. *Journal of Arabic EdTech, 2*(1), 1–15. https://doi.org/XXXXX

Godwin-Jones, R. (2021). Big data and language learning: Opportunities and challenges. *Language Learning & Technology, 25*(1), 4–19.

Habash, N. (2010). *Introduction to Arabic natural language processing.* Morgan & Claypool Publishers.

Hassan, A. R., & Ibrahim, S. Y. (2024). Gamification in Arabic vocabulary learning: A quasi-experimental study. *Journal of Arabic EdTech, 1*(1), 1–18. https://doi.org/XXXXX

Holmes, W., Bialik, M., & Fenwick, J. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning.* Center for Curriculum Redesign.

Izzan, A. (2009). *Metodologi pembelajaran bahasa Arab.* Humaniora.

Ismail, Moh. (2019). Design objective tests using the computer program in developing the Arabic language for non-native speakers. *At-Ta'dib*, *14*(2), 1. https://doi.org/10.21111/at-tadib.v14i2.4824

Ismail, Moh., Ahmad, F. S., & Ma'ruf, M. A. (2023). The Impact of Learning Management System "arabi.id" Web-Based Application on Developing Arabic Language Skills. *Al-*

Ta'rib : Jurnal Ilmiah Program Studi Pendidikan Bahasa Arab IAIN Palangka Raya, *11*(2), 213–232. https://doi.org/10.23971/altarib.v11i2.6591

Krashen, S. D. (1985). *The input hypothesis: Issues and implications.* Longman.

Madhfar, M. A., & Qamar, A. M. (2021). Neural Arabic text diacritization: State of the art results and a novel approach for machine translation. *IEEE Access, 9*, 88107–88119. https://doi.org/10.1109/ACCESS.2021.3089733

Mayer, R. E. (2021). *Multimedia learning* (3rd ed.). Cambridge University Press.

Obeid, O., Zalmout, N., Khalifa, S., Taji, D., Oudah, M., Alhafni, B., Inoue, G., Eryani, F., Erdmann, A., & Habash, N. (2020). CAMeL Tools: An open source python toolkit for Arabic natural language processing. In *Proceedings of the Twelfth Language Resources and Evaluation Conference* (pp. 7022–7032). European Language Resources Association.

OpenAI. (2023). *GPT-4 technical report.* arXiv. https://doi.org/10.48550/arXiv.2303.08774

Reich, J. (2020). *Failure to disrupt: Why technology alone can't transform education.* Harvard University Press.

Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 33–58). Lawrence Erlbaum Associates.

Ryding, K. C. (2005). *A reference grammar of Modern Standard Arabic.* Cambridge University Press.

Settles, B., & Meeder, B. (2016). A trainable spaced repetition model for language learning. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* (pp. 1848–1858). Association for Computational Linguistics.

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research, 104*, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039

Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly, 16*(1), 32–71. https://doi.org/10.2307/747348

Torraco, R. J. (2005). Writing integrative literature reviews: Guidelines and examples. *Human Resource Development Review, 4*(3), 356–367. https://doi.org/10.1177/1534484305278283

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

Warschauer, M., & Healey, D. (1998). Computers and language learning: An overview. *Language Teaching, 31*(2), 57–71. https://doi.org/10.1017/S0261444800012970

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1), Article 39. https://doi.org/10.1186/s41239-019-0171-0