

Generative AI in Arabic Grammar Learning: A Critical Review of Pedagogical Benefits and Linguistic Limitations

Muhammad Fariq Heemal Attruk*, Nik Hanan Mustapha

International Islamic University Malaysia, Malaysia

* fariqhemal@gmail.com (Primary Contact)

ABSTRACT

Keywords

generative artificial intelligence, Arabic grammar learning, nahwu, shorof, AI-assisted learning, pedagogical implications

Article History

Received: 2026-05-07

Accepted: 2026-06-11

The rapid development of generative artificial intelligence (AI) has significantly influenced Arabic language education, particularly in the learning of *nahwu* (syntax) and *shorof* (morphology). This study aims to critically examine the pedagogical benefits and linguistic limitations of generative AI in Arabic grammar learning through a narrative literature review approach. The study analyzes recent scholarly works on AI-assisted Arabic language instruction and grammatical analysis. The findings reveal that generative AI supports independent learning, provides immediate feedback, and increases learner engagement. However, persistent limitations remain in handling complex syntactic structures, contextual ambiguity, and accurate grammatical interpretation. The review also highlights concerns regarding learners' overreliance on AI and the continuing need for teacher supervision. The study concludes that generative AI should function as a supportive instructional tool rather than a fully autonomous learning authority in Arabic grammar education.

Copyright © 2026, Attruk et al.

Published by MAN 4 Kota Pekanbaru

DOI: [10.56113/takuana.v5i1.493](https://doi.org/10.56113/takuana.v5i1.493)

1. INTRODUCTION

The rapid development of generative artificial intelligence (AI) has significantly influenced various domains of education, including language learning. In recent years, AI-based systems have increasingly been utilized as interactive tools that support learners in understanding linguistic concepts, providing immediate feedback, and facilitating more flexible learning experiences. In the context of Arabic language learning, several studies indicate that generative AI has begun to play an important role in assisting students in engaging with complex grammatical structures, particularly in the areas of *nahwu* (syntax) and *shorof* (morphology) (Rahmaddani & Naifah, 2025; Rahmouni, 2024; Sa'idah et al., 2024; Albantani et al., 2025). In this review, the term “generative AI” refers primarily to large language model (LLM)-based conversational agents—such as ChatGPT—that are employed as interactive instructional tools. However, the discussion also encompasses other AI-powered applications with generative capabilities, including automatic harakat

detection systems and Arabic grammatical error correction tools, which are explicitly noted where relevant.

Arabic grammar is widely recognized as one of the most challenging components of language learning due to its structural complexity, including rich morphological patterns and context-dependent syntactic rules. These characteristics often make it difficult for learners to achieve a deep and consistent understanding of grammatical concepts. In this context, the emergence of generative AI offers new opportunities to support learning by enabling students to access explanations, practice exercises, and feedback in a more immediate and interactive manner (Adawiyah, 2025; Linur et al., 2025; Zubaidi et al., 2025). As a result, AI has been increasingly adopted as a supplementary tool in both formal educational settings and independent learning environments.

A growing body of research highlights the pedagogical potential of generative AI in Arabic grammar learning. Studies have shown that AI-based tools can enhance students' conceptual understanding, increase engagement, and support self-directed learning (Albantani et al., 2025; Sa'idah et al., 2024; Rahmaddani & Naifah, 2025). In addition, AI systems have been utilized as practice tools for grammatical analysis and morphological transformation, allowing learners to interact with language structures more intensively and repeatedly (Zubaidi et al., 2025; Rizki et al., 2025). These findings suggest that generative AI can play a meaningful role in addressing some of the challenges associated with learning Arabic grammar.

However, despite these advantages, the literature also reveals significant limitations related to linguistic accuracy. Several studies report that generative AI systems may produce incorrect or inconsistent analyses of Arabic grammatical structures, particularly in tasks involving syntactic parsing and morphological processing (Alkaabi & Almaamari, 2025; Tamam et al., 2024; Karima et al., 2025; Othman & Asbulah, 2025). The probabilistic nature of these models, combined with the inherent complexity of Arabic grammar, makes it difficult to ensure reliable and consistent outputs across different contexts (Adel et al., 2026; Abdelrehim et al., 2025). These limitations raise important concerns regarding the reliability of AI as a learning resource.

Beyond technical limitations, the use of generative AI also has pedagogical implications. Research indicates that students may develop misconceptions when relying on inaccurate AI-generated explanations and may become overly dependent on AI tools without engaging in critical evaluation (Al-Jamali & Abdalla, 2025; Linur et al., 2025). At the same time, the increasing integration of AI in education requires a redefinition of the role of teachers, who are expected to guide students in using AI tools critically and effectively (Nugraha & Syafe'i, 2025).

Taken together, these findings point to a central issue in the literature: the coexistence of pedagogical benefits and linguistic limitations in the use of generative AI for Arabic grammar learning. Although previous studies have discussed the educational advantages and challenges of generative AI, existing research remains fragmented and often focuses on general language learning contexts or students' perceptions rather than the specific domain of Arabic grammar learning. Moreover, limited studies critically examine the interaction between pedagogical benefits and linguistic limitations, particularly regarding the reliability of AI-generated grammatical analysis in handling complex syntactic and morphological structures. As a result, there is still a need for a more integrated, critical, and domain-specific perspective that systematically synthesizes these dimensions within the

context of *nahwu* and *shorof* learning. The novelty of this study lies in integrating pedagogical and linguistic perspectives in examining generative AI within Arabic grammar education, thereby offering a more comprehensive and context-sensitive understanding of its potential and limitations.

