Online Learning in Educational Research
Volume 6, Issue 2, 187-204.
e_ISSN: 2798-2351
https://www.journal.foundae.com/index.php/oler/index
DOI: 10.58524/oler.v6i2.1116

AI-Driven Arabic Grammatical Error Correction for Inshā’ Writing: A Systematic Review of Current Models and Pedagogical Directions

Abdurahman*
Universitas Islam Depok, INDONESIA

Siti Nurul Azkiyah
UIN Syarif Hidayatullah Jakarta, INDONESIA

Muhbib Abdul Wahab
UIN Syarif Hidayatullah Jakarta, INDONESIA

Maswani
UIN Syarif Hidayatullah Jakarta, INDONESIA

Article Info

Article history:

Received: February 27, 2026
Revised: May 1, 2026
Accepted: May 28, 2026

Keywords:

Arabic Grammatical Error Correction;
Artificial Intelligence;
Arabic Writing;
Inshā’ Learning;
Systematic Literature Review;

Abstract

Artificial intelligence (AI) has expanded opportunities for supporting Arabic writing through automated grammatical error correction (GEC). However, existing research remains fragmented, with limited synthesis connecting learner error patterns, AI-based correction models, and their pedagogical implications for Inshā’ instruction. This study systematically reviews AI-based Arabic GEC research to identify dominant nahwu–ṣarf error patterns, evaluate the evolution of AI correction models, and examine their educational implications for non-native Arabic learners. Following the PRISMA framework, eleven studies published between 2020 and 2025 were synthesized using qualitative thematic analysis and VOSviewer-assisted keyword mapping. The review identifies recurrent learner difficulties in agreement, definiteness, morphological ambiguity, syntactic dependencies, orthographic variation, and diacritical accuracy. Arabic GEC has progressed from rule-based and hybrid approaches to Transformer- and large language model (LLM)-based systems that provide more context-sensitive corrections. Nevertheless, current models remain constrained by challenges related to i‘rāb, learner-specific error profiles, explainable grammatical feedback, and limited annotated corpora. The review proposes a human-in-the-loop pedagogical framework integrating AI-generated feedback with teacher validation to foster formative assessment, learner autonomy, and reflective revision in Arabic writing instruction.

To cite this article: Abdurahman, A., Azkiyah, S. N., Wahab, M. A., & Maswani, M. (2026). AI-Driven Arabic grammatical error correction for Inshā’ writing: A systematic review of current models and pedagogical directions. *Online Learning in Educational Research, 6*(2), 187-204. https://doi.org/10.58524/oler.v6i2.1116

INTRODUCTION

Learning Arabic as a foreign or second language still proves difficult for nonnative speakers, especially in writing. The errors in grammar that learners of Arabic make most often are those associated with the structure and formation of words (Syafe’i et al., 2022). From the perspective of Error Analysis Theory, learner errors should not be viewed merely as signs of failure, but as evidence of learners’ developing linguistic systems and interlanguage competence (Corder, 1967; Ellis, 2008). These kinds of mistakes are found in relation to *i‘rāb*, gender and number agreement, definiteness, sentence structure in terms of nouns and verbs, verb pattern, and diacritical marks accuracy in Arabic composition. These errors influence not only the grammatical correctness of the student’s writing but also the clarity and coherence of the writing. The aim of *inshā’* teaching is to write an Arabic composition that is grammatically correct and semantically and rhetorically right. It is hard for students to do so because of their lack of competence in *nahwu-ṣarf* and the absence of feedback.

In terms of implementation in the classroom setting, the correction of Arabic written texts remains reliant on the role of teachers and/or lecturers. While manual correction can offer useful and valuable feedback, it is very time-consuming, unreliable, and often hard to maintain with a class

* **Corresponding Author**
Abdurahman Abdurahman, Universitas Islam Depok, INDONESIA ✉ abdurahman@uidepok.ac.id

having many students. Written corrective feedback research highlights the importance of feedback in learning writing skills, although it relies on clarity and consistency among others (Ferris, 2011; Hyland & Hyland, 2006; Karatay & Karatay, 2024). Such a requirement calls for the development of feedback mechanisms that help teachers in detecting the common grammatical mistakes that students make, while at the same time enabling students to correct their mistakes with minimal intervention. This has led to the rising importance of automated grammar checking systems in teaching Arabic writing skills (Karatay & Karatay, 2024; Weng & Chiu, 2023).

With the advancement of Artificial Intelligence (AI) and NLP, there have been some possibilities that can be explored for solving this problem. The use of a grammar checker based on AI is one such application that can be used for solving this problem (Cano et al., 2024; Hady et al., 2026; Rusmansyah et al., 2025; Zainuddin et al., 2025). In English writing instruction, tools such as Grammarly and Write & Improve have been reported to support students' writing development by offering automated feedback and revision assistance (Fitria, 2021). Nevertheless, there exist additional complexities in developing and teaching similar applications for the Arabic language owing to the complexity of morphology, flexibility of syntax, presence of diacritics which may be optional, as well as existence of syntax such as *i'rāb* (Antoun et al., 2020; ElSabagh et al., 2025). Consequently, the correction of grammatical errors in the Arabic language should not just be viewed as a technical approach whereby mistakes in terms are replaced with the right ones but entails sensitivity towards sentence structures and grammatical rules (Alhafni et al., 2023; Kwon et al., 2022).

Recent studies provide an important foundation for the development of AI-based Arabic grammatical error correction. Hanandeh et al., (2024) show that AI implementations within the realm of the Arabic language provide both potential opportunities and also pose some challenges regarding linguistic resources and language identity. The work of Al-Haddad et al. emphasizes the importance of NLP systems, machine translations, and the Arabic models of language, including AraBERT, CAMeLBERT, and AraELECTRA (Hanandeh et al., 2024). The emergence of Arabic-specific pre-trained language models has strengthened Arabic NLP research because these models are trained to better represent Arabic linguistic features than general multilingual models (Alhafni et al., 2023; Antoun et al., 2020, 2021). Earlier work by Shaalan, (2005) introduced Arabic GramCheck, which was developed as a grammar checker based on the Prolog programming language that can detect errors in nominal and verbal sentence structure. While the development of this grammar checker was indeed significant, it was still narrow in scope in that it only covered certain structures of standard Arabic. (Alhafni et al., 2023; Shaalan, 2005).

More recent research has shifted towards neural, Transformer-based, and large language model approaches. Solyman, Zhenyu, Qian, Elhag, Toseef, and Abdelaal et al, (2024), for example, developed the SCUT AGEC model for Arabic grammatical error correction using convolutional neural machine translation, synthetic data, and fine-tuning techniques. Their work demonstrates the potential of neural models to improve Arabic grammatical correction performance. Subsequent studies have shown that Transformer-based models, including AraT5 and AraBART, can improve Arabic grammatical error detection and correction, especially when supported by learner corpora, error tagging, and high-quality synthetic data (Alhafni et al., 2023; Kwon et al., 2022; Moatez et al., 2022). Other studies have also shown that large language model approaches can improve context-sensitive correction, although they continue to face difficulties in handling diacritics, syntactic dependency, *i'rāb*, and fine-grained morphological distinctions (Alrehili & Alhothali, 2025; Kwon et al., 2022). In addition, research on AI-assisted feedback in writing, including studies involving Arab EFL learners, suggests that AI-generated feedback may support grammatical accuracy and learner autonomy (Mansoor et al., 2025). However, findings from EFL contexts should be interpreted carefully because English writing feedback does not fully represent the linguistic and pedagogical complexity of Arabic *inshā'* learning (Hyland & Hyland, 2006; Karatay & Karatay, 2024).

Nevertheless, there is a lack of consistency in the research literature about automatic correction of Arabic grammar using AI. Most of the current research tends to emphasize either the technical aspects related to model building or data collection or benchmarking, but not necessarily the pedagogical implications (Alhafni et al., 2023; Karatay & Karatay, 2024; Kwon et al., 2022). Although there has been some discussion in existing literature regarding the models of NLP for Arabic, grammatical error correction tools, learner corpora, and AI feedback, the issue in question is generally addressed in isolation from other issues mentioned above. It is crucial since the grammar checking tool will become beneficial in a pedagogical setting only if the error correcting technique of the model matches the real grammatical problems that the students have and if its feedback can be used in class (Alhafni et al., 2023; Corder, 1967; Weng & Chiu, 2023).

