Global Education Journal
Vol. 4, 2 (May-August, 2026), pp. 255-270
EISSN: 2986-6529, DOI: https://doi.org/10.59525/gej.1714

Critical AI Literacy in Arabic Language Education: Rethinking Teacher Roles, Ethics, and Curriculum in the Age of Generative AI

Muhammad Fariq Heemal Attruk[1], Nik Hanan Mustapha[1]

[1] International Islamic University Malaysia, Malaysia

ARTICLE INFO

Keywords:

Critical AI Literacy;
Arabic Language Education;
Ethics;
Teacher Roles

Article history:

Received 2026-04-11
Revised 2026-05-11
Accepted 2026-06-16

ABSTRACT

This study examines the growing role of generative artificial intelligence (AI) in Arabic language education and highlights the increasing need for critical AI literacy among learners and teachers. While AI tools offer opportunities to support grammar learning, vocabulary development, writing practice, and independent learning, concerns remain regarding accuracy, overreliance, academic integrity, and the decline of critical thinking. The study aims to explore the need for critical AI literacy in Arabic language learning, examine the changing role of teachers, identify institutional and ethical challenges, and propose a conceptual framework for future practice and research. A Narrative Literature Review (NLR) approach was employed to synthesize and critically interpret relevant studies. Literature was collected from Scopus and Google Scholar, resulting in 26 primary studies that were analyzed through thematic synthesis. The review identified six major themes: critical AI literacy as a pedagogical necessity, the repositioning of teachers, infrastructure and policy challenges, academic ethics and integrity, teacher professional development and AI-responsive curriculum design, and the development of a critical AI literacy framework. Based on these findings, the study proposes a preliminary framework consisting of evaluative linguistic skills, awareness of AI limitations, human–AI collaboration, the use of morphology-aware and adaptive AI tools, and the integration of digital technologies with sound pedagogical practices. The study concludes that effective AI integration in Arabic language education requires critical engagement, informed human judgment, teacher preparedness, and context-sensitive educational policies.

Corresponding Author:
Muhammad Fariq Heemal Attruk
International Islamic University Malaysia, Malaysia; fariqhemal@gmail.com

1. INTRODUCTION

Generative artificial intelligence (AI) is increasingly reshaping educational practice, including the ways languages are taught and learned. Tools such as ChatGPT, Gemini, and other AI-powered platforms are now widely used by students and teachers. They offer immediate responses, personalized feedback, and flexible access to learning materials. Studies have shown that generative AI can support grammar learning, vocabulary development, writing practice, and independent study (Rahmaddani & Naifah, 2025; Zubaidi et al., 2025; Albantani et al., 2025). Despite these advantages, researchers have also identified several concerns. AI tools may produce inaccurate information, and learners can become overly dependent on them (Karataş et al., 2024; Law, 2024). There are also worries about academic dishonesty, plagiarism, and the weakening of critical thinking skills when students rely too much on AI (Perkins, 2023; Alkaabi & Almaamari, 2025). These concerns have led to a growing discussion about the need for AI literacy—the ability to understand, use, and critically evaluate AI tools in learning contexts.

In response to these concerns, scholars have proposed various definitions and conceptualizations of AI literacy. Pang et al. (2025) described AI literacy as encompassing not only the ability to understand and use AI effectively, but also the capacity to evaluate its outputs critically and consider its ethical and cultural implications. They argued that AI literacy has become a necessity in modern language education, identifying key competencies such as prompt engineering, critical evaluation of AI outputs, and ethical awareness. Yet the ability to use AI effectively does not necessarily imply the ability to evaluate its outputs critically. For this reason, the concept of critical AI literacy has gained attention. Wu et al. (2025) discussed a Critical AI Literacy (CAIL) framework specifically tailored for students of applied linguistics and language education. Their framework includes technical understanding, critical thinking, ethical awareness, and practical application. This emphasis on criticality is particularly relevant because AI-generated responses may appear fluent and convincing despite containing factual or linguistic inaccuracies. Learners who lack critical AI literacy may accept inaccurate information without question, which can lead to misconceptions and passive learning habits (Linur et al., 2025; Al-Jamali & Abdalla, 2025).

As AI becomes more integrated into education, the role of teachers is also changing. As AI increasingly mediates access to information, teachers are being encouraged to rethink their traditional roles in the learning process. Instead, they are expected to become mediators who help students evaluate AI-generated content, facilitators who design meaningful learning experiences with AI, and ethical guardians who ensure responsible use of technology (Creely & Carabott, 2025; Nugraha & Syafe'i, 2025). Such a repositioning, however, presents new professional and pedagogical challenges. Many language instructors feel uncertain about how to integrate AI effectively and worry about being replaced by technology (Kohnke et al., 2023). They need institutional support, training, and clear guidelines to perform these new roles. At the same time, educational systems must address broader issues such as infrastructure, equity, and policy if AI is to be used ethically and effectively (Jayasinghe et al., 2026; Adawiyah, 2025).

While these discussions have gained considerable attention in general language education and English language teaching, the specific context of Arabic language learning has received comparatively less attention. Arabic has a highly complex grammatical system, characterized by a rich morphology built on root-and-pattern formations, flexible syntax, and context-dependent grammatical rules (Alayba, 2025; Boulesnam & Boucetti, 2025). The learning of *nahwu* (syntax) and *shorof* (morphology) has long been recognized as one of the most challenging aspects of Arabic language study (Abidin & Sain, 2025). Recent studies suggest that current generative AI systems still struggle to handle many aspects of this linguistic complexity. Even advanced models like GPT-4o achieve only partial accuracy on Arabic grammar tasks, and specialized Arabic models perform even worse (Mubarak et al., 2026). AI systems frequently make errors in morphological analysis, case endings, and the handling of diacritics, which are essential for understanding Arabic texts (Tamam et al., 2024; Al-Jarf, 2025; Karima et al., 2025). These limitations are not merely technical; they also carry important pedagogical consequences. When learners use AI to learn

Arabic grammar without adequate critical skills, they may internalize incorrect rules and develop misconceptions (Othman & Asbulah, 2025; Rahmouni, 2024).