Therefore, this study aims to provide a critical narrative literature review of generative AI usage in Arabic grammar learning by examining both its pedagogical contributions and its limitations in linguistic accuracy. Specifically, this review seeks to: (1) analyze how generative AI is used in Arabic grammar learning; (2) identify its pedagogical benefits; (3) examine its limitations in linguistic accuracy; and (4) explore its pedagogical implications for learners and educators. By synthesizing these aspects, this study aims to offer a more comprehensive and nuanced understanding of the role of generative AI in Arabic grammar learning and to identify key directions for future research.

2. METHOD

This study employs a narrative literature review (NLR) approach to critically examine the use of generative AI in Arabic grammar learning. Unlike reviews that merely map existing research, an NLR enables critical analysis, evaluation, and synthesis of findings, generating conclusions that extend beyond individual empirical reports (Grant & Booth, 2009; Baumeister & Leary, 1997). The review focuses on both pedagogical benefits and linguistic limitations of generative AI, particularly in relation to *nahwu* (syntax) and *shorof* (morphology) (Baalbaki, 2019). This dual focus reflects the NLR principle that critical appraisal—including identification of weaknesses—is integral to rigorous literature review (Baumeister & Leary, 1997). The analytical framework encompasses search, appraisal, synthesis, and analysis, structured around four objectives: identifying the roles of generative AI, examining its pedagogical benefits, analyzing its linguistic limitations, and synthesizing the tension between these dimensions along with their pedagogical implications (Grant & Booth, 2009).

The literature search was conducted using two major academic databases, namely Scopus and Google Scholar, in order to ensure broad and credible coverage of relevant studies, consistent with the principle of exhaustive and comprehensive searching in rigorous literature reviews (Grant & Booth, 2009). Scopus was used to identify peer-reviewed international publications, while Google Scholar was employed to capture a wider range of relevant studies that may not be indexed in Scopus but are still academically significant (Falagas et al., 2008). The search strategy was developed by combining key thematic components, including generative AI and artificial intelligence, education and pedagogy, Arabic language and linguistics, as well as grammar, syntax (*nahwu*), and morphology (*shorof*). The primary search string applied in Scopus consisted of a combination of these keywords using Boolean operators (Ferrari, 2015). In Google Scholar, a more flexible keyword variation strategy was adopted to capture diverse terminologies used in the literature, including phrases such as "generative AI in Arabic grammar learning," "AI for Arabic syntax learning," "AI for Arabic morphology learning," "ChatGPT in Arabic language education," "artificial intelligence in Arabic grammar teaching," and "AI-assisted Arabic language learning." This approach was intended to capture variations in terminology and identify relevant studies that may not be indexed using standardized keywords. The literature search was limited to publications from 2020 to 2026 to ensure that the review

reflects recent developments in generative AI and remains relevant to current educational practices.

To ensure the relevance and quality of the selected studies, this review applied specific inclusion and exclusion criteria (Ferrari, 2015). Studies were included if they discussed the use of generative AI or artificial intelligence in educational contexts, focused on language learning with particular emphasis on Arabic, addressed aspects of Arabic grammar such as *nahwu* and *shorof*, and presented empirical findings, case studies, or academic analyses. In addition, included studies were required to discuss either the pedagogical benefits or the limitations of AI and to be published between 2020 and 2026. Only studies written in English were included to maintain consistency and ensure alignment with international academic standards, as restricting literature reviews to English-language publications is a widely accepted methodological practice that does not substantially alter review conclusions (Morrison et al., 2012). Conversely, studies were excluded if they did not specifically address generative AI, focused solely on technical development without educational context, were not related to Arabic language learning, did not address grammatical aspects, consisted of opinion-based or non-academic writings, were duplicates across databases, or were published before 2020.

The study selection process was conducted in three stages to ensure systematic filtering of relevant literature (Ferrari, 2015). Initially, articles were screened based on their titles to eliminate studies that were clearly irrelevant to generative AI, Arabic language learning, or grammar. This was followed by abstract screening, where the remaining articles were evaluated to determine their relevance to the research focus, particularly whether they addressed AI in educational contexts, Arabic language learning, and grammatical aspects. Finally, full-text review was conducted on the selected articles to ensure their alignment with the inclusion criteria, their relevance to the research questions, and the presence of meaningful analytical content. Through this multi-stage process, a corpus of relevant studies was identified and used as the basis for further analysis. The selected studies were also critically evaluated in terms of their methodological relevance, analytical depth, and contribution to the research objectives. The initial searches across Scopus and Google Scholar yielded several hundred records in total. Following the removal of duplicates, title and abstract screening, and full-text assessment against the inclusion and exclusion criteria, a final corpus of 19 studies was included in the review. It should be noted that, consistent with the narrative rather than aggregative purpose of this review, a fully itemized record of the number of studies excluded at each screening stage was not maintained. Unlike a systematic review, which aims to exhaustively catalogue all available evidence, this NLR prioritized the purposive selection of studies that meaningfully address the intersection of generative AI, Arabic grammar learning, and the pedagogical–linguistic tension under investigation (Grant & Booth, 2009; Baumeister & Leary, 1997).

Data extraction was conducted systematically to identify key information from each selected study (Ferrari, 2015; Grant & Booth, 2009). The extracted data included the authors and year of publication, research focus and objectives, context of AI use in Arabic grammar learning, identified pedagogical benefits, reported limitations — particularly those related to linguistic accuracy — and key findings relevant to the research questions. This process enabled the organization of data into comparable categories, facilitating thematic analysis across studies.