To bridge this gap, the current review combines computational studies on the subject of Arabic Grammatical Error Correction (GEC) with pedagogical insights into *inshā'* learning by adopting a human-in-the-loop perspective. Unlike most literature on GEC that tends to concentrate purely on the performance of the models, this review focuses on integrating evidence from three perspectives (Kitchenham, 2004; Page et al., 2021; Xiao & Watson, 2019). In contrast to the research that deals solely with the performance of systems in a technical manner, this review brings together three interrelated dimensions: Arabic grammatical errors commonly made by learners, advantages and disadvantages of the AI-based grammatical error correction systems, as well as the implications of the use of such systems for teaching the Arabic language. In this sense, the uniqueness of this research is in its effort to integrate Arabic error analysis, AI-based correction, and pedagogy in one review (Alhafni et al., 2023; Corder, 1967; Weng & Chiu, 2023)

Therefore, the purpose of this study is to: (1) determine the prevalent errors in the area of *nahwu* and *ṣarf* observed in research on non-native Arabic written works; (2) provide an overview of the performance, advantages, and limitations of Arabic grammar checking using AI; and (3) derive pedagogical recommendations on the use of AI grammar checkers in the instruction of *inshā'*. Thus, the contribution of this study to research on the education of the Arabic language is that it provides insights into the appropriate utilization of AI grammar checkers in the process of writing instruction (Ferris, 2011; Karatay & Karatay, 2024; Kwon et al., 2022; Weng & Chiu, 2023).

METHOD

The Systematic Literature Review (SLR), coupled with qualitative thematic synthesis, was adopted in this study. The SLR method was chosen due to its capability of allowing previous studies to be located, evaluated, and synthesized in a systematic manner (Kitchenham, 2004; Xiao & Watson, 2019). This research being a literature review, no empirical validation of a particular Arabic grammar checker using AI is made. Rather, the existing information on AI-based Arabic grammatical error correction in relation to *nahwu–ṣarf* error pattern detection and associated challenges is compiled.

The review process followed the PRISMA 2020 framework, particularly in documenting the stages of identification, screening, eligibility, and inclusion (Page et al., 2021). This review is not pre-registered since the study is concerned with language acquisition, Arabic natural language processing, and education technology rather than clinical studies. However, the search strategy, inclusion criteria, screening process, quality assessment, data collection, and synthesis methods have been detailed below for transparency and reproducibility purposes.

The following literature review has been conducted to answer three research questions: (1) What are the main errors in *nahwu-ṣarf* in Arabic writing studies of non-native speakers? (2) Which AI approaches have been utilized for identifying and correcting Arabic grammatical errors, and what are their benefits and weaknesses? (3) What implications does the use of AI-based grammar checkers have for i*nshā'* education of non-native Arabic speakers?

A literature review has been conducted in the Scopus database, DOAJ database, and Google Scholar. Scopus was used to find journal papers and proceedings that are indexed internationally, while DOAJ has been included to locate open-access studies on the topics under discussion. Moreover, Google Scholar was employed as an additional database to locate additional scholarly works pertaining to Arabic natural language processing, grammatical error correction, and assisted language learning using artificial intelligence. The literature search has been restricted to the period from 2020 to 2025.

The search string included three keyword clusters: structure of Arabic linguistics, grammatical error correction, and technologies based on artificial intelligence. The main search string used was: ("Arabic grammar" OR "Arabic morphology" OR "Arabic syntax" OR "*nahwu*" OR "*ṣarf*") AND ("grammar checker" OR "grammatical error correction" OR "error correction" OR "error detection" OR "error tagging" OR "GEC" OR "GED") AND ("artificial intelligence" OR AI OR NLP OR "machine learning" OR "deep learning" OR Transformer OR "large language model" OR ChatGPT). The keywords were refined through preliminary searches to reduce irrelevant results, particularly studies focusing only on speech recognition, machine translation, sentiment analysis, or general AI in education. Search results were exported in CSV or RIS format and managed using Zotero to organise bibliographic data and remove duplicates.

Table 1. Inclusion and Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
Studies published between 2020 and 2025	Studies published before 2020, except foundational works used outside the SLR dataset
Studies focusing on Arabic grammatical error detection, grammatical error correction, error tagging, learner corpus development, or AI-based grammar checking	Studies focusing only on general Arabic linguistics without AI, NLP, or error correction components
Studies discussing Arabic grammar, morphology, syntax, *nahwu–ṣarf*, *i'rāb*, sentence structure, diacritics, or learner writing errors	Studies focusing only on unrelated NLP tasks, such as speech recognition, sentiment analysis, or machine translation without grammar correction
Peer-reviewed journal articles, indexed conference proceedings, or reputable scholarly publications	Opinion pieces, editorials, blog posts, and non-peer-reviewed sources
Studies reporting model development, corpus construction, evaluation metrics, error categories, or pedagogical implications	Studies that mention AI generally but do not provide evidence related to Arabic grammatical correction
Full-text articles available in English or accessible scholarly format	Articles without full-text access or with insufficient methodological information

The article selection process followed the PRISMA stages, see Figure 1. During the first round of literature search, 135 documents were found in Scopus, DOAJ, and Google Scholar databases. Out of them, 24 duplicates were excluded, leaving 111 documents to be examined based on their titles and abstracts. In the second step of screening, 69 out of 111 documents were eliminated because they had no connection to Arab grammar error correction, artificial intelligence (AI)-based grammar checking, or learners' writing tasks. Next, 42 documents went through a full-text analysis; 31 of them were left out because they did not have a clear link to Arab grammar error correction, did not contain information about error categories and model efficiency, and failed to pass the quality assessment criteria.

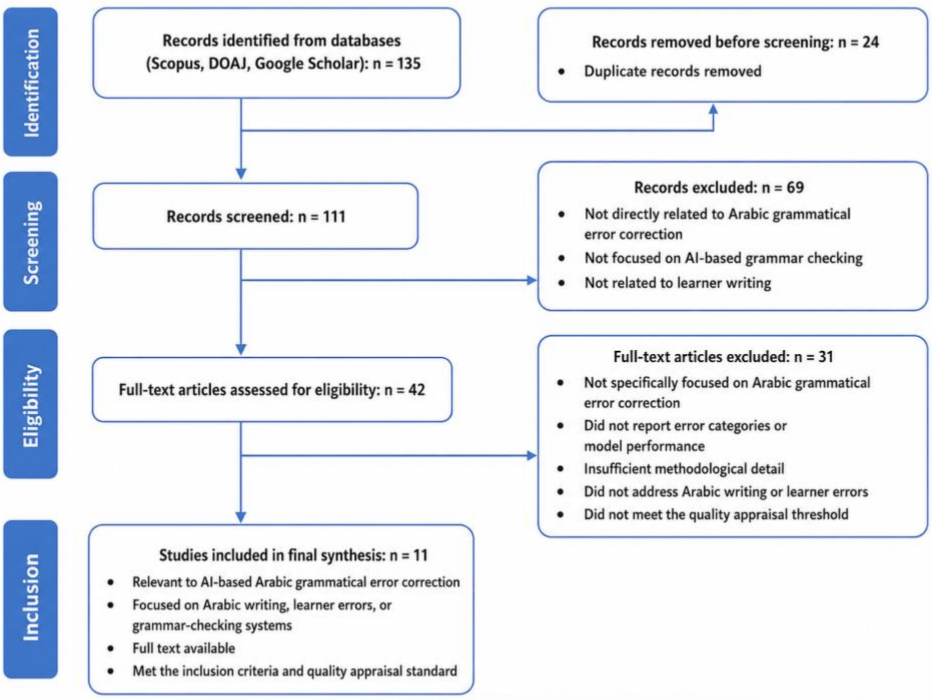

Figure 1. PRISMA Flow Chart

In order to verify the quality and credibility of the selected articles, a quality appraisal checklist was used before conducting the synthesis. This was done through the application of six criteria for assessing each study, which were borrowed from systematic review methodologies; namely, clarity of research aim, relevancy to correction of errors in Arabic grammar, clarity of source or corpus of data, clarity of method/model, results/evaluation criteria, and relevancy to Arabic writing instruction/feedback using AI (Kitchenham, 2004; Xiao & Watson, 2019). Each criterion was scored on a three-point scale: 2 = clearly fulfilled, 1 = partially fulfilled, and 0 = not fulfilled. Studies scoring 8 or higher out of 12 were included in the final synthesis. Table 2 summarises the quality appraisal criteria and the data extraction framework used to evaluate the methodological quality of the selected studies and to organise the information required for thematic synthesis.