Despite the growing body of research on AI in Arabic language education, relatively little attention has been given to developing a comprehensive framework for critical AI literacy that addresses the unique linguistic and cultural dimensions of Arabic language learning. Existing studies tend to focus either on the pedagogical benefits of AI (e.g., Sa'idah et al., 2024; Khouli et al., 2026) or on its technical limitations (e.g., Adel et al., 2026; Alhafni & Habash, 2025). Only a few works have touched upon the need for critical engagement with AI in Arabic grammar learning (Nugraha & Syafe'i, 2025; Dahlan et al., 2023). Despite these contributions, the literature remains fragmented, with limited efforts to integrate pedagogical, ethical, and technological perspectives into a coherent approach to critical AI literacy in Arabic language education. This gap is particularly important in Arabic language education, where the complexity of grammatical structures and the current limitations of AI increase the need for learners to evaluate AI-generated outputs critically.

In response to this gap, the present study conducts a narrative literature review to synthesize recent research on generative AI in Arabic language learning. Accordingly, this review examines the need for critical AI literacy in Arabic language education, explores the changing role of teachers in AI-supported learning environments, identifies the institutional and ethical challenges associated with AI integration, and proposes a conceptual framework to guide future practice and research. By bringing these perspectives into dialogue, this study proposes a preliminary framework that may inform future curriculum development, teacher training, and educational policy. Ultimately, the review seeks to contribute to ongoing discussions about how generative AI can be integrated into Arabic language education in ways that promote informed judgment rather than passive dependence.

2. METHODS

This study employed a Narrative Literature Review (NLR) approach to examine the emerging discourse on critical AI literacy in Arabic language education and to develop a conceptual understanding of its implications for teachers, ethics, and curriculum design. A narrative review was considered appropriate because the topic remains relatively new, multidisciplinary, and conceptually fragmented. Unlike systematic reviews that seek to answer narrowly defined questions through exhaustive procedures, narrative reviews allow researchers to synthesize diverse forms of evidence, critically interpret findings, identify emerging themes, and generate new conceptual perspectives (Baumeister & Leary, 1997; Ferrari, 2015). This approach is particularly suitable when the objective extends beyond summarizing previous studies to exploring broader theoretical and pedagogical implications (Torraco, 2005; Snyder, 2019).

The literature search was conducted using two major academic sources: Scopus and Google Scholar. Scopus was selected because it provides access to peer-reviewed international research with established quality standards, while Google Scholar was used to broaden coverage and capture relevant studies that may not yet be indexed in Scopus, particularly in emerging areas such as generative AI and Arabic language education. Search terms included combinations of keywords such as “generative AI,” “artificial intelligence,” “ChatGPT,” “Arabic language learning,” “Arabic language education,” “*nahwu*,” “*shorof*,” “AI literacy,” “teacher roles,” “curriculum,” and “academic ethics.”

Following the recommendations of narrative review methodology (Ferrari, 2015; Ahmad, 2025), a purposive and iterative selection process was employed. Studies were included if they: (1) focused on the use of artificial intelligence or generative AI in Arabic language learning and teaching; (2) discussed pedagogical, ethical, curricular, institutional, or teacher-related implications of AI integration; and (3) were published in peer-reviewed journals, conference proceedings, or other reputable academic sources. This process resulted in a final corpus of 26 primary studies that were considered directly relevant to the objectives of the review and formed the basis of the thematic analysis. Additional methodological and conceptual references were consulted to support the theoretical framing and interpretation of findings. Consistent with the nature of narrative reviews, the emphasis was placed on

selecting studies that offered conceptual richness and relevance rather than achieving exhaustive coverage (Grant & Booth, 2009).

The selected studies were analyzed through thematic synthesis. Repeated reading and comparison of the literature enabled the identification of recurring concepts, areas of convergence, and critical tensions across studies. Following the recommendations of Torraco (2005), the analysis moved beyond descriptive summarization by critically examining assumptions, limitations, and unresolved issues within the literature. This process generated six major themes, which subsequently informed the development of a preliminary framework for Critical AI Literacy in Arabic Language Education. Rather than presenting a definitive model, the proposed framework serves as a conceptual foundation for future empirical investigation and theoretical refinement.

3. FINDINGS AND DISCUSSION

3.1. Critical AI Literacy as a Pedagogical Necessity

The growing use of generative AI in Arabic language learning has created a new challenge for educators and learners. On one hand, AI tools are widely accepted as helpful learning assistants. On the other hand, research shows that learners often accept AI-generated outputs without sufficient critical evaluation. This situation makes critical AI literacy an urgent pedagogical need. Critical AI literacy means that learners must develop the ability to question, verify, and reflect on the information provided by AI, rather than simply accepting it as correct.

Several studies reveal that AI tools, although useful, are not always accurate in handling Arabic language tasks. Mubarak et al. (2026) developed the *Nahwu* benchmark to test how well language models understand Arabic grammar. They found that the best-performing model, GPT-4o, achieved 67% average accuracy across all tasks. An Arabic specific model, ALLaM-7B, scored even lower at 42%. These results show that current AI systems still have significant gaps in Arabic grammar competence. If learners do not know about these limitations, they may treat AI explanations as fully trustworthy. Linur et al. (2025) also reported that while students found AI helpful for understanding grammar and vocabulary, some students noted that AI gave inaccurate word meanings or sentence translations, especially for culturally specific expressions. This shows that learners are already exposed to errors, but without proper training, they may not always recognize them.

The problem is not only about accuracy. It is also about how learners interact with AI. Albantani et al. (2025) found that generative AI was effective for vocabulary acquisition, but its role in developing writing skills and maintaining focus was limited. They explicitly stated that "intentional instructional guidance is necessary to address its limitations, especially for complex tasks like morphological analysis and written production." This statement highlights that AI cannot replace human judgment. Learners must be guided to use AI as a starting point, not as a final answer. Magdy et al. (2024) introduced the Gazelle dataset for Arabic writing assistance and argued for a "Human AI co writing" approach. Their work suggests that AI should work together with human oversight, where learners remain responsible for checking grammar, style, and accuracy. This collaborative mindset is an important part of critical AI literacy.