The analysis followed a thematic synthesis approach consistent with the narrative literature review design (Grant & Booth, 2009). The selected studies were examined to identify patterns, similarities, and differences in their findings (Ferrari, 2015). The synthesis was structured into four main analytical themes: the role of generative AI as a learning tool, its pedagogical benefits, its limitations in terms of linguistic accuracy, and its pedagogical implications for learners. These themes were then integrated into a higher-level conceptual synthesis, which highlights the tension between pedagogical utility and linguistic reliability as the central finding of this review, consistent with the principle that narrative literature reviews should present conclusions at a higher level of abstraction beyond individual empirical reports (Baumeister & Leary, 1997).

3. RESULTS AND DISCUSSION

Table 1 provides an overview of the 19 reviewed studies, covering their contexts, methodologies, and key findings.

Table 1. Summary of Reviewed Studies

No	Author(s) & Year	Context	Methodology	Key Findings
1	Abdelrehim et al. (2025)	Hybrid LLM and rule-based synthetic data for Arabic grammatical error correction	Experimental; model development	Hybrid models improve synthetic data generation but require further fine-tuning for pedagogical use
2	Adawiyah (2025)	AI implementation in Arabic learning in Islamic higher education	Qualitative; case study	AI perceived as helpful but faces challenges in adoption and accuracy
3	Adel et al. (2026)	Arabic morphosyntactic tagging and dependency parsing using LLMs	Experimental; model evaluation	LLMs show potential but need fine-tuning and rule-based integration for optimal performance
4	Albantani et al. (2025)	Influence of generative AI on student engagement and academic outcomes	Quantitative; survey	AI enhances engagement and is perceived to improve learning outcomes
5	Al-Jamali & Abdalla (2025)	Behavioural determinants of AI-driven Arabic learning (UTAUT2 model)	Quantitative; survey	Social factors and habits influence AI adoption; risks of overreliance identified
6	Al-Jarf (2025)	Linguistic questions AI cannot answer accurately	Qualitative analysis	AI often fails on complex and ambiguous linguistic questions
7	Alkaabi & Almaamari (2025)	Generative AI implementation and assessment in Arabic language teaching	Mixed-methods	AI is beneficial but not always accurate; critical evaluation strategies needed

No	Author(s) & Year	Context	Methodology	Key Findings
8	Karima et al. (2025)	Accuracy comparison of generative AI models in <i>tashrif</i> and <i>wazan</i> interpretation	Comparative; model evaluation	AI accuracy is limited in <i>tashrif</i> and <i>wazan</i> interpretation tasks
9	Kwon et al. (2023)	ChatGPT for Arabic grammatical error correction	Experimental; performance evaluation	ChatGPT can correct errors but performance is inconsistent
10	Linur et al. (2025)	Student perception of AI's impact on Arabic learning from a personality perspective	Quantitative; survey	Students perceive AI positively but trust does not always align with accuracy
11	Nugraha & Syafe'i (2025)	Curriculum foundations for Arabic education in the AI era	Qualitative; policy analysis	Curriculum adaptation needed for critical AI integration
12	Othman & Asbulah (2025)	Problems of AI applications in digitizing Arabic grammar and morphology	Qualitative; problem analysis	AI faces fundamental challenges in handling Arabic grammar and morphology
13	Rahmaddani & Naifah (2025)	Students' perspectives on ChatGPT in phonology, morphology, and syntax	Qualitative; survey	Students find ChatGPT helpful for understanding linguistic components
14	Rahmouni (2024)	Using ChatGPT in teaching Arabic case endings (<i>i'rab</i>)	Small-scale experimental	ChatGPT facilitates understanding of <i>i'rab</i> but has accuracy limitations
15	Rishanda et al. (2025)	AI in independent <i>nahwu</i> learning	Qualitative; case study	Students actively use AI for autonomous <i>nahwu</i> study
16	Rizki et al. (2025)	Automatic harakat detection for Arabic sentence learning (Qalam AI)	System development; trial	Harakat detection supports orthographic and grammatical accuracy
17	Sa'idah et al. (2024)	Enhancing Arabic language teaching through AI	Mixed-methods	AI improves teaching effectiveness and student engagement
18	Tamam et al. (2024)	Utilizing ChatGPT for analysing Arabic texts in <i>nahwu</i> study	Qualitative; case analysis	ChatGPT shows high accuracy (99%) in <i>i'rab</i> detection but has errors in certain particle categories
19	Zubaidi et al. (2025)	Enhancing Arabic writing skills using ChatGPT-based AI learning models	Experimental; human-AI framework	Human-AI collaboration improves writing and grammatical skills

3.1. Generative AI as a Tool in Arabic Grammar Learning

Generative AI has been increasingly integrated into Arabic grammar learning as a supportive instructional tool across various educational contexts. Evidence from multiple studies indicates that AI-based systems are not only used as sources of information but also function as interactive learning assistants that facilitate students' engagement with complex grammatical concepts in Arabic, particularly in the domains of *nahwu* (syntax) and *shorof* (morphology) (Rahmaddani & Naifah, 2025; Rahmouni, 2024; Sa'idah et al., 2024; Albantani et al., 2025; Adawiyah, 2025).

In practice, generative AI performs multiple pedagogical roles. First, it operates as a virtual tutor, providing explanations of grammatical rules, sentence structures, and linguistic patterns. Learners are able to request step-by-step analyses of sentence constructions or grammatical cases (*i'rab*), enabling immediate clarification without direct teacher intervention. Rahmaddani and Naifah (2025) demonstrate that students actively use AI tools to explore linguistic components such as phonology, morphology, and syntax, while Rahmouni (2024) highlights the use of AI systems such as ChatGPT in assisting learners' understanding of Arabic case endings. Additional studies also indicate that AI can provide structured linguistic explanations that support learners in navigating complex grammatical patterns (Sa'idah et al., 2024).