Table 2. Quality Appraisal and Data Extraction Framework

Component	Criteria / Extracted Data
Quality appraisal	Research aim, relevance to Arabic GEC, clarity of data source/corpus, clarity of method/model, reporting of findings or metrics, pedagogical relevance
Scoring system	2 = clearly fulfilled; 1 = partially fulfilled; 0 = not fulfilled
Inclusion threshold	Studies scoring 8 or higher out of 12
Extracted bibliographic data	Author, year, publication type
Extracted research data	Research focus, data source/corpus, AI model or approach, error type, evaluation metric
Extracted synthesis data	Main findings, technical limitations, pedagogical relevance, contribution to the research questions

The selected studies were analysed using thematic analysis, which is suitable for identifying and interpreting patterns across qualitative data (Braun & Clarke, 2006). Four major stages of analysis were adopted, which consisted of reading through the included studies, generating initial codes, putting together related codes to create themes, and interpretation of the themes in regard to the research questions. These include *nahwu-ṣarf* error patterns, AI methods of correction and their performance, limitations of technology and language, and pedagogical implications of *inshā'.*

VOSviewer was also employed as an additional tool to visualize the co-occurrence of keywords and research trends among the selected literature. Before the analysis, synonymous keywords like "Arabic GEC," "Arabic grammatical error correction," and "grammar correction" were clustered together in order to eliminate any repetitions. The VOSviewer findings were solely utilized for visualizing research trends, whereas the analysis itself was mainly focused on a close analysis of the 11 selected papers.

Some limitations that need to be taken into consideration are: Firstly, the search for information via Scopus, DOAJ, and Google Scholar may be biased due to the fact that there might be other relevant studies in the Arabic language that are not available in these databases. Secondly, there is an issue of language bias due to the fact that only studies available in the English language were used. Thirdly, publication bias might arise since only studies showing good performance of the models in question get published, while those with poor or negative results do not. Lastly, only 11 studies satisfied the requirements; hence, the conclusion of this paper cannot be considered comprehensive.

RESULTS AND DISCUSSION

Research Mapping and Thematic Overview

Before presenting the thematic synthesis, this section first maps the general direction of the reviewed literature using VOSviewer. The VOSviewer analysis was used as a supplementary tool to visualise keyword co-occurrence among the eleven studies included in this review. Therefore, the visualisation should be interpreted as a descriptive map of the reviewed corpus rather than as a comprehensive bibliometric representation of the entire Arabic NLP field. Synonymous and closely related terms, such as “Arabic GEC,” “Arabic grammatical error correction,” “grammar correction,” and “error correction,” were harmonized before interpretation to reduce duplication and improve thematic clarity

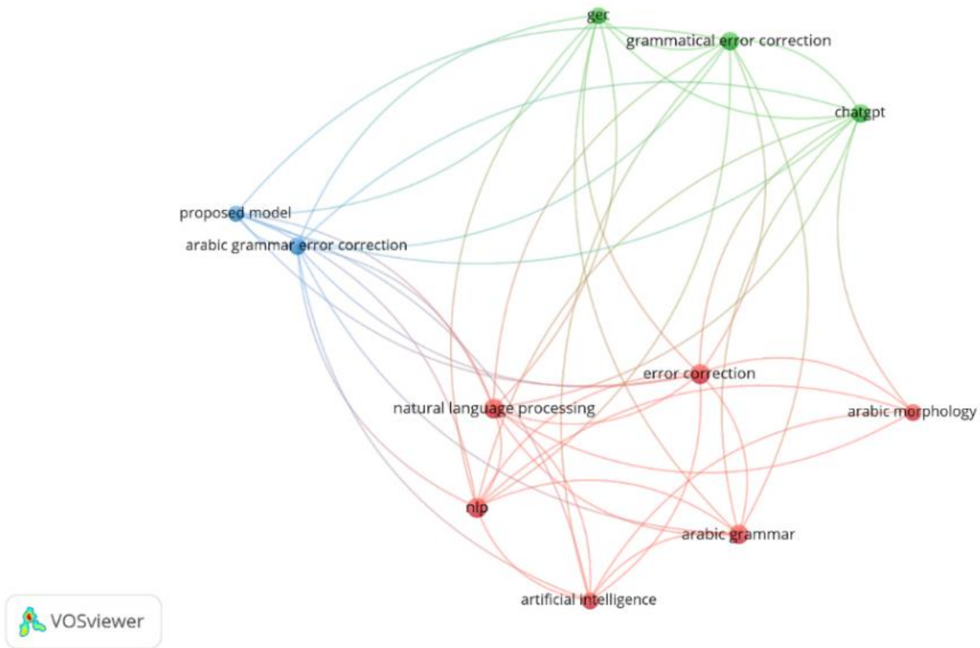

Figure 2. VOS-Viewer Display for Type of Analysis Network Visualization

Although the network visualization shows the structure and connections between the major research topics, it does not show the development of these topics through time. In order to enrich the findings with a further view, an overlay visualization was constructed using VOSviewer to show the temporal distribution of the keywords of the reviewed literature. Through colouring according to the average year of publication, the overlay visualization depicts the trend of the interests from the early rule-based and corpus-based approaches to more recent papers focusing on the Transformer architecture, LLMs, and context-based Arabic grammatical error correction.

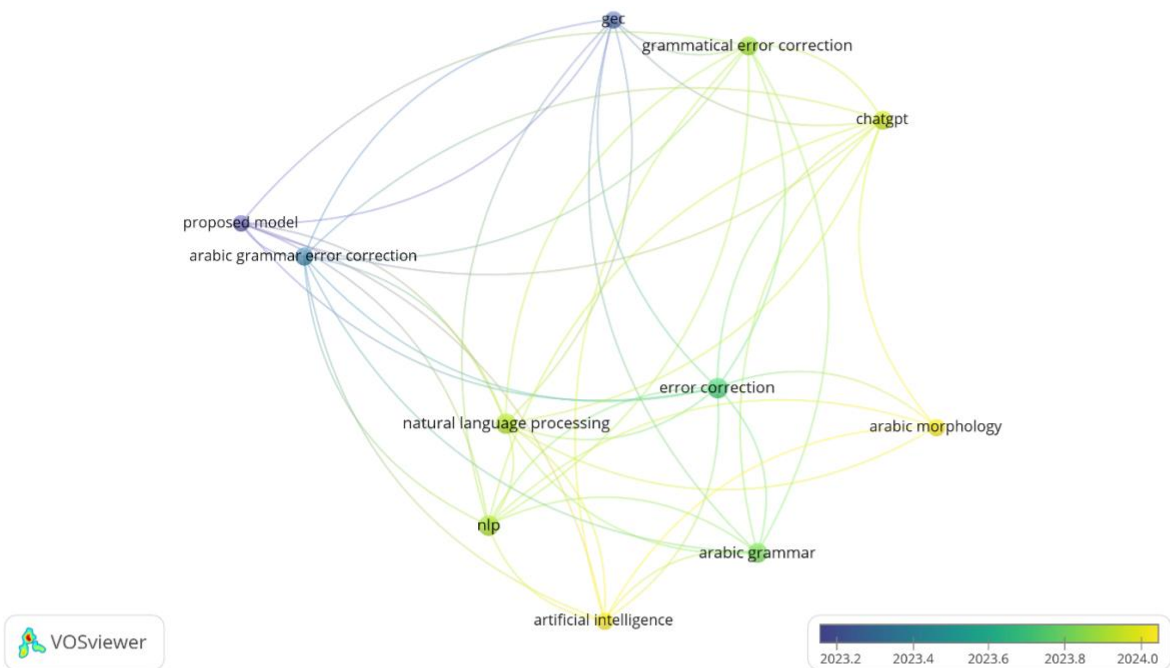

Figure 3. VOS-Viewer Display for Type of Analysis Overlay Visualization

Based on the network visualisation in Figure 2, the reviewed studies can be grouped into several interconnected clusters. The first cluster is linked with “grammatical error correction,” “GEC,” and “ChatGPT,” which means that recent studies about Arabic grammar correction have been going beyond just identifying errors (Alhafni et al., 2023; Kwon et al., 2022). The second cluster has to do with the area of correction of errors in Arabic grammar, as well as proposed models for this process. This particular cluster shows how technical the field is when it comes to designing correction models (Abdelaal et al., 2024; Solyman et al., 2021). The third category is related to NLP, Arabic morphology, Arabic grammar, and AI. This cluster emphasizes the linguistic aspect of Arabic GEC, which includes the challenge of handling the morphology, syntactic relations, diacritics, and grammatical dependencies at the sentence level (Alrehili & Alhothali, 2025; Belkebir & Habash, 2021).