Another reason why critical AI literacy is needed is that AI systems often struggle with cultural and contextual aspects of language. Sibaee et al. (2025) evaluated several language models on complex Arabic tasks and found that even the best model, Claude 3.5 Sonnet, achieved only 30% accuracy. All models had difficulty with culturally nuanced topics. This means that AI is not yet capable of fully understanding the cultural dimensions of Arabic, which are deeply connected to its grammar and expressions. If learners are not aware of this, they may accept AI outputs that are linguistically correct but culturally inappropriate.

For critical AI literacy to become a reality, support at the institutional level is also needed. Nugraha and Syafe'i (2025) emphasized that AI should be integrated into Arabic language curricula in a way that reflects Islamic spiritual values and national educational goals. They stressed the importance of

investing in teacher training and ensuring equitable digital access. This shows that critical AI literacy cannot be left to individual students to develop on their own. It must be systematically taught and supported by the educational system. Adawiyah (2025) also found that although lecturers in Islamic higher education viewed AI positively, they faced several barriers. These included inadequate technological infrastructure, limited student preparedness, and difficulty in adapting to digital tools. Such challenges make it difficult to foster critical engagement with AI in the classroom.

More fundamentally, the growing emphasis on critical AI literacy may reflect assumptions derived largely from Western traditions of critical pedagogy. Whether these assumptions can be transferred directly into Arabic language education remains an open question. In many contexts, Arabic grammar instruction has historically been associated with the transmission of established knowledge and the mastery of authoritative texts. Consequently, fostering a habit of questioning AI-generated explanations may require more than technical training; it may demand a shift in underlying educational epistemologies. Future research should therefore examine how critical AI literacy can be adapted to the linguistic, cultural, and pedagogical traditions of Arabic language education rather than simply imported from existing AI literacy frameworks.

Taken together, these findings show that critical AI literacy is no longer an optional skill. Learners need to know how to evaluate AI outputs, recognize the limitations of AI models, and engage with AI as a learning partner rather than an unquestioned authority. At the same time, teachers and institutions must create the conditions that make critical engagement possible. Without these competencies and supports, the risks of misconceptions and passive learning will continue to grow. The next section will discuss how the role of teachers must be repositioned to foster this critical and reflective use of AI.

3.2. Repositioning the Role of Teachers: Mediator, Facilitator, and Ethical Guardian

The growing integration of generative AI into Arabic language learning does not diminish the importance of teachers. Rather, it reshapes and, in many respects, strengthens their pedagogical role. The reviewed studies show that the role of teachers becomes even more essential. As AI tools handle routine tasks such as providing definitions or correcting surface errors, teachers are repositioned as mediators who help students evaluate AI outputs, facilitators who design meaningful learning with AI, and ethical guardians who ensure responsible use of technology.

One of the most important new roles for teachers is acting as mediators between AI-generated content and accurate linguistic knowledge. Abidin and Sain (2025) emphasized that the integration of generative AI into Arabic learning faces serious obstacles. They found that AI often fails to maintain the integrity of *nahwu* and *shorof*, producing syntactic inconsistencies. They argued that the TPACK (Technological Pedagogical Content Knowledge) framework is essential for training teachers to combine technology with strong content and pedagogical knowledge. With this training, teachers can critically evaluate AI-generated content before it reaches students. Samiya (2025) also noted that while intelligent tutoring systems and automated error detection tools are available, they still require teacher supervision to ensure that the feedback given to learners is correct and pedagogically appropriate. Without teachers acting as mediators, students may accept inaccurate grammatical explanations without realizing it.

Nevertheless, the growing emphasis on teacher preparedness raises an important question about the practicality of these expectations. Much of the literature assumes that frameworks such as TPACK can equip teachers to mediate AI-supported learning effectively. Yet this assumption may overlook the realities of many Arabic language classrooms, where teachers have had limited exposure to AI technologies and may be encountering such tools for the first time. Moreover, the suitability of TPACK itself warrants further reflection. Originally developed within broader educational contexts, the framework may not fully account for the pedagogical and epistemological traditions of Arabic language education, particularly in *madrasah* and *pesantren* settings. This suggests that teacher repositioning may require not only technical training but also a reconsideration of how existing educational traditions can engage with emerging AI literacies.

Teachers are also expected to become facilitators who guide learners in using AI effectively and reflectively. Nugraha and Syafe'i (2025) stressed that AI should be integrated into Arabic language curricula in a way that reflects Islamic spiritual values and national educational goals. They argued that culturally responsive AI integration can increase student engagement and retention, but only when teachers are prepared to design and oversee such integration. Adawiyah (2025) also found that although lecturers in Islamic higher education viewed AI positively, many faced difficulties in adapting to digital tools due to inadequate infrastructure and limited preparedness. These findings suggest that effective AI-supported learning is unlikely to emerge automatically from the availability of technology alone. Teachers need institutional support and continuous professional development to carry out this role effectively.

Beyond mediation and facilitation, teachers serve as ethical guardians in AI-supported learning environments. Law (2024), in a broad review of generative AI in language education, found that alongside the benefits, serious concerns have been raised about academic dishonesty, plagiarism, and the weakening of critical thinking skills. These concerns apply directly to Arabic language learning, where students may use AI to complete grammar exercises or writing assignments without genuine effort. Al-Jamali and Abdalla (2025) found that while students engaged actively with AI for grammar correction, concerns about content accuracy and usability prevented them from fully trusting AI outputs. This cautious attitude suggests that students continue to rely on teachers to help navigate the often-blurred boundaries between legitimate academic support and academic misconduct.

The ethical role of teachers also involves preserving the cultural and spiritual dimensions of Arabic language learning. Moustafa et al. (2026) highlighted that the adoption of generative AI in Arabic education must be accompanied by ethical considerations and culturally sensitive approaches. AI tools, which are often developed with English or Western contexts in mind, may not fully capture the values embedded in Arabic and Islamic education. Dahlan et al. (2023) showed that AI can effectively improve Arabic linguistic competence, particularly in *nahwu* and *sharf*, but challenges remain regarding pronunciation accuracy, dialectal variation, and cultural relevance. They argued that AI should be used to support the integration of Islamic values rather than replace the human and spiritual elements of learning. Teachers, in this view, are not only language instructors but also guardians of the cultural and spiritual context in which Arabic is learned. AI cannot replicate this dimension, which makes the teacher's presence essential.