Second, generative AI functions as a learning assistant that supports independent and self-directed learning. Through interactive dialogue, learners can pose questions, test their understanding, and receive instant feedback, allowing AI to act as a form of scaffolding in the learning process. This role is supported by findings from Albantani et al. (2025) and Sa'idah et al. (2024), which indicate that AI-enhanced environments contribute to increased student engagement and more flexible learning experiences. Similarly, Adawiyah (2025) and Linur et al. (2025) report that learners perceive AI as a helpful and accessible tool that supports their ongoing study of Arabic and facilitates autonomous learning.

Third, generative AI is widely utilized as a practice tool, particularly for exercises involving *i'rab* analysis and *tashrif* (morphological transformation). Learners can input sentences or word forms and receive grammatical breakdowns or conjugation patterns, enabling repeated practice and reinforcement. Zubaidi et al. (2025) emphasize that AI-supported writing and grammar tools enhance students' ability to construct and analyze Arabic sentences through iterative human-AI interaction. In addition, Rizki et al. (2025) demonstrate that AI-based systems for automatic *harakat* detection can support learners in addressing one of the most challenging aspects of Arabic grammar.

In terms of learning contexts, the use of generative AI extends across both formal and informal settings. Within formal education, AI is typically integrated as a supplementary instructional tool that complements teacher-led instruction. In informal contexts, however, AI plays a more central role by enabling autonomous exploration of grammatical concepts. Studies such as Rishanda et al. (2025) and Linur et al. (2025) indicate that students increasingly rely on AI tools for independent learning of *nahwu*, reflecting a broader shift toward learner-centered approaches facilitated by digital technologies.

While these studies collectively demonstrate the diverse pedagogical roles of generative AI, several methodological limitations should be noted. Much of the evidence for AI as a virtual tutor and learning assistant is drawn from small-scale qualitative studies and self-reported data (Rahmaddani & Naifah, 2025; Adawiyah, 2025; Linur et al., 2025), which,

although rich in contextual detail, may not be generalizable across different educational settings. Studies exploring AI as a practice tool, such as Zubaidi et al. (2025) and Rizki et al. (2025), are based on specific AI systems in controlled conditions, limiting their applicability to other tools or informal learning contexts. Furthermore, the rapid evolution of generative AI models means that findings from studies conducted even two or three years ago may not fully reflect the capabilities or limitations of current systems. These factors suggest that the portrayal of generative AI in the literature, while promising, rests on an evidence base that is still developing and requires more robust, large-scale, and longitudinal investigation.

Overall, the literature indicates that generative AI has evolved into a multifunctional tool in Arabic grammar learning, encompassing roles as tutor, assistant, and practice medium. Its ability to provide instant feedback, interactive explanations, and flexible access positions it as a valuable resource for learners dealing with the inherent complexity of Arabic grammatical systems. However, while these functions demonstrate its pedagogical utility, they also raise important questions regarding the accuracy and reliability of AI-generated outputs, which will be critically examined in the following section.

3.2. Pedagogical Benefits of Generative AI in Arabic Grammar Learning

Generative AI has demonstrated considerable potential in enhancing the pedagogical processes involved in Arabic grammar learning. Evidence across multiple studies indicates that AI-based tools contribute to improved comprehension, increased learner engagement, and greater flexibility in accessing grammatical knowledge (Rahmaddani & Naifah, 2025; Rahmouni, 2024; Albantani et al., 2025; Sa'idah et al., 2024; Adawiyah, 2025; Linur et al., 2025). These benefits are particularly relevant in the context of Arabic grammar, which is often perceived as complex due to its intricate syntactic (*nahwu*) and morphological (*shorof*) structures.

One of the most prominent benefits identified in the literature is the enhancement of students' conceptual understanding. Generative AI enables learners to receive immediate explanations of grammatical rules and examples, which supports deeper comprehension. Rahmaddani and Naifah (2025) report that students perceive AI tools as helpful in understanding linguistic components such as syntax and morphology, while Rahmouni (2024) highlights how AI-assisted explanations of Arabic case endings can facilitate learners' grasp of *i'rab*. Additional evidence from Albantani et al. (2025) and Sa'idah et al. (2024) suggests that AI-supported learning environments contribute to improved learning outcomes and conceptual clarity.

In addition to improving understanding, generative AI also promotes active and independent learning. The interactive nature of AI systems allows learners to engage in self-directed exploration by asking questions, testing hypotheses, and receiving instant feedback. Studies by Albantani et al. (2025) and Sa'idah et al. (2024) demonstrate increased student engagement, while Linur et al. (2025), and Rishanda et al. (2025) further indicate that learners actively use AI tools to support independent study of Arabic grammar, particularly in *nahwu*.

Another key advantage lies in the ability of generative AI to function as an effective practice medium. Through repeated interaction, learners can practice grammatical analysis and receive immediate corrections or alternative formulations. Zubaidi et al. (2025) show that AI-based learning models enhance students' writing and grammatical skills through

continuous human–AI interaction. Similarly, Rizki et al. (2025) demonstrate that AI systems designed for automatic *harakat* detection can support learners in addressing orthographic and grammatical accuracy, which are essential components of Arabic grammar mastery.

The literature also highlights the role of generative AI in increasing accessibility to learning resources. Unlike traditional classroom settings that are constrained by time and instructor availability, AI tools provide on-demand assistance that can be accessed anytime and anywhere. Studies such as Adawiyah (2025), Linur et al. (2025), and Rishanda et al. (2025) indicate that students increasingly rely on AI for flexible and autonomous learning, suggesting that generative AI contributes to more inclusive learning environments.