From the overlay visualisation provided in Figure 3, it can be seen that there is a gradual change in the orientation of the studies reviewed. In general, earlier studies in the reviewed collection of papers have been found to highlight approaches based on rules or sequence labelling or corpora, but more recent studies are increasingly focused on exploring Transformer models, data synthesis, and applications in large language models. This does not mean that rule-based approaches have lost their importance; rather, it means that Arabic GEC research is progressing from rule-based to hybrid approaches combining linguistic rules, learner corpora annotation, and neural networks. This development is essential as grammatical errors in Arabic languages cannot be fixed using word-level correction alone; they need sentence-level processing.

In addition, the keyword map reveals that the reviewed literature is not strictly technical in nature. The keywords that demonstrate the relationship between the development of AI-based GEC for Arabic and writing pedagogy include learner errors, writing assessment, and essay scoring. Thus, there is an indication that Arabic GEC based on AI has started to be considered not merely as a computational tool but also as a future tool for teaching Arabic writing skills. However, the pedagogical aspect of Arabic GEC still requires further exploration because the majority of the reviewed articles concentrate more on the issues related to model performance, development of corpus, and correction of learner errors. This idea corresponds with the research gap indicated in the Introduction section.

The thematic synthesis of the eleven studies resulted in six major themes: neural, Transformer, and LLM-based model development; Arabic error corpus and error tagging; GEC for non-native Arabic learners; automatic grammaticality assessment of essays; structure-based correction; and synthetic data generation for low-resource Arabic GEC. These themes show that Arabic GEC research has developed in three interrelated directions: improving correction models, strengthening error data resources, and exploring possible pedagogical applications for Arabic writing instruction. The synthesis is summarised in Table 3.

Table 3. Thematic Synthesis of the Analysed Articles

Main Theme	Related Studies	Research Focus	Key Findings and Pedagogical Relevance
Neural-, Transformer-, and LLM-based GEC model development	Abdelaal et al., (2024); Musyafa et al., (2024); Solyman et al., (2021)	Development of Arabic GEC models using NMT, Transformer-based models, and LLM-based approaches.	These studies report improved correction performance through neural architectures, fine-tuning, synthetic data, and prompt-based correction. However, challenges remain in *iʿrāb*, diacritics, and fine-grained syntactic relations. Pedagogically, these models may support initial corrective feedback in *inshā’* learning, but teacher validation remains necessary.
Arabic error corpus development and error tagging	Alrehili & Alhothali, (2025); Belkebir &	Development of Arabic error annotation systems and learner-oriented corpora.	Error-tagging systems and learner corpora provide important resources for identifying recurrent grammatical errors and training Arabic GEC models. Their pedagogical relevance lies in supporting error-based

Main Theme	Related Studies	Research Focus	Key Findings and Pedagogical Relevance
	Habash, (2021)		remedial instruction and local learner corpus development.
GEC for non-native Arabic learners	Althafir & Ghnemat, (2022); Madi & Al-Khalifa, (2020)	Detection and correction of typical L2 Arabic learner errors.	The reviewed studies indicate that non-native Arabic learners show error patterns that differ from native-speaker writing, especially in agreement, definiteness, morphology, and verb forms. This finding supports the need for Arabic grammar checkers adapted to L2 learner profiles.
Automatic grammaticality assessment and essay scoring	Kwon et al., (2022)	Use of Arabic models for automatic assessment of essay grammaticality.	Automatic grammaticality assessment may help lecturers conduct preliminary diagnosis of students' writing. However, it should be used as a supporting tool rather than as a final scoring authority.
Structure-based and dependency-aware correction	Essa et al., (2025)	Use of POS tagging, dependency parsing, and sequence models for sentence-level correction.	Structure-based approaches are relevant because many Arabic errors involve sentence-level relations, such as *mubtada'–khabar*, *fi'l–fā'il,* and *maf'ūl* structures. This supports the need for grammar checkers that go beyond token-level correction.
Synthetic data generation for low-resource Arabic GEC	Alrehili & Alhothali, (2025); Musyafa et al., (2024)	Use of synthetic data to overcome limited annotated Arabic GEC resources.	Synthetic data can expand training resources for Arabic GEC, especially in low-resource settings. For educational use, this approach may inform the development of local datasets based on Indonesian learners' *nahwu–ṣarf* errors.

In general, the findings from this mapping study and thematic synthesis clearly reveal the fact that GEC of Arabic has evolved from an error detection-only framework to a more complex correction system using various techniques such as learner corpora, neural models, synthetic data, and a pedagogical feedback system. However, it is also evident from the review of the literature that there are some concerns in the field which remain unaddressed, such as learner dataset scarcity, *i'rāb* and diacritics, classroom validation, and human supervision. Therefore, the following sections discuss the findings in relation to the three research questions: dominant *nahwu–ṣarf* error patterns, AI model performance and limitations, and pedagogical implications for *inshā'* learning among non-native Arabic learners.

Dominant *Nahwu–Ṣarf* Error Patterns in Non-Native Arabic Learners' Writing

The thematic synthesis indicates that non-native Arabic learners tend to experience recurring difficulties in grammatical accuracy, particularly in areas related to morphology, agreement, definiteness, and sentence-level interpretation. However, the reviewed studies do not examine learner errors from a single perspective. Some studies focus directly on learner-generated errors, while others examine Arabic grammatical error correction through model development, corpus construction, or error annotation. Therefore, the error patterns discussed in this section should be understood as a synthesis of learner-related findings and error categories used in Arabic GEC research, rather than as direct classroom findings from all eleven studies.

Morphological and agreement-related errors are among the clearest patterns identified in the reviewed literature. Althafir & Ghnemat, (2022) show that non-native speakers learning Arabic as a second language often produce grammatical errors involving feminine–masculine agreement and definite–indefinite forms. Their study is relevant because it develops an Arabic grammatical auto-correction approach for non-native Arabic learners using a hybrid rule-based and AraBERT model. This suggests that AI-based grammar checkers for L2 Arabic learners should be sensitive to agreement and definiteness errors, rather than relying only on general spelling or surface-level correction.

The other significant pattern has to do with morphological complexity and word-form interpretation. Error correction of Arabic grammar continues to be challenging since Arabic is a language with a high degree of morphological complexity and therefore needs considerations of the relationship between root and pattern, word formation, grammatical properties, and sentence context. Research done on Arabic GEC indicates that the effectiveness of the models is influenced by the complexity of Arabic grammar (Alhafni et al., 2023; Kwon et al., 2022; Solyman et al., 2021). For this very reason, mistakes made in relation to ṣarf, including selection of word forms, agreement features, and derived structures, must be seen as aspects of morphological processing problems rather than superficial errors alone.