Taken together, these findings suggest that the expansion of AI in Arabic language education does not marginalize teachers; rather, it redefines their professional responsibilities. The teacher's role increasingly extends beyond transmitting linguistic knowledge to mediating AI-generated information, facilitating meaningful learning experiences, and safeguarding academic and cultural values. At the same time, the literature raises important questions about whether existing teacher preparation models are sufficient for these emerging demands. As AI becomes more embedded in educational practice, the challenge is not merely to train teachers to use technology, but to prepare them to exercise informed pedagogical and ethical judgment in environments where technological outputs may appear authoritative but remain fallible.

3.3. Infrastructure, Institutional Readiness, and Policy Challenges

The previous section discussed how teachers must be repositioned as mediators, facilitators, and ethical guardians in AI supported Arabic language learning. However, even the most skilled and motivated teachers cannot perform these roles effectively if the educational system does not provide the necessary infrastructure, institutional readiness, and policy support. The studies reviewed here suggest that the challenges of integrating generative AI into Arabic language education extend far beyond individual classrooms and reflect broader systemic constraints.

One of the most detailed accounts of these obstacles comes from Adawiyah (2025), who studied the implementation of AI in Islamic higher education. She identified five major barriers: inadequate technological infrastructure, limited student preparedness, difficulty among lecturers in adapting to

digital tools, scarcity of Arabic specific AI datasets, and technical limitations of AI applications. These findings suggest that access to AI tools alone is insufficient for meaningful implementation. Institutions may lack reliable internet, updated devices, or the technical support needed for smooth AI integration. Without addressing these basic needs, calls for critical AI literacy and teacher repositioning will remain theoretical.

The problem of resource scarcity is echoed by other researchers. Samiya (2025) noted that the development of AI for Arabic is held back by limited high quality linguistic resources. Arabic datasets are smaller and less diverse compared to those available for languages like English. This scarcity affects the accuracy of AI tools used in educational settings. Doohee (2024) also pointed out that there is a lack of widespread Arabic content online and poor coordination among institutions working on Arabic language technology. Many existing programs have not yet reached full computational and cognitive accuracy. As a result, even highly motivated teachers and learners may struggle to engage critically with AI when the underlying tools remain constrained by linguistic and technical limitations.

Abidin and Sain (2025) added another dimension to this challenge by highlighting the digital divide. They argued that the effectiveness of digital innovation in Arabic language learning depends on mitigating this divide. Not all institutions have the same level of access to technology, and this inequality affects how AI can be adopted. Their study also emphasized the importance of the TPACK framework for teacher training, but they noted that such training can only be effective if institutions have the technological foundation to support it.

The policy level is equally important. Nugraha and Syafe'i (2025) called for the formalization of AI integration into national Arabic language curricula. They argued that Indonesia needs to invest in teacher training and ensure equitable digital access so that AI can be used in a way that is pedagogically sound and culturally appropriate. Their comparative analysis with Malaysia, the UAE, and Saudi Arabia showed that successful AI integration depends on tailoring policies to specific cultural, ideological, and pedagogical contexts. Without clear policies and institutional commitment, AI adoption in Arabic language education will remain uneven and unguided.

At the same time, the growing emphasis on AI policy raises important questions about how such policies are formulated and implemented. Much of the current discourse focuses on technological infrastructure and digital access, yet comparatively less attention is given to what might be termed intellectual infrastructure: the professional cultures, pedagogical traditions, and institutional capacities needed to support meaningful AI integration. There is also a risk of policy borrowing, where educational institutions adopt AI frameworks and strategies developed in different cultural and educational contexts without sufficient adaptation. In Arabic language education, where linguistic, religious, and pedagogical traditions play a central role, the effectiveness of imported AI policies cannot be taken for granted. Without active involvement from teachers and local educational stakeholders, AI policies may remain largely symbolic, producing compliance on paper rather than meaningful pedagogical change in practice.

Taken together, these findings suggest that the challenges surrounding AI in Arabic language education are not merely technological but systemic in nature. Efforts to promote critical AI literacy and redefine teachers' roles are unlikely to succeed without parallel investments in infrastructure, institutional capacity, linguistic resources, and context-sensitive policy development. More importantly, the literature indicates that sustainable AI integration depends not only on technological readiness but also on the ability of educational institutions to align innovation with their own pedagogical and cultural realities.

3.4. Academic Ethics and Integrity in AI-Assisted Arabic Learning

As generative AI becomes more embedded in Arabic language education, questions surrounding academic ethics and integrity have become increasingly difficult to ignore. The reviewed studies show that while AI can support learning, it also opens the door to academic dishonesty, plagiarism, and a

weakening of critical thinking. Addressing these ethical concerns is an essential part of building critical AI literacy and repositioning teachers as ethical guardians.

One of the most direct discussions of this issue comes from Law (2024), who conducted a scoping review of generative AI in language teaching and learning. The review found that researchers generally held positive attitudes toward AI, but they also raised serious concerns about academic dishonesty, plagiarism, and the hindrance of critical thinking skills. These concerns are not limited to English or general education; they apply directly to Arabic language learning, where students may use AI to complete grammar exercises, generate essays, or analyze texts without genuinely engaging with the material. Without clear ethical guidelines, learners may cross the line between using AI as a learning aid and using it as a shortcut to avoid real effort.

The risk of overreliance on AI was also observed in Arabic language classrooms. Alkaabi and Almaamari (2025) studied how generative AI functioned as a teaching assistant in Arabic language courses. Instructors reported that many students used AI for basic grammar exercises and relied too heavily on AI for writing assignments. This made it difficult for teachers to assess genuine learning. When students submit AI-generated work as their own, the purpose of assessment is undermined. Teachers cannot know whether the student has truly mastered a grammatical concept or simply copied an AI output. This situation highlights the need for clear institutional policies on acceptable AI use, as well as teacher training on how to detect and respond to AI-assisted academic dishonesty.

At the same time, the distinction between legitimate assistance and academic misconduct is becoming increasingly difficult to define in AI-supported learning environments. If a student uses AI to check case endings, refine *i'rab* analysis, or improve grammatical accuracy before submitting an assignment, it is not always clear whether this constitutes appropriate learning support or an unfair academic advantage. This ambiguity raises broader questions about whether traditional definitions of plagiarism and authorship remain fully applicable in the age of generative AI. Rather than viewing AI use solely through the lens of compliance and violation, future discussions may need to reconsider how learning, originality, and intellectual contribution are understood in technology-mediated educational contexts.