Overall, the findings indicate that generative AI offers multiple pedagogical advantages in Arabic grammar learning, including enhanced understanding, increased engagement, support for independent learning, and expanded access to educational resources. These benefits position AI as a valuable supplementary tool in both formal and informal learning contexts. However, it is important to note that a considerable portion of the evidence supporting these pedagogical benefits is derived from learners' self-reported perceptions rather than from objectively measured learning outcomes. Studies by Albantani et al. (2025), Sa'idah et al. (2024), Adawiyah (2025), and Linur et al. (2025), for instance, primarily rely on surveys, questionnaires, or interviews that capture students' subjective views on the usefulness and accessibility of generative AI tools. While such perceptual data offer valuable insights into learners' experiences and attitudes, they do not necessarily reflect actual gains in grammatical competence. Perceived effectiveness and empirically measured effectiveness represent different levels of evidence, and claims of improved conceptual understanding or enhanced learning should be interpreted with appropriate caution when they are based solely on self-report measures. Future research would benefit from longitudinal and experimental designs that assess the impact of generative AI on Arabic grammar learning through objective, performance-based assessments. However, despite these contributions, the literature also points to important limitations—particularly regarding the accuracy and reliability of AI-generated grammatical analysis—which will be critically examined in the following section.

3.3. Limitations of Generative AI in Arabic Grammar Learning: Issues of Linguistic Accuracy

Despite the pedagogical advantages outlined in the previous section, a substantial body of studies highlights important limitations of generative AI in the context of Arabic grammar learning, particularly concerning linguistic accuracy. Evidence across multiple studies indicates that while AI systems can generate grammatically plausible outputs, they often struggle to provide consistently accurate analyses of Arabic syntactic (*nahwu*) and morphological (*shorof*) structures (Alkaabi & Almaamari, 2025; Othman & Asbulah, 2025). It should be noted that many of these limitations have been observed primarily in general-purpose large language models (LLMs), such as ChatGPT-3.5 and ChatGPT-4, rather than in specialized Arabic natural language processing (NLP) systems or models specifically fine-tuned for Arabic grammatical tasks. The extent to which these limitations apply across different AI architectures therefore varies, and caution is warranted when generalizing findings from one type of system to generative AI as a whole.

One of the most frequently reported issues relates to errors in syntactic analysis, especially in tasks involving i'rab. Several studies demonstrate that generative AI systems may incorrectly assign grammatical roles, misidentify sentence structures, or provide incomplete explanations of syntactic relationships. Tamam et al. (2024) show that while AI-based tools can achieve high accuracy in i'rab detection, they still produce errors in certain particle categories and complex sentence constructions, while Alkaabi and Almaamari (2025) report that AI-generated grammatical analyses are not always reliable in instructional contexts. Related findings from Othman and Asbulah (2025) further indicate that AI systems face persistent challenges in handling Arabic syntactic structures accurately.

In addition to syntactic challenges, morphological inaccuracies also represent a major limitation. Arabic morphology, which involves intricate root-and-pattern transformations (*tashrif*), poses significant difficulties for generative AI systems. Karima et al. (2025) find that AI-generated outputs in tasks related to *tashrif* and *wazan* interpretation often contain inaccuracies, suggesting that these systems do not fully capture the underlying morphological rules of the language. This limitation is particularly critical because accurate morphological understanding is essential for mastering Arabic grammar.

Another important issue concerns the inconsistency of AI-generated outputs. Generative AI models may produce different responses for similar inputs, influenced by variations in prompts or internal probabilistic processes. Kwon et al. (2023) demonstrate that while AI can be used for Arabic grammatical error correction, its performance is not always stable, and the quality of outputs can vary. Additional observations from Al-Jarf (2025) indicate that AI systems may produce misleading or partially correct explanations, further reinforcing concerns about reliability.

Furthermore, several studies point to the complexity of Arabic grammar itself as a key factor underlying these limitations. Arabic is characterized by rich morphology, flexible syntax, and context-dependent grammatical rules, making it particularly challenging for AI systems to process accurately. Studies such as Adel et al. (2026) and Abdelrehim et al. (2025) suggest that even advanced language models require additional mechanisms—such as fine-tuning or hybrid rule-based integration—to improve performance in Arabic grammatical tasks. This indicates that current generative AI models are not yet fully capable of handling the depth and nuance of Arabic linguistic structures.

Importantly, some studies explicitly highlight that there are types of linguistic questions that generative AI cannot answer accurately. Al-Jarf (2025) identifies cases in which AI systems fail to provide correct grammatical explanations, particularly when dealing with complex or ambiguous constructions. Similarly, Othman and Asbulah (2025) emphasize that AI continues to face fundamental problems in handling Arabic grammar and morphology, reinforcing the view that its capabilities remain limited in this domain.

While these studies provide important evidence of AI's linguistic limitations, several methodological considerations affect the strength of their conclusions. First, many of the error analyses are based on a limited set of sentence structures or grammatical phenomena (Tamam et al., 2024; Karima et al., 2025), which may not fully represent the range of challenges posed by Arabic grammar. Second, the criteria used to assess accuracy vary considerably across studies, with some relying on expert judgment (Al-Jarf, 2025) and others on automated metrics (Kwon et al., 2023), making direct comparisons difficult. Third, most of the reviewed studies evaluate general-purpose LLMs rather than models

specifically fine-tuned for Arabic, meaning the reported inaccuracies may partly reflect model choice rather than an inherent limitation of generative AI. These inconsistencies highlight the need for standardized evaluation frameworks and caution against overgeneralizing findings across different AI systems and linguistic tasks.