Sentence-level and syntactic issues should also be addressed, although this category is supported by the reviewed literature to varying degrees. Even though there are some studies that do not analyze L2 Arabic classroom writing, they prove that any Arabic GEC system should take into consideration sentence structure and grammatical relations to correct the mistakes accurately. The use of structure-aware methods, such as dependency and sequence-to-sequence correction models, proves that Arabic grammar correction should go beyond token-level replacement (Essa et al., 2025). It has relevance to *nahwu* because many of the grammatical mistakes in the Arabic language are based on the relationship between components of the sentence, including the subject-predicate relationship, verbal sentence construction, and case-based interpretation. Syntactic mistakes need to be considered as one of the most important problems for correction, but not as a discovery made in all articles.

Orthographic and diacritic-related ambiguity is another recurring issue in Arabic GEC research. The reviewed studies on Arabic error annotation and corpus development show that Arabic error categories include orthographic, morphological, syntactic, and semantic dimensions. Belkebir & Habash, (2021), through ARETA, demonstrate the importance of structured error-type annotation for Arabic texts, while Alrehili & Alhothali, (2025), through the Tibyan Corpus, include multiple error categories such as orthography, morphology, syntax, semantics, punctuation, merge, and split. These studies support the view that spelling, orthography, and diacritic-related ambiguity should not be separated from *nahwu–ṣarf* learning because they may affect grammatical interpretation.

Corpus-based and error-tagging studies further show that Arabic learner errors are increasingly being analysed through systematic annotation rather than general impressionistic observation. ARETA and the Tibyan Corpus demonstrate the importance of structured error categories for identifying recurrent grammatical patterns and improving Arabic GEC models (Alrehili & Alhothali, 2025; Belkebir & Habash, 2021). For Arabic writing pedagogy, this suggests that learner corpora can support more targeted remedial instruction, especially when the corpus reflects the actual error profile of the learners being taught.

Table 4 summarises the dominant error patterns identified from the reviewed literature. The table should be read as a qualitative synthesis of error categories and model-relevant correction challenges, not as a frequency-based ranking of learner errors.

Table 4. Dominant *Nahwu–Ṣarf* Error Patterns and Pedagogical Relevance

Error Pattern	Main Forms Identified in the Reviewed Literature	Supporting Studies	Interpretation for Arabic GEC	Pedagogical Relevance for *Inshā’* Learning
Agreement and definiteness errors	Feminine–masculine agreement, definite–indefinite confusion, agreement inconsistency	Althafir & Ghnemat, (2022)	These errors indicate that L2 Arabic learners may struggle with grammatical features that are structurally different from their first language.	Grammar checkers should provide rule-based explanations, not only corrected forms, so learners understand why agreement or definiteness must change.

Error Pattern	Main Forms Identified in the Reviewed Literature	Supporting Studies	Interpretation for Arabic GEC	Pedagogical Relevance for *Inshā'* Learning
Morphological / *ṣarf*-related errors	Word-form selection, derived patterns, verb-form confusion, root-pattern interpretation	Alhafni et al., (2023); Kwon et al., 2022; Solyman et al., (2021)	Arabic GEC requires sensitivity to morphological richness and sentence context, not merely word-level substitution.	*Inshā'* instruction should include targeted practice on *taṣrīf,* root-pattern relations, and word formation in meaningful sentence contexts.
Syntactic / *nahwu*-related correction challenges	Sentence-level grammatical relations, subject–predicate relations, verbal sentence structure, case-related interpretation	Alhafni et al., 2023; Essa et al., (2025)	Some studies show that accurate correction depends on understanding sentence structure and grammatical dependency.	AI feedback should help learners recognise sentence relations, such as *mubtada'–khabar* and *fi'l–fā'il* relations, while teachers validate the grammatical explanation.
Orthographic and diacritic-related ambiguity	Orthographic ambiguity, punctuation, merge/split issues, possible ambiguity in unvocalised forms	Alrehili & Alhothali, (2025); Belkebir & Habash, (2021)	Orthographic and diacritic-related ambiguity can affect how a system interprets morphology, syntax, and meaning.	Orthography and diacritisation should be integrated into *nahwu–ṣarf* feedback because they may influence grammatical interpretation.
Corpus-based error categories	Morphological, syntactic, orthographic, semantic, punctuation, merge, and split errors	Alrehili & Alhothali, (2025); Belkebir & Habash, (2021)	Structured annotation makes learner errors more visible and supports the development of more reliable Arabic GEC systems.	Learner corpora can be used to design remedial materials, class error maps, and local grammar-checking tools for specific learner groups.

Overall, the reviewed studies suggest that the dominant error patterns relevant to non-native Arabic writing include agreement-related errors, definiteness errors, morphological ambiguity, sentence-level grammatical relations, and orthographic or diacritic-related ambiguity. These results are useful for answering RQ1 since they reveal that the Arabic writing mistakes are not limited to individual words. Rather, they arise from the complex combination of *ṣarf, nahwu*, orthography, and the sentence context. This means that for *inshā'* teaching purposes, the AI-powered grammar checkers have to be developed not only as tools for identifying errors, but also for giving explanations as to why they occur. However, teachers still need to verify such information in order to confirm its correctness.

AI-Based Arabic GEC Models: Performance, Strengths, and Limitations

Over the past decade, the capacity of AI systems for detecting and fixing errors in Arabic grammar has evolved. The reviewed studies reveal that Arabic grammatical error detection and correction with the help of AI has evolved from a number of modeling methods, including rule-based, sequence-labeling, neural network, Transformer-based, and LLM-based. Whereas the previous section is concerned with the error patterns made by learners, this section deals with the way various AI models try to find and fix these errors and what advantages and disadvantages have been mentioned in the reviewed studies. As the reviewed studies employed different sets of data, measures of evaluation, and correction tasks, it would be wrong to compare their results.

Early Arabic GEC approaches relied on rule-based, sequence-labelling, or hybrid methods. These approaches are useful for targeting specific grammatical categories because they can incorporate explicit linguistic rules. Madi & Al-Khalifa, (2020) for example, demonstrate the use of neural sequence labelling for Arabic grammatical error detection, while Althafir & Ghnemat, (2022) use a rule-based correction system along with AraBERT to correct certain errors that arise from non-native Arabic speakers. The key advantage of using this approach is that they can explain why a certain grammatical structure might be wrong. However, they lack effectiveness when it comes to broader context, ambiguity in meaning, or syntax.

More recent studies have shifted towards neural machine translation and Transformer-based models. Solyman et al, (2021) show that synthetic data can support Arabic grammar correction in low-resource conditions, while Alhafni et al, (2023) show that Transformer architectures can help to enhance context-aware correction. In comparison with previous systems, these systems are better at creating more natural and fluent corrections as well as processing broader sentence contexts. Still, the performance of these systems is very much dependent on the quality of the training data and annotated error corpora for Arabic learners.

LLM-based correction represents a further development in Arabic GEC. Kwon et al, (2022) show that Correction suggestions can be made using approaches that are based on ChatGPT regarding the grammatical mistakes in the Arabic language. The key advantage of LLMs is that they provide flexible corrections in a context-aware manner. However, LLM-based correction suggestions need to be used with caution since they may provide misleading yet plausible explanations, overcorrect acceptable constructions, or lack fine-grained application of the grammatical rules of the Arabic language.

The reviewed studies also show that model performance is strongly shaped by corpus quality and error annotation. Belkebir & Habash, (2021), through Alrehili & Alhothali, (2025), through the Tibyan Corpus, demonstrate that structured error annotation is crucial for improving Arabic GEC. It appears, therefore, that in addition to architectural features, the success of Arabic grammar checkers requires having accurate error types, learner data, and annotated texts validated by experts; otherwise, even the most sophisticated systems would not be able to detect learners' errors.

Another important limitation concerns sentence-level correction. Essa et al, (2025) indicate that the correction of grammatical mistakes in Arabic is made more effective by the use of structure-sensitive approaches because certain mistakes can only be corrected by understanding the meaning of the grammatical relationship between words rather than just replacing them. Arabic systems for grammar error correction must move from token-based correction to sentence-level correction when it comes to syntax dependency and grammatical role identification.

Table 5 summarises the main model types identified in the reviewed studies. The table focuses on model performance characteristics and limitations, rather than repeating the learner error categories discussed in the previous section.