Despite these risks, students themselves are not entirely uncritical. Rahmouni (2024) found that the majority of students viewed ChatGPT as a supplementary tool rather than a standalone one for comprehensive learning, preferring traditional human instruction. This finding suggests that learners are not simply passive consumers of AI-generated information. They have a basic awareness of ethical boundaries. However, this awareness needs to be reinforced and guided by teachers, because the line between appropriate help and dishonest use is not always clear.

A broader ethical perspective is offered by Moustafa et al. (2026), who reviewed the opportunities, challenges, and ethical considerations of generative AI in Arabic learning and teaching. They argued that the adoption of AI must be accompanied by culturally sensitive approaches. Ethical considerations should not be treated as an afterthought but as a central part of integrating AI into education. This includes respecting the cultural and religious values embedded in Arabic language learning. For example, an AI tool might suggest phrasing that is grammatically correct but culturally inappropriate for an Islamic educational context. Teachers, therefore, play an important role in helping students understand not only what is linguistically correct, but also what is ethically and culturally acceptable.

Another ethical concern that receives comparatively less attention in the literature relates to the biases embedded within AI training data. Generative AI systems are trained on particular linguistic corpora and therefore may privilege certain grammatical conventions, textual traditions, or language varieties over others. In the context of Arabic language learning, this raises questions about how less dominant grammatical interpretations, regional linguistic practices, or alternative pedagogical traditions are represented. Ethical AI use therefore involves not only preventing academic dishonesty but also developing awareness of whose linguistic knowledge is being amplified and whose may be marginalized through AI-generated outputs.

In summary, the reviewed studies suggest that academic ethics in AI-assisted Arabic learning cannot be reduced to simple questions of rule compliance. While concerns about dishonesty and overreliance remain important, the literature also reveals deeper questions about authorship, intellectual responsibility, and the changing boundaries between human and AI contributions to learning. Addressing these challenges will require not only clearer institutional guidelines but also ongoing dialogue among educators, students, and policymakers regarding what constitutes meaningful and ethical learning in the age of generative AI.

3.5. Teacher Professional Development and AI-Responsive Curriculum Design

For teachers to fulfill their roles as mediators, facilitators, and ethical guardians, professional development and curriculum adaptation have become increasingly important. The reviewed studies show that teacher training and thoughtful curriculum design are not optional extras; they are necessary foundations for the responsible integration of generative AI into Arabic language education.

One of the most important frameworks identified in the literature is TPACK, which stands for Technological Pedagogical Content Knowledge. Abidin and Sain (2025) argued that TPACK is essential for training Arabic language teachers to combine advanced technology with strong content and pedagogical knowledge. Without this combination, teachers may know how to use AI tools but lack the ability to evaluate their outputs critically or integrate them meaningfully into grammar instruction. More importantly, the framework encourages teachers to think beyond the technical operation of AI and consider how it can be integrated into language instruction in pedagogically meaningful ways.

The direct impact of teacher training on learning outcomes has been measured empirically. Sa'idah et al. (2024) investigated the integration of AI in Arabic language teaching and found that teacher training significantly enhanced both AI usage and student learning. Their analysis showed that teacher training had a direct effect on AI usage with a path coefficient of 0.50, and it also directly boosted learning outcomes with a path coefficient of 0.30. These findings provide empirical support for the argument that teacher preparation plays a significant role in shaping the effectiveness of AI integration. When teachers are trained, they use AI more effectively, and their students learn better.

Nugraha and Syafe'i (2025) also emphasized the importance of investment in teacher training and equitable digital access. They reported on a pilot project in Indonesia where culturally responsive AI integration led to a 77% rise in student engagement and a 67% increase in long-term retention. These results show that when AI is integrated thoughtfully—with attention to cultural and pedagogical values—it can have a powerful positive effect. However, the authors stressed that such success depends on formalizing AI integration in national curricula and ensuring that teachers receive proper training.

While these findings are encouraging, they also invite a degree of caution. Successful pilot projects often demonstrate what is possible under favorable conditions, but it remains uncertain whether such outcomes can be sustained over time or replicated across diverse educational settings. In some cases, increased engagement may partly reflect the novelty of interacting with new technologies rather than a lasting improvement in learning. More fundamentally, there is a risk of viewing AI as a solution to challenges that are pedagogical in nature. Difficulties in learning *nahwu* and *shorof*, for example, may stem not only from limited instructional resources but also from deeply rooted teaching practices, assessment cultures, and learner beliefs about language learning. Without broader pedagogical reform, the integration of AI may do little more than digitize existing forms of rote learning rather than transform them.

An important insight from the literature is that curriculum design should not abandon traditional methods. Instead, effective curricula can combine traditional approaches with digital tools. Baharun et al. (2025) studied the *I'rab* method of Habib Hasan Baharun, a traditional orally based approach to teaching Arabic syntax. While the method itself is not technological, the authors noted that it can be complemented with digital reinforcement such as audio recordings, voice recognition exercises, and digital flashcards. This example illustrates a broader principle: AI-responsive curriculum design does

not mean replacing everything with technology. It means thoughtfully blending traditional pedagogical strengths with digital tools to create richer learning experiences.

The use of adaptive learning systems represents another promising direction for curriculum design. Akhtar et al. (2025) developed an AI-powered adaptive learning platform for Classical Arabic that uses natural language processing and reinforcement learning algorithms. The system tracks learners' morphological and syntactic skills in fine detail and adjusts the difficulty of exercises and instructional approaches in real time. Their preliminary results showed statistically significant gains in grammatical accuracy among learners using the adaptive system, particularly for those with low metacognitive awareness. This kind of technology, when embedded in a well-designed curriculum and guided by trained teachers, can personalize learning while still allowing space for critical reflection.

Taken together, the reviewed studies suggest that teacher professional development and curriculum design remain central to the development of critical AI literacy in Arabic language education. However, the literature also indicates that technological innovation alone is unlikely to transform learning outcomes. The effectiveness of AI depends largely on how it is embedded within broader pedagogical practices, cultural contexts, and institutional structures. Consequently, the challenge is not simply to introduce AI into Arabic language classrooms, but to ensure that its use supports deeper learning, critical engagement, and meaningful educational change.