Taken together, these findings suggest that while generative AI can support Arabic grammar learning, its limitations in linguistic accuracy remain significant. Errors in syntactic and morphological analysis, inconsistencies in output, and difficulties in handling linguistic complexity all contribute to concerns about the reliability of AI in educational contexts. It must be acknowledged, however, that these findings largely reflect the performance of general-purpose LLMs, and that specialized or fine-tuned Arabic NLP models may perform differently; thus, broad generalizations about all generative AI systems should be made with caution. These limitations not only challenge the effectiveness of AI as a learning tool but also raise important pedagogical questions regarding how such technologies should be used, monitored, and integrated into language learning environments.

3.4. Pedagogical Implications and Learner Risks in the Use of Generative AI

The limitations of generative AI in linguistic accuracy, as discussed in the previous section, carry important implications for pedagogical practices in Arabic grammar learning. While AI tools offer convenience and accessibility, a growing body of studies indicates that their use also introduces a range of risks related to learners' conceptual understanding, learning behavior, and cognitive engagement (Al-Jamali & Abdalla, 2025; Adawiyah, 2025; Linur et al., 2025).

One of the primary concerns identified in the literature is the potential for misconceptions in grammatical understanding. Because generative AI systems may produce inaccurate or incomplete analyses of *nahwu* and *shorof*, learners who rely heavily on these tools may internalize incorrect grammatical rules. This issue is particularly critical for novice learners who lack sufficient prior knowledge to critically evaluate AI-generated responses. Al-Jamali and Abdalla (2025) show that AI-assisted learning can lead to confusion when outputs are inconsistent or misaligned with established grammatical principles. This concern is further supported by studies highlighting inaccuracies in AI-generated grammatical analysis, which may indirectly contribute to misunderstandings among learners.

In addition to misconceptions, the literature highlights the risk of overreliance on AI tools. The immediacy and accessibility of AI-generated responses may encourage learners to depend on AI rather than engage in deeper cognitive processing. Rishanda et al. (2025) report that students increasingly use AI tools for independent learning of *nahwu*, while Al-Jamali and Abdalla (2025) note that strong engagement with AI-assisted learning may create pedagogical challenges when learners rely heavily on AI-generated support. Similarly, Alkaabi and Almaamari (2025) emphasize that inaccuracies in AI-generated grammatical analysis highlight the importance of critical evaluation when using AI tools for Arabic grammar learning.

Another important implication concerns the changing role of teachers in AI-supported learning environments. As AI becomes more integrated into instructional practices, teachers are no longer the sole providers of knowledge but instead act as

facilitators who guide students in using AI critically and effectively. Nugraha and Syafe'i (2025) emphasize the need for curriculum adaptation in response to the growing presence of AI, while Alkaabi and Almaamari (2025) highlight the importance of instructional strategies that help learners critically evaluate AI-generated outputs and recognize potential inaccuracies. This shift underscores the need for a more guided and reflective use of AI in educational contexts.

Furthermore, the literature suggests that the integration of generative AI necessitates the development of critical AI literacy among learners. Students must be equipped not only with linguistic knowledge but also with the ability to assess the reliability of AI-generated content. Without such skills, learners may struggle to distinguish between correct and incorrect grammatical explanations. Linur et al. (2025) highlight that although students generally perceive AI positively, their trust in AI outputs does not always align with actual accuracy, indicating a gap between perception and reliability.

It should be noted, however, that much of the evidence on pedagogical risks—such as misconceptions and overreliance—is inferred from general observations or student self-reports rather than from controlled studies that isolate the effects of AI use on grammatical competence over time. Studies by Al-Jamali and Abdalla (2025) and Linur et al. (2025) identify potential risks based on survey data, but do not empirically demonstrate a causal link between AI use and the development of misconceptions or dependency. Similarly, discussions on the changing role of teachers (Nugraha & Syafe'i, 2025) are largely conceptual rather than grounded in classroom-based experimental research. While these insights are valuable, the pedagogical implications drawn from them should be regarded as preliminary and in need of more rigorous empirical validation.

Overall, the integration of generative AI into Arabic grammar learning presents a complex pedagogical landscape. While AI offers meaningful support for learning, it also introduces risks related to misunderstanding, dependency, and shifting educational roles. These findings underscore the importance of adopting a critical and balanced approach to AI use in education. Rather than replacing traditional instruction, generative AI should be positioned as a supplementary tool that requires careful supervision, guided use, and ongoing critical evaluation. This duality ultimately reinforces the need to examine the broader tension between the benefits and limitations of generative AI, which will be synthesized in the following section.

3.5. The Tension Between Pedagogical Utility and Linguistic Reliability

The findings of this review reveal a fundamental tension in the use of generative AI for Arabic grammar learning: while these technologies offer significant pedagogical benefits, they simultaneously exhibit limitations in linguistic accuracy that challenge their reliability as learning tools. This duality emerges consistently across the reviewed studies and provides a critical lens for understanding the role of AI in Arabic language education.

On the one hand, generative AI has been widely recognized as a facilitative tool that enhances learning experiences. Studies indicate that AI systems support learners in understanding complex grammatical concepts, provide immediate feedback, and enable flexible, self-directed learning (Rahmaddani & Naifah, 2025; Rahmouni, 2024; Albantani et al., 2025; Sa'idah et al., 2024; Adawiyah, 2025; Linur et al., 2025; Rishanda et al., 2025). These features are particularly valuable in Arabic grammar learning, where learners often

struggle with the abstract and rule-intensive nature of *nahwu* and *shorof*. The accessibility and interactivity of AI tools allow learners to engage with grammatical structures in ways that extend beyond traditional instructional settings.