Table 5. AI-Based Arabic GEC Models: Strengths and Limitations

Model / Approach	Related Studies	Main Strengths	Main Limitations
Rule-based and hybrid correction	Althafir & Ghnemat (2022); Madi & Al-Khalifa (2020)	Interpretable and useful for targeting selected grammatical categories.	Limited in handling wider context, semantic ambiguity, and complex sentence relations.
Neural sequence labelling and NMT-based models	Madi & Al-Khalifa (2020); Solyman et al. (2021)	Can improve automatic detection and correction beyond purely rule-based systems.	Dependent on dataset quality and may struggle with learner-specific or sentence-level errors.

Model / Approach	Related Studies	Main Strengths	Main Limitations
Transformer-based Arabic GEC models	Alhafni et al. (2023); Abdelaal et al. (2024); Musyafa et al. (2024)	Better at modelling sentence context and producing fluent corrections.	Still challenged by limited annotated data, diacritics, morphology, and syntactic relations.
LLM-based correction	Kwon et al. (2023)	Flexible and capable of generating context-aware correction suggestions.	May produce overcorrection, inconsistent explanations, or grammatically plausible but inaccurate outputs.
Corpus and error-tagging systems	Belkebir & Habash (2021); Alrehili & Alhothali (2025)	Provide structured error categories and improve model training resources.	Require expert validation and may not represent all learner populations.
Structure-aware correction	Essa et al. (2025)	Supports sentence-level correction and grammatical relation analysis.	Needs further validation in learner-writing and classroom contexts.

However, from an analysis of the literature reviewed above, AI-powered Arabic GEC has achieved considerable development, especially via Transformers, Large Language Model-driven correction, synthesis of training data, and structured error tagging. Despite the advancements, there are various challenges facing this research field, such as a lack of learner-related datasets, Arabic grammar intricacies, inconsistency in explanations provided, and the requirement for expert validation. These findings answer RQ2 by showing that AI-based Arabic GEC has strong potential, but it should not yet be treated as a fully autonomous correction system.

Pedagogical Implications for *Inshā'* Learning

These studies indicate that there is significant pedagogical value of AI-based Arabic Grammatical Error Correction (GEC) in *inshā'* learning. However, such value should be seen as supplementary rather than substitutionary. The role of Arabic grammar checkers cannot be viewed as that of an independent authority which takes the place of the teacher but that of an assistance which enables the learners to detect errors in their writings, revise them, and reflect upon the grammatical principles behind them.

Firstly, AI-assisted GEC of Arabic can be used for formative feedback during the process of teaching Arabic writing skills. As it happens in traditional *inshā'* lessons, students usually have to rely on teachers' correction of their written works prior to any revision of them. The use of automated feedback could reduce the time needed for this since it would detect potential errors and give preliminary corrections (Alhafni et al., 2023; Alrehili & Alhothali, 2025; Kwon et al., 2022). However, the above feedback should only be taken as a basis for improvement and not as a definitive assessment of correctness. The students should learn how to evaluate AI feedback against *nahwu–ṣarf* principles, dictionaries, textbooks, and even teacher feedback.

Second, Arabic GEC can support error-based remedial instruction. Error-tagging and learner corpus studies show that Arabic errors can be classified into orthographic, morphological, syntactic, semantic, and structural categories (Alrehili & Alhothali, 2025; Belkebir & Habash, 2021). This is beneficial in the classroom since the teacher can design exercises based on repetitive mistakes made by students. In the case where there are repetitive mistakes in areas such as agreement, definiteness, or verbs, the teacher may design exercises in these areas. Consequently, error analysis is not only a diagnostic tool but also one used in teaching grammar.

Third, grammar checkers that use AI might enhance learner autonomy and revision skills. In many writing classes, students generally see corrections as the teacher's job. Automated feedback might encourage students to self-correct their writing before submitting their papers. This might make students think about what they write in terms of their grammar, syntax, and meanings. Nevertheless, autonomy does not mean that AI should be used without any supervision from teachers. Learners should learn how to judge the feedback generated by AI and justify why they have made some changes based on Arabic grammar.

Fourth, the discussed research indicates the need for constructing local learner corpora in the context of non-native Arabic. Many Arabic grammar error correction systems depend on benchmark corpora or general corpora, which do not fully capture the mistakes of the particular learners. In the case of the teaching institutions that offer Arabic classes to Indonesian learners, the *inshā'* texts written by the students would be an asset in developing the learner corpora (Musyafa et al., 2024; Solyman et al., 2021), but local learner data and expert validation remain necessary to ensure that the system reflects actual classroom needs.

Fifth, AI-based assessment tools may assist teachers in preliminary diagnosis, but they should not replace teacher evaluation (Putri et al., 2025). Automatic grammaticality assessment and essay scoring can help lecturers identify general weaknesses in students' writing more efficiently (Tang & Mahmoud, 2021). Nevertheless, *inshā'* testing is not limited to the aspect of grammatical correctness but encompasses other aspects such as meaning, coherence, rhetoric, vocabulary usage, and learning targets. Consequently, machine-based scoring or diagnostics should only serve as supplemental data, leaving the evaluation process up to the teacher's discretion.

Lastly, the implementation of Arabic GEC technology into learning systems ought to be done using a human-in-the-loop approach. In such an approach, the system uses AI technology to suggest possible mistakes and solutions; the learners work on those suggestions and reflect on them, while the teachers ensure that the corrections and the grammar of the output product are accurate. This is especially necessary since there are high chances that AI might have problems with issues like *i'rāb*, diacritics, syntactic dependency, semantics, and traditional grammar. The pedagogical implications identified from the reviewed studies can be organised into several instructional areas. Table 6 summarises these implications together with their supporting studies and recommended classroom applications.

Table 6. Pedagogical Implications of Arabic GEC for *Inshā'* Learning

Pedagogical Area	Implication for *Inshā'* Learning	Supporting Studies	Recommended Use
Formative feedback	AI can provide preliminary correction suggestions before teacher feedback.	Abdelaal et al., (2024); Alhafni et al., (2023); Kwon et al., (2022)	Use AI feedback as the first revision layer, followed by teacher validation.
Error-based remediation	Error categories can help teachers design focused grammar exercises.	Alrehili & Alhothali, (2025); Belkebir & Habash, (2021)	Use class error patterns to plan remedial lessons on agreement, definiteness, morphology, or syntax.
Learner autonomy	Automated feedback can encourage students to revise their own writing more actively.	Alhafni et al., (2023); Kwon et al., (2022)	Ask students to explain why they accept or reject AI suggestions.
Local learner corpus development	Students' *inshā'* texts can be collected and annotated to identify recurring local error patterns.	Alrehili & Alhothali, (2025); Belkebir & Habash, (2021); Musyafa et al., (2024)	Build institutional learner corpora to support contextual Arabic GEC systems.
Preliminary essay diagnosis	AI-based assessment may help identify grammatical weaknesses in Arabic essays.	Tang & Mahmoud, (2021)	Use automatic scoring or diagnostic comments as support, not as the final grade.
Human-in-the-loop validation	Teacher validation remains necessary for grammatical accuracy and pedagogical appropriateness.	Alhafni et al., (2023); Essa et al., (2025); Kwon et al., (2022)	Combine AI correction, student revision, and teacher explanation in a staged writing process.

Based on Table 6, these implications, AI-based Arabic GEC should be integrated into *inshā'* learning through a staged pedagogical model. In the first stage, students draft their Arabic text independently. In the second stage, they use AI-based feedback to identify possible errors and make initial revisions. In the third stage, students compare the feedback with relevant *nahwu–ṣarf* rules and explain their revisions. In the final stage, the teacher reviews the revised text, validates the correction, and provides deeper grammatical explanation. This staged model prevents AI from becoming a shortcut for writing tasks and instead turns it into a tool for guided revision, grammatical reflection, and learner development.

Proposed Pedagogical Framework

In order to implement this model, this research presents some design examples for prompt designs for LLM-based Arabic grammar feedback. The prompt designs here have not been empirically tested, yet these examples are used only to show the instructional scenarios based on the findings from the SLR on Arabic GEC, error annotation, automatic essay evaluation, learner corpora, and human-in-the-loop feedback. The purpose of the prompts is to facilitate better systematic use of generative AI technology by the teachers and learners. To operationalise the proposed human-in-the-loop model, Table 7 presents example prompt designs that illustrate how large language models (LLMs) may be used to support Arabic grammatical feedback in *inshā’* learning. These prompts are conceptual examples derived from the findings of this review rather than empirically validated instructional tools.