3.6. Toward a Framework of Critical AI Literacy for Arabic Language Learning

The findings discussed in the previous sections point to a clear need for a structured approach to critical AI literacy in Arabic language education. Rather than treating critical engagement with AI as an optional skill, the reviewed literature suggests that it should be systematically developed. Based on the synthesis of the studies in this review, several key components of a critical AI literacy framework can be identified. Rather than presenting a definitive model, the framework should be understood as a starting point for further discussion and development.

The first component is the ability to evaluate AI outputs across multiple linguistic levels. Al-Jarf (2025) examined specific linguistic questions that AI cannot answer accurately and recommended that students be trained in five areas: phonological verification, morphological analysis, lexical validation, pragmatic awareness, and bibliographic literacy. These skills enable learners to cross-check AI-generated content against established knowledge rather than accepting it at face value. In Arabic language learning, this means that when AI provides a grammatical analysis, learners should be able to compare it with trusted sources, identify possible errors, and seek clarification when needed. This evaluative ability is the foundation of critical AI literacy.

The second component is awareness of the actual performance limits of AI models. Mubarak et al. (2026) developed the *Nahwu* benchmark and showed that even advanced models achieve only partial accuracy on Arabic grammar tasks. Familiarizing learners and teachers with such benchmarks can help challenge the common assumption that AI outputs are inherently reliable. This awareness can be integrated into classroom practice, where students compare AI analyses with their own understanding or with teacher feedback. When learners see for themselves where AI makes mistakes, they develop a more realistic understanding of what AI can and cannot do.

The third component involves adopting a human-AI collaboration mindset. Zubaidi et al. (2025) proposed a tridimensional human-AI collaboration model in which AI functions as a tool, while humans act as conceptors, validators, and executors. In this model, AI is not the final authority but a support system that requires human judgment at every stage. This approach is echoed in the work of Chirkunov et al. (2025), who developed ARWI (Arabic Write and Improve), a system that provides targeted feedback on Arabic grammar while leaving final decisions to the human user. Similarly, Nandatasia and Nugrahawan (2026) showed that AI served best as a complementary tool in writing instruction, helping students apply grammatical structures more accurately while the teacher remained in control of the learning process. Together, these examples illustrate how human judgment can remain central even within AI-supported learning environments.

The fourth component is the use of morphology-aware and adaptive AI tools within guided learning environments. Bourouba (2025) argued that an authentic Arabic AI model must move beyond translation-based adaptation to concept-driven design that incorporates morphology-aware processing. This means that AI tools designed for Arabic should be built with the root-and-pattern system in mind, not simply adapted from English models. Al-Said (2026) introduced Mizan, a computational tool that automatically extracts and analyzes Arabic morphological patterns with high accuracy. Tools like Mizan provide learners and teachers with reliable morphological information that can support critical analysis. Akhtar et al. (2025) presented an adaptive learning platform that tracks learners' morphological and syntactic skills and adjusts content accordingly. When such systems are used under teacher supervision, they can personalize learning while still allowing room for critical reflection.

The fifth component is the thoughtful integration of digital tools with sound pedagogical methods. Baharun et al. (2025) showed that the traditional *I'rab* method, while not technological, can be complemented with digital tools such as audio recordings, voice recognition exercises, and digital flashcards. This blending of traditional and digital approaches preserves the strengths of established pedagogical methods while adding the benefits of technology. Borham et al. (2024) stressed that developing AI based Arabic learning software requires careful handling of morphological and syntactic ambiguities. This implies that user interface and experience design must actively support learners in navigating these grammatical challenges rather than ignoring them.

At the same time, this proposed framework should be viewed with appropriate caution. Although it is grounded in findings from the reviewed literature, much of the existing research on AI in Arabic language education has been conducted in higher education settings and often reflects contexts with relatively greater access to digital resources. As a result, the applicability of these components to other educational environments, such as secondary schools, *madrasahs*, or *pesantren*, cannot be assumed. Furthermore, the framework remains conceptual rather than empirically validated. Its effectiveness will depend on how it is adapted to specific linguistic, cultural, and institutional contexts. Future research should therefore examine how these components interact in practice and whether additional dimensions emerge in educational settings that are currently underrepresented in the literature.

Taken together, these five components provide a preliminary framework for conceptualizing critical AI literacy in Arabic language learning. Rather than offering a universal solution, the framework seeks to organize key insights emerging from the literature and translate them into practical directions for curriculum development, teacher preparation, and learner support. It also highlights the multidimensional nature of critical AI literacy, which depends not only on learner competencies but also on the contributions of teachers, institutions, and technology developers. As research in this area continues to evolve, the framework should be refined, tested, and adapted across diverse educational contexts.

4. CONCLUSION

This narrative literature review examined the growing role of generative AI in Arabic language education and explored the implications of its integration for learners, teachers, institutions, and curriculum development. The review identified six major themes: the necessity of critical AI literacy, the repositioning of teachers as mediators, facilitators, and ethical guardians, the importance of institutional readiness and policy support, concerns related to academic ethics and integrity, the need for teacher professional development and AI-responsive curriculum design, and the development of a preliminary framework for critical AI literacy. Collectively, these findings suggest that while generative AI offers significant opportunities for enhancing Arabic language learning, its effective use requires more than technical competence. Learners must develop the capacity to critically evaluate AI-generated outputs, while teachers and institutions must create conditions that support responsible, reflective, and contextually appropriate AI use.

A key contribution of this review is the proposal of a preliminary framework for Critical AI Literacy in Arabic Language Education. The framework highlights five interconnected components: evaluative linguistic skills, awareness of AI limitations, human–AI collaboration, the use of morphology-aware and adaptive AI tools, and the integration of digital technologies with sound pedagogical practices. Rather than positioning AI as an authoritative source of knowledge, the framework emphasizes the importance of maintaining human judgment at the center of the learning process. The review also argues that AI integration should be aligned with the linguistic, cultural, and educational traditions of Arabic language learning, particularly in contexts where religious, ethical, and pedagogical values play a significant role.