However, this pedagogical utility is counterbalanced by persistent concerns regarding linguistic reliability. Multiple studies demonstrate that generative AI systems frequently produce inaccurate or inconsistent analyses of Arabic grammar, particularly in tasks involving syntactic parsing and morphological transformation (Alkaabi & Almaamari, 2025; Karima et al., 2025; Othman & Asbulah, 2025; Al-Jarf, 2025). These inaccuracies are not merely technical limitations but have direct implications for learning, as they may lead to the internalization of incorrect grammatical knowledge. Furthermore, the variability of AI outputs, as highlighted by Kwon et al. (2023), complicates its use as a stable reference for learners.

This tension becomes particularly critical when viewed from the learner's perspective. Students often perceive AI as an authoritative and trustworthy source due to its fluency and responsiveness, yet they may lack the expertise needed to critically evaluate its outputs (Linur et al., 2025). As a result, the very features that make AI attractive—such as immediacy, accessibility, and ease of use—can also amplify its risks, particularly when accuracy is compromised. Evidence from studies on learner behavior (Rishanda et al., 2025) further suggests that reliance on AI tools may reduce critical engagement if not properly guided.

Moreover, the literature indicates that this tension is not incidental but reflects deeper technological and linguistic constraints. The probabilistic nature of generative AI models, combined with the structural complexity of Arabic grammar, makes it difficult to ensure consistent and fully accurate outputs across contexts (Adel et al., 2026; Abdelrehim et al., 2025). This suggests that the limitations observed are not solely due to user behavior but are also rooted in the current capabilities of AI systems.

Taken together, these findings underscore the need for a balanced and critical approach to the use of generative AI in Arabic grammar learning. Rather than viewing AI as either entirely beneficial or fundamentally flawed, it should be understood as a tool with both strengths and limitations. Its effectiveness depends not only on its technical capabilities but also on how it is pedagogically integrated and critically mediated. This perspective aligns with broader discussions on the role of educators in guiding AI use and ensuring that learners engage with AI outputs reflectively.

This tension, while consistently evident across the reviewed literature, must be interpreted with awareness of the methodological diversity and limitations of the underlying studies. The pedagogical benefits are predominantly documented through perception-based and small-scale studies, while the linguistic limitations are largely identified through technical evaluations of specific AI models on constrained tasks. These two bodies of evidence stem from different research paradigms—one primarily educational and qualitative, the other computational and experimental—which limits their direct comparability. Future research that integrates both perspectives within unified study designs, such as classroom-based experiments that simultaneously measure learning outcomes and AI accuracy, would provide stronger evidence for understanding and resolving this tension.

Ultimately, the tension between pedagogical utility and linguistic reliability constitutes the central contribution of this review. By synthesizing this duality, the study provides a more nuanced understanding of generative AI in Arabic grammar learning and offers a foundation for future research and pedagogical development.

3.6. Research Gaps and Future Directions

Despite the growing body of research on the use of generative AI in education, the findings of this review reveal several critical gaps, particularly in relation to Arabic grammar learning. These gaps emerge directly from the tension identified between the pedagogical benefits of AI and its limitations in linguistic accuracy, indicating areas where current research remains insufficient.

First, there is a lack of domain-specific research focusing explicitly on Arabic grammar, particularly in *nahwu* and *shorof*. While many studies examine generative AI in language learning more broadly, relatively few directly engage with the structural and conceptual complexity of Arabic grammar. Studies such as Karima et al. (2025) and Othman and Asbulah (2025) highlight the challenges posed by Arabic morphological and syntactic systems, yet these complexities are not consistently addressed in broader AI-in-education research. This suggests a need for more targeted investigations that focus specifically on the linguistic characteristics of Arabic.

Second, there remains a significant gap in the systematic evaluation of linguistic accuracy in generative AI outputs. Although multiple studies report errors in syntactic and morphological analysis (Alkaabi & Almaamari, 2025; Tamam et al., 2024; Al-Jarf, 2025), these findings are largely descriptive and lack standardized evaluation frameworks. While emerging work (Adel et al., 2026; Abdelrehim et al., 2025) begins to address technical evaluation, there is still a lack of widely accepted benchmarks for assessing AI performance in Arabic grammar. This limits the comparability and generalizability of findings across studies.

Third, the literature reveals limited attention to the long-term pedagogical impact of generative AI use in Arabic grammar learning. Most existing studies focus on immediate outcomes such as student perceptions or short-term learning improvements (Albantani et al., 2025; Adawiyah, 2025), while neglecting how sustained AI use may influence learners' conceptual understanding, critical thinking, and grammatical competence over time. This gap is particularly important given the identified risks of overreliance and misconceptions among learners (Linur et al., 2025).

Fourth, there is insufficient exploration of pedagogical integration strategies for generative AI in Arabic grammar teaching. While existing studies acknowledge both the benefits and limitations of AI, few provide concrete instructional models or frameworks for its effective use in classroom contexts. Nugraha and Syafe'i (2025) emphasize the need for curriculum adaptation in response to AI integration, yet practical guidance for educators remains limited. Future research should therefore focus on developing pedagogically grounded models that balance AI-assisted learning with teacher guidance and critical evaluation.

Finally, the review identifies a gap in understanding the interaction between technical factors and learning outcomes. Although several studies highlight the importance of prompt design, model training, and system configuration (Adel et al., 2026; Abdelrehim et al., 2025),

their implications for actual learning processes are not sufficiently explored. Bridging this gap requires interdisciplinary research that connects computational perspectives with pedagogical outcomes.