Table 7. Prompt Design for LLM-Assisted Arabic Grammar Feedback in *Inshā’* Learning

No	Prompt Function	Pedagogical Purpose	Prompt Template	Required Safeguard
1	*Nahwu–Ṣarf* Correction	To identify and correct syntax and morphology errors in students' Arabic writing.	"Check the following Arabic text for *nahwu* and *ṣarf* errors. Provide: (1) corrected version, (2) list of errors, (3) explanation of each correction based on Arabic grammar rules, and (4) alternative correct sentences. Text: «…»"	Teacher must verify whether the correction follows accepted *nahwu–ṣarf* rules.
2	Error Tagging	To classify student errors into systematic categories.	"Identify all errors in the following Arabic text using these categories: morphology, syntax, orthography, diction, sentence structure, and meaning. Present the result in a table: Error – Category – Explanation – Correction. Text: «…»"	Categories should be checked against the error taxonomy used in the course or study.
3	*I'rāb* and Diacritisation Check	To help students understand case endings and diacritic-sensitive grammar.	"Add appropriate diacritics to the following Arabic text and identify words that may cause *i'rāb* errors. Present the result in a table: Word – *I'rāb* position – Reason – Correction if needed. Text: «…»"	AI-generated *i'rāb* must be reviewed by the lecturer because this is a high-risk area for error.
4	Morphological Analysis	To analyse word formation, root, pattern, and function.	"Analyse the following Arabic words from a *ṣarf* perspective. Provide: root, *wazan*, form, morphological change, and function in the sentence. Words/Text: «…»"	Use as learning support, not as the only source for *ṣarf* explanation.
5	Sentence Structure Analysis	To analyse nominal and verbal sentence relations.	"Analyse the sentence structure of the following Arabic text. Identify nominal sentences, verbal sentences, *mubtada'*, *khabar, fi'l*, *fā'il*, *maf'ūl*, and possible structural errors. Text: «…»"	Lecturer should validate sentence relations, especially in complex sentences.

No	Prompt Function	Pedagogical Purpose	Prompt Template	Required Safeguard
6	Adaptive *Inshā'* Practice	To generate exercises based on students' recurring weaknesses.	"Create *inshā'* exercises focusing on these weaknesses: [agreement/definiteness/*i'rāb*/*idāfah* /verb forms]. Include: (1) short model paragraph, (2) completion exercise, (3) correction task, and (4) answer key with explanation."	Exercises should be reviewed before being assigned to students.
7	Preliminary Essay Feedback	To provide formative feedback before teacher assessment.	"Evaluate the following Arabic essay for grammatical accuracy, coherence, vocabulary choice, and semantic clarity. Do not give a final grade. Provide diagnostic comments and revision suggestions. Essay: «…»"	Avoid using AI output as final grading; use it only for formative diagnosis.
8	Local Learner Corpus Preparation	To prepare structured data from student errors.	"From the following student sentences, create a corpus table with these columns: original sentence, corrected sentence, error category, grammatical explanation, and possible remedial focus. Sentences: «…»"	Student data must be anonymised and checked by experts before reuse.
9	Reflective Revision	To train students to explain their revisions.	"Compare my original Arabic sentence and the corrected version. Explain what changed, why it changed, and which *nahwu–ṣarf* rule is involved. Original: «…» Corrected: «…»"	Students should justify the correction using course materials.
10	Non-Hallucinatory Feedback Control	To reduce unsupported AI explanations.	"Provide feedback only based on the Arabic text below. Do not invent errors that are not present. If uncertain, write 'needs teacher verification'. Text: «…»"	Useful for reducing overconfident or fabricated explanations.

Table 7 shows the prompts that can operationalize the human-in-the-loop model in *inshā'*. In this model, generative AI offers initial feedback to the learners, who subsequently improve their writing through reflective processes, after which their teacher verifies the correctness of grammar and teaching in their end product. The prompts should therefore not be seen as a replacement of the Arabic teacher, but as feedback aids in Arabic writing lessons.

The educational significance of AI-enabled Arabic GEC, therefore, consists of its capacity to facilitate feedback, diagnosis, revision, and awareness of errors. The use of AI grammar checkers in the process of *inshā'* should always be contextualized, guided, and pedagogically driven. The results obtained from this study provide an answer to RQ3 by proving that AI grammar checkers can become an integral part of Arabic writing instruction under certain conditions.

LIMITATIONS

Despite this being an evidence-based analysis of the literature regarding the topic of AI-based Arabic GEC, some constraints must be mentioned. Firstly, the number of studies analyzed is rather limited, since this paper considers only eleven studies carried out from 2020 to 2025. Therefore, the results presented might not cover the fast-developing area of Arabic natural language processing (NLP) and might not include preliminary research works, industry-related innovations, articles published in the Arabic language, as well as local studies which were not included in the chosen databases. Secondly, this paper uses solely secondary data collected from previous research and lacks observations and experiments conducted in the classrooms, as well as interviews of the participants of the learning process. Thirdly, most of the reviewed studies have been done beyond the Indonesian Arabic learning context, which implies that the datasets, profiles of learners, and institutions involved may be different from those of Indonesian learners, especially Arabic Education students. Fourth, the reviewed studies use different datasets, error typologies, model architectures, and evaluation measures, implying that comparisons between models' performances are difficult to make; hence, readers should interpret the discussion in this paper as thematic synthesis and not statistical analysis or meta-analysis. Lastly, due to rapid advancements in AI-based language correction, especially through the application of Transformer and Large Language Models, some technical results may need updating in light of emerging models, datasets, and measures. Therefore, future research work needs to construct a local Arabic learner corpus, test AI-based grammar checkers based on the *inshā'* writings of Indonesian students, assess their effects on writing development, and construct human-in-the-loop learning models.

CONCLUSION

This systematic literature review synthesised eleven studies published between 2020 and 2025 to examine AI-based Arabic Grammatical Error Correction (GEC) and its implications for *inshā'* learning among non-native Arabic learners. The review indicates that the most recurrent learner errors involve agreement, definiteness, morphological ambiguity, sentence-level grammatical relations, and orthographic or diacritical ambiguity. Across the reviewed studies, Arabic GEC has evolved from rule-based and hybrid approaches to neural, Transformer-based, and large language model (LLM)-based systems, enabling more context-sensitive correction while continuing to face challenges related to *i'rāb*, diacritics, syntactic dependencies, learner-specific error profiles, and the reliability of grammatical explanations.

This review also demonstrates that AI-based Arabic GEC has considerable pedagogical potential when implemented as a formative feedback tool that supports learner autonomy, revision, and error-based instruction. Rather than replacing teachers, AI should be integrated through a human-in-the-loop approach in which automated feedback is critically evaluated by learners and validated by instructors. By synthesizing learner error patterns, AI-based correction models, and pedagogical perspectives within a single framework, this review contributes to a more comprehensive understanding of how AI can support Arabic writing instruction. Future research should focus on developing learner corpora based on Indonesian Arabic learners, empirically evaluating AI-assisted grammar correction in classroom settings, and designing instructional models that effectively integrate AI with teacher-guided feedback.

AUTHOR CONTRIBUTION

Conceptualization: AA, SNA; Methodology: AA, MAW; Literature search and data curation: AA; Formal analysis: AA, MAW; Visualization: AA; Writing original draft preparation: AA; Writing review and editing: SNA, MAW, and MM; Supervision: MAW and MM. All authors have read and approved the final version of the manuscript.

ACKNOWLEDGEMENT

The authors would like to thank colleagues who provided valuable comments and suggestions during the preparation of this manuscript. Their feedback has contributed to improving the quality of this study.