Despite these contributions, the proposed framework remains conceptual and should be interpreted as a starting point for further inquiry rather than a definitive model. Future research should empirically examine the applicability of the framework across diverse educational settings, including secondary schools, *madrasahs*, *pesantren*, and higher education institutions. Further studies are also needed to investigate the relationships among the framework's components, develop and validate instruments for measuring critical AI literacy in Arabic language learning, and evaluate the long-term effectiveness of AI-supported pedagogical interventions. In addition, comparative and cross-cultural research would help determine how critical AI literacy can be adapted to different educational traditions while preserving the linguistic and cultural distinctiveness of Arabic language education.

REFERENCES

Abidin, H., & Sain, Z. H. (2025). The transformation of Arabic language learning in the digital era: A critical review of technological innovation, linguistic complexity, and the pedagogical imperatives of TPACK (2020–2025). *JETech: Journal of Education and Technology, 1*(3), 97–106. https://doi.org/10.65678/jetech.v1i3.231

Adawiyah, R. (2025). Implementing AI in Arabic language learning: Challenges and insights from Islamic higher education. *Al-Ishlah: Jurnal Pendidikan, 17*(3), 3729–3739. https://doi.org/10.35445/alishlah.v17i3.7390

Adel, M., Alhafni, B., & Habash, N. (2026). Arabic morphosyntactic tagging and dependency parsing with large language models. *arXiv*. https://doi.org/10.48550/arXiv.2603.16718

Ahmad, M. N. (2025). Narrative literature reviews in scientific research: Pros and cons. *Jordan Journal of Agricultural Sciences, 21*(1), 1–4. https://doi.org/10.35516/jjas.v21i1.4143

Akhtar, M., Rajwana, M. A., Malik, J. A., Baloch, M. A., Hussain, A., & Soomro, A. M. (2025). Adaptive learning systems for Classical Arabic: Personalised pathways to proficiency. *Southern Journal of Arts & Humanities, 3*(2), 57–67.

Alayba, A. M. (2025). Arabic natural language processing (NLP): A comprehensive review of challenges, techniques, and emerging trends. *Computers, 14*(11), 497. https://doi.org/10.3390/computers14110497

Albantani, A. M., Rozak, A., Sahrir, M. S., & Khoiriyyah, B. (2025). The influence of generative AI on student engagement and academic outcomes in Arabic language learning. *Asalibuna, 9*(2), 149–163. https://doi.org/10.30762/asalibuna.v9i02.6810

Alhafni, B., & Habash, N. (2025). Enhancing text editing for grammatical error correction: Arabic as a case study. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 17892–17914). Association for Computational Linguistics.

Al-Jamali, S., & Abdalla, S. Z. S. (2025). Behavioral determinants of AI-driven Arabic language learning: Insights from the extended UTAUT2 model. *Educational Process: International Journal, 16*, e2025282. https://doi.org/10.22521/edupij.2025.16.282

Al-Jarf, R. (2025). Specific linguistic questions that artificial intelligence (AI) cannot answer accurately: Implications for digital didactics. *Frontiers in Computer Science and Artificial Intelligence, 4*(4), 43–61. https://doi.org/10.32996/fcsai.2025.4.4.4

Alkaabi, M. H., & Almaamari, A. S. (2025). Generative AI implementation and assessment in Arabic language teaching. *International Journal of Online Pedagogy and Course Design, 15*(1), 1–18.

Al-Said, A. B. (2026). Mizan: An innovative tool for building Arabic morphological pattern dictionaries and enhancing lexical resource development. *ACM Transactions on Asian and Low-Resource Language Information Processing, 25*(2), 11. https://doi.org/10.1145/3786605

Baharun, S., Solehudin, M., Masnun, & Syaheed, S. M. (2025). The I'rab method of Habib Hasan Baharun: Impact on Arabic grammar instruction. *Al-Muhawarah: Jurnal Pendidikan Bahasa Arab, 1*(1), 23–35.

Baumeister, R. F., & Leary, M. R. (1997). Writing narrative literature reviews. *Review of General Psychology, 1*(3), 311–320. https://doi.org/10.1037/1089-2680.1.3.311

Borham, S. R., Ramli, S., & Abdul Ghani, M. T. (2024). AI concepts integration in developing E-Muhadathat kits for non-Arabic speakers. *Ijaz Arabi: Journal of Arabic Learning, 7*(3). https://doi.org/10.18860/ijazarabi.V7i3.26568

Boulesnam, I., & Boucetti, R. (2025). Arabic language characteristics that make its automatic processing challenging. *The International Arab Journal of Information Technology, 22*(4), 814–831. https://doi.org/10.34028/iajit/22/4/14

Bourouba, H. (2025). The impact of language on the development of artificial intelligence: An analytical study on the possibility of building an authentic Arabic AI model. *ZAOULI, 6*(12), 315–337.

Chirkunov, K., Alhafni, B., Qwaider, C., Habash, N., & Briscoe, T. (2025). ARWI: Arabic Write and Improve. In *Proceedings of the Fourth Workshop on Intelligent and Interactive Writing Assistants (In2Writing 2025)* (pp. 11–18). Association for Computational Linguistics.

Creely, E., & Carabott, K. (2025). Teaching and learning with AI: An integrated AI-oriented pedagogical model. *The Australian Educational Researcher*. Advance online publication. https://doi.org/10.1007/s13384-025-00913-6

Dahlan, J., Abdurrahman, M., & Rif'at, A. D. (2023). The utilization of artificial intelligence in Arabic language learning: Between linguistic competence and Islamic spirituality. *International Journal of Islamic Thought and Humanities, 2*(2), 380–394. https://doi.org/10.54298/ijith.v2i2.733

Doohee, A. H. (2024). The contribution of artificial intelligence to learning the Arabic language. *European Journal of Language and Culture Studies, 3*(3), 17–23. https://doi.org/10.24018/ejlang.2024.3.3.120

Ferrari, R. (2015). Writing narrative style literature reviews. *Medical Writing, 24*(4), 230–235. https://doi.org/10.1179/2047480615Z.000000000329

Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information and Libraries Journal, 26*(2), 91–108. https://doi.org/10.1111/j.1471-1842.2009.00848.x