In light of these gaps, future research should prioritize: (1) domain-specific investigations of *nahwu* and *shorof* in AI-supported learning; (2) the development of standardized frameworks for evaluating linguistic accuracy; (3) longitudinal studies examining the long-term impact of AI on learners' grammatical competence; and (4) the design of pedagogically informed models for integrating generative AI into Arabic language education. Addressing these areas will contribute to a more comprehensive and critically grounded understanding of the role of generative AI in Arabic grammar learning.

3.7. Limitations of the Review

Several limitations should be noted. The literature search was limited to English-language publications, which may have excluded relevant studies published in Arabic. Although language restrictions are common in review studies and do not necessarily affect overall conclusions (Morrison et al., 2012), this issue is particularly important in Arabic grammar learning, where a substantial portion of the literature is available in Arabic. Consequently, the findings reported here primarily represent evidence from English-language research. Future reviews should consider including Arabic-language publications to provide a more comprehensive picture of the field.

4. CONCLUSION

This review critically examined the role of generative AI in Arabic grammar learning, particularly its pedagogical benefits and limitations in linguistic accuracy. The findings indicate that generative AI has become an increasingly influential tool in supporting the learning of *nahwu* and *shorof* by enhancing conceptual understanding, providing immediate feedback, and facilitating flexible access to learning resources. These advantages demonstrate the potential of generative AI as a valuable supplementary tool in both formal and informal educational contexts. However, the review also reveals persistent limitations related to linguistic reliability, including inaccuracies in syntactic analysis, morphological processing, and contextual interpretation. These issues reflect both the complexity of Arabic grammar and the probabilistic nature of current AI systems, raising concerns about misconceptions and excessive learner dependence on AI-generated explanations.

The synthesis of the reviewed studies suggests that the integration of generative AI in Arabic grammar learning is characterized by a fundamental tension between pedagogical utility and linguistic reliability. While AI can meaningfully support learning, its limitations require careful and critical use. Therefore, generative AI should not be regarded as a replacement for traditional instruction, but rather as a complementary educational tool that must be integrated within guided pedagogical practices. This study contributes to the literature by providing a critical synthesis that simultaneously examines pedagogical opportunities and linguistic limitations within the specific context of Arabic grammar learning.

From a pedagogical perspective, the findings emphasize the continuing role of educators in guiding students to critically evaluate AI-generated content and use AI tools

responsibly. The review also highlights the importance of developing critical AI literacy among learners to reduce uncritical dependence on AI systems. Furthermore, the identified gaps indicate the need for more domain-specific research on Arabic grammar, standardized frameworks for evaluating linguistic accuracy, and pedagogically grounded models for AI integration. Overall, although generative AI offers significant potential for Arabic grammar education, its implementation requires a balanced, critical, and context-sensitive approach to ensure that AI enhances rather than undermines the quality of language learning.

ACKNOWLEDGMENT

The authors sincerely acknowledge the support and contributions of all individuals and institutions involved in the completion of this article. The authors also express their appreciation to the researchers and scholars whose studies contributed to the conceptual and theoretical foundation of this work. Any shortcomings or remaining limitations in this article remain entirely the responsibility of the authors.

REFERENCES

- Abdelrehim, M., Torki, M., & El-Makky, N. (2025). Hybrid LLM and rule-based synthetic data generation for Arabic grammatical error correction. In *2025 International Conference on Machine Intelligence and Smart Innovation (ICMISI)*. IEEE. <https://doi.org/10.1109/ICMISI65108.2025.11115884>
- 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* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2603.16718>
- 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>
- 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, Article 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 (IJOPCD)*, 15(1), 1–18.
- Baalbaki, R. (2019). Arabic linguistic tradition I: Nahw and şarf. In J. Owens (Ed.), *The Oxford handbook of Arabic linguistics* (pp. 7–21). Oxford University Press.

- 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>
- Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, Web of Science, and Google Scholar: Strengths and weaknesses. *The FASEB Journal*, 22(2), 338–342. <https://doi.org/10.1096/fj.07-9492LSF>
- 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>
- 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>
- Kwon, S. Y., Bhatia, G., Nagoudi, E. M. B., & Abdul-Mageed, M. (2023). *ChatGPT for Arabic grammatical error correction* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2308.04492>
- Linur, R., Rusdi, T., 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>
- Morrison, A., Polisena, J., Husereau, D., Moulton, K., Clark, M., Fiander, M., Mierzwinski-Urban, M., Clifford, T., Hutton, B., & Rabb, D. (2012). The effect of English-language restriction on systematic review-based meta-analyses: A systematic review of empirical studies. *International Journal of Technology Assessment in Health Care*, 28(2), 138–144. <https://doi.org/10.1017/S0266462312000086>
- 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. <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>
- 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), 1–18.

- Rishanda, A. T., Koderi, F. G., & Mizan, A. N. (2025). Artificial intelligence dalam pembelajaran nahwu secara mandiri. *Al Mi'yar: Jurnal Ilmiah Pembelajaran Bahasa Arab dan Kebahasaaraban*, 8(1), 27–39. <https://doi.org/10.35931/am.v8i1.4594>
- Rizki, R. B., Rizal, M. F., Rahmawati, C., Bachrudin, M. A., Farhani, S., & Batul, Z. (2025). Qalam Al: A study on the potential of automatic harakat detection for Arabic sentence learning. *Alsina: Journal of Arabic Studies*, 7(2), 285–316. <https://doi.org/10.21580/alsina.7.2.27500>
- 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>
- 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>
- Zubaidi, A., Munip, A., Widodo, S. A., & Zerrouki, T. (2025). Enhancing Arabic writing skills using ChatGPT-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>