REFERENCES

Abdelaal, A., Medhat, A. A., Elsayad, M., Foad, M., Khaled, S., Tamer, A., & Medhat, W. (2024). Text correction for modern standard Arabic. *Procedia Computer Science*, *244*, 371–377. https://doi.org/10.1016/j.procs.2024.10.211

Alhafni, B., Inoue, G., Khairallah, C., & Habash, N. (2023). Advancements in Arabic grammatical error detection and correction : An empirical investigation. *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 6430–6448. https://doi.org/10.18653/v1/2023.emnlp-main.396

Alrehili, A., & Alhothali, A. (2025). Tibyan corpus: Balanced and comprehensive error coverage corpus using ChatGPT for Arabic grammatical error correction. *PeerJ Computer Science*, *11*, e2724. https://doi.org/10.7717/peerj-cs.2724

Althafir, Z., & Ghnemat, R. (2022). A Hybrid approach for auto-correcting grammatical errors generated by non-native Arabic speakers. *2022 International Conference on Emerging Trends in Computing and Engineering Applications (ETCEA)*, 1–6. https://doi.org/10.1109/ETCEA57049.2022.10009874

Antoun, W., Baly, F., & Hajj, H. (2020). AraBERT : Transformer-based model for Arabic language understanding. *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools*, *May*, 9–15.

Antoun, W., Baly, F., & Hajj, H. (2021). AraELECTRA: Pre-training text discriminators for Arabic language understanding. *Proceedings of the Sixth Arabic Natural Language Processing Workshop*, 191–195.

Belkebir, R., & Habash, N. (2021). Automatic error type annotation for Arabic. *Proceedings of the 25th Conference on Computational Natural Language Learning*, 596–606. https://doi.org/10.18653/v1/2021.conll-1.47

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

Cano, P. M. A., Catiggay, J. M. L., Florida, J. M. M., Pua, J. M. R., & Samonte, M. J. C. (2024). Natural language processing of grammar checker tools for academic writing : A systematic literature review. *Journal Electrical Systems*, *20*(7), 1151–1156. https://doi.org/10.52783/jes.3611

Corder, S. P. (1967). The significance of learner's errors. *International Review of Applied Linguistics in Language Teaching*, *5*(1–4), 161–170. https://doi.org/doi:10.1515/iral.1967.5.1-4.161

Ellis, R. (2008). *The study of second language acquisition*. Oxford University Press.

ElSabagh, A. A., Azab, S. S., & Hefny, H. A. (2025). A comprehensive survey on Arabic text augmentation: Approaches, challenges, and applications. *Neural Computing and Applications*, *37*(10), 7015–7048. https://doi.org/10.1007/s00521-025-11020-z

Essa, N., El-Gayar, M., & El-Daydamony, E. (2025). Arabic grammar correction for Arabic text summaries. *Mansoura Journal for Computer and Information Sciences*, *20*. https://doi.org/10.21608/mjcis.2025.353920.1009

Ferris, D. (2011). Treatment of error in second language student writing. *Bibliovault OAI Repository, the University of Chicago Press*. https://doi.org/10.3998/mpub.2173290

Fitria, T. N. (2021). " Grammarly " as AI -powered English writing assistant : Students ' alternative for English writing. *Metathesis: Journal of English Language, Literature, and Teaching*, *5*(1), 65–78. https://doi.org/10.31002/metathesis.v5i1.3519

Hady, Y., Azkiyah, S. N., Muhbib, M., & Maswani, M. (2026). Artificial intelligence-based Arabic language learning: A systematic study of the development and challenges of pedagogical innovation. *Online Learning in Educational Research (OLER)*, *6*(1), 49–63. https://doi.org/10.58524/oler.v6i1.976

Hanandeh, A., Ayasrah, S., Kofahi, I., & Qudah, S. (2024). Artificial Intelligence in Arabic linguistic landscape : Opportunities , challenges , and future directions. *TEM Journal.*, *13*(4), 3137–3145. https://doi.org/10.18421/TEM134-48

Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching*, *39*(2), 83–101. https://doi.org/10.1017/S0261444806003399

Karatay, Y., & Karatay, L. (2024). Automated writing evaluation use in second language classrooms: A research synthesis. *System*, *123*, 103332. https://doi.org/10.1016/j.system.2024.103332

Kitchenham, B. (2004). Procedures for performing systematic reviews. In *Keele, UK, Keele Univ.* (Vol. 33).

Kwon, S. Y., Bhatia, G., Nagoud, E. M. B., & Abdul-Mageed, M. (2022). ChatGPT for Arabic grammatical error correction. *ArchivePrefix*, *23*(8).

Madi, N., & Al-Khalifa, H. (2020). Error detection for Arabic text using neural sequence labeling. In *Applied Sciences* (Vol. 10, Issue 15, p. 5279). https://doi.org/10.3390/app10155279

Mansoor, K. R., Mousavi, K., & Shokri, A. (2025). Investigating the effects of AI and teacher-based explicit correction on learner autonomy and grammatical accuracy. *Forum for Linguistic Studies*, *07*(08), 498–510. https://doi.org/10.30564/fls.v7i8.9643

Moatez, E., Nagoudi, B., & Abdul-mageed, M. (2022). AraT5 : Text-to-text transformers for Arabic language generation. *Proceedings of the 60th Annual Meeting of the Association for*

Computational Linguistics (Volume 1: Long Papers), *1*, 628–647. https://doi.org/10.18653/v1/2022.acl-long.47

Musyafa, A., Gao, Y., Solyman, A., Khan, S., Cai, W., & Khan, M. F. (2024). Dynamic decoding and dual synthetic data for automatic correction of grammar in low-resource scenario. *PeerJ Computer Science*, *10*, e2122. https://doi.org/10.7717/peerj-cs.2122

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*(1), 71. https://doi.org/10.1136/bmj.n71

Putri, R. K. A., Kusaeri, K., & Suparto, S. (2025). Deep learning in automated essay scoring for Islamic education: A systematic review. *Online Learning in Educational Research (OLER), 5*(2), 319–338. https://doi.org/10.58524/oler.v5i2.753

Rusmansyah, Syahmani, Erika, F., Kusuma, A. E., & TienTien, L. (2025). Integrating AR–AI–STEAM for 6C skill development in wetland learning: Trends and knowledge mapping from 2019–2025. *Online Learning in Educational Research*, *5*(2), 457–477. https://doi.org/10.58524/oler.v5i2.945

Shaalan, K. F. (2005). Arabic gramcheck: A grammar checker for Arabic. *Software: Practice and Experience*, *35*(7), 643–665. https://doi.org/10.1002/spe.653

Solyman, A., Zhenyu, W., Qian, T., Elhag, A. A. M., Toseef, M., & Aleibeid, Z. (2021). Synthetic data with neural machine translation for automatic correction in arabic grammar. *Egyptian Informatics Journal*, *22*(3), 303–315. https://doi.org/10.1016/j.eij.2020.12.001

Syafe'i, I., Nada, N., Fauziah, P., & Azizah, Z. (2022). Tahlīl al-akhtā' al-sharfiyyah wa al-nahwiyya fī al-kitāb al-'arabiyyah li dars al-insyā. *Tadris Al- ' Arabiyyah*, *1*(1), 54–73. https://doi.org/10.15575/ta.v1i1.17383

Tang, L., & Mahmoud, Q. H. (2021). A survey of machine learning-based solutions for phishing website detection. *Machine Learning and Knowledge Extraction*, *3*(3), 672–694. https://doi.org/10.3390/make3030034

Weng, X., & Chiu, T. K. F. (2023). Instructional design and learning outcomes of intelligent computer assisted language learning: Systematic review in the field. *Computers and Education: Artificial Intelligence*, *4*, 100117. https://doi.org/10.1016/j.caeai.2022.100117

Xiao, Yu, & Watson, Maria. (2019). Guidance on conducting a systematic literature review. *Journal of Planning Education and Research*, *39*(1), 93–112. https://doi.org/10.1177/0739456X17723971

Zainuddin, Z., Misbah, M., Mastuang, M., Qamariah, Q., Zaidi, M., Amiruddin, B., Farahwahidah, N., & Rahman, A. (2025). Mapping a decade of research on artificial intelligence and augmented reality in physics education: A bibliometric analysis (2016–2025). *Online Learning in Educational Research*, *5*(2), 537–557. https://doi.org/10.58524/oler.v5i2.901