Jayasinghe, S., Gamage, K. A. A., Yang, D., Cheng, C., Disanayake, C., & Apeji, U. D. (2026). Six institutional intervention areas to support ethical and effective student use of generative AI in higher education: A narrative review. *Education Sciences, 16*(1), 137. https://doi.org/10.3390/educsci16010137

Karataş, F., Abedi, F. Y., Gunyel, F. O., Karadeniz, D., & Kuzgun, Y. (2024). Incorporating AI in foreign language education: An investigation into ChatGPT's effect on foreign language learners. *Education and Information Technologies, 29*, 19343–19366. https://doi.org/10.1007/s10639-024-12574-6

Karima, S. R., Edidarmo, T., & Raswan. (2025). A comparison of the accuracy of generative artificial intelligence models in taṣrīf and the explanation of wazan meanings: A study on their application in Arabic morphology (ṣarf). *JALSAT Journal of Arabic Language Studies and Teaching, 5*(2), 234–250. https://doi.org/10.15642/jalsat.2025.5.2.234-250

Khouli, A., Esa, A. A., Hamamra, B., Nierat, S., & Khlaif, Z. N. (2026). From vocabulary to voice: How generative AI shapes Arabic learning for non-native speakers. *Research Square*. https://doi.org/10.21203/rs.3.rs-9249983/v1

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). Exploring generative artificial intelligence preparedness among university language instructors: A case study. *Computers and Education: Artificial Intelligence, 5*, 100156. https://doi.org/10.1016/j.caeai.2023.100156

Law, L. (2024). Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping literature review. *Computers and Education Open, 6*, 100174. https://doi.org/10.1016/j.caeo.2024.100174

Linur, R., Rusdi, Lubis, T., & Rohman, K. S. (2025). Student perception analysis of AI's impact on Arabic language learning: A personality perspective. *Jurnal Al Bayan: Jurnal Jurusan Pendidikan Bahasa Arab, 17*(2), 312–328. https://doi.org/10.24042/fya8h073

Magdy, S. M., Alwajih, F., Kwon, S. Y., Abdel-Salam, R., & Abdul-Mageed, M. (2024). Gazelle: An instruction dataset for Arabic writing assistance. *arXiv*. https://doi.org/10.48550/arXiv.2410.18163

Moustafa, A. H. A., Al-Hamad, M. F., Qureshi, M. A., & Qadir, J. (2026). Generative AI for learning and teaching Arabic: Opportunities, challenges, and ethical considerations. *IEEE Access, 14*. https://doi.org/10.1109/ACCESS.2026.3651864

Mubarak, H., Hawasly, M., & Mohamed, A. (2026). Nahw: A comprehensive benchmark of Arabic grammar understanding, error detection, correction, and explanation. In *Proceedings of the 19th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 6310–6328). Association for Computational Linguistics.

Nandatasia, R. A., & Nugrahawan, A. R. (2026). Fostering mahārah kitābah through picture-based controlled writing with AI-generated visual stimuli. *Alibbaa': Jurnal Pendidikan Bahasa Arab, 7*(1), 158–179. https://doi.org/10.19105/ajpba.v7i1.23463

Nugraha, A., & Syafe'i, I. (2025). Curriculum foundations for Arabic language education in the AI era: Holistic, juridical, and technological perspectives. *Journal of Arabic Language Learning and Teaching (JALLT), 3*(2), 151–160. Retrieved from https://ejournal.iain-palangkaraya.ac.id/index.php/jallt

Othman, A. M. W., & Asbulah, L. H. B. (2025). Problems of artificial intelligence applications in digitizing the Arabic language in the areas of grammar and morphology and methods to resolve it. *IJAZ ARABI: Journal of Arabic Learning, 8*(2), 721–740. https://doi.org/10.18860/ijazarabi.V8i2.31596

Pang, S., Nol, E., & Heng, K. (2025). Artificial intelligence (AI) literacy: A necessity in modern language education. *Journal of English Learner Education, 17*(1), 8. Retrieved from https://stars.library.ucf.edu/jele/vol17/iss1/8

Perkins, M. (2023). Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching & Learning Practice, 20*(2), 7. https://doi.org/10.53761/1.20.02.07

Rahmaddani, A., & Naifah. (2025). Students' perspectives on the use of Chat-GPT in the fields of phonology, morphology, and syntactics in Arabic language learning. *Journal of Islamic Education Thought and Development, 1*(1), 48–60.

Rahmouni, K. (2024). Exploring the use of ChatGPT in teaching Arabic case endings: Effectiveness, challenges and recommendations. *Journal of Educational Technology and Innovation, 6*(4), 1–20. https://doi.org/10.61414/jeti.v6i4.198

Sa'idah, M. A., Dolan, E., Diantoro, K., Santoso, N. A., Mahmudah, U., & Junaedi, S. R. P. (2024). Enhancing Arabic language teaching through artificial intelligence: Assessing effectiveness and educational implications. In *2024 3rd International Conference on Creative Communication and Innovative Technology (ICCIT)* (pp. 1–8). IEEE. https://doi.org/10.1109/ICCIT62134.2024.10701089

Samiya, R. (2025). Artificial intelligence in Arabic language education: Current applications, challenges, and future perspectives. *Studies in Education Sciences, 6*(4), 1–21. https://doi.org/10.54019/sesv6n4-011

Sibaee, S., Nacar, O., Ammar, A., Al-Habashi, Y., Al-Batati, A., & Boulila, W. (2025). From guidelines to practice: A new paradigm for Arabic language model evaluation. *arXiv*. https://doi.org/10.48550/arXiv.2506.01920

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

Tamam, M. I., Ilahi, M. M. K., Cholilah, Z., Taufiqurrochman, R., & Machmudah, U. (2024). Utilizing ChatGPT for analyzing Arabic texts in the study of nahwu (Arabic grammar). *KITABA: Journal of Interdisciplinary Arabic Learning, 2*(3), 193–208. https://doi.org/10.18860/kitaba.v2i3.28463

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

Wu, J. G., Lee, S., Zhang, D., & Szabo, C. (2025). Cultivating AI literacy in language education: Theoretical discussions and practical applications. *Journal of China Computer-Assisted Language Learning, 5*(2), 169–174. https://doi.org/10.1515/jccall-2025-0023

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