Urwatul Wutsqo: Jurnal Studi Kependidikan dan Keislaman
Vol. 15 No. 2, July 2026
P-ISSN: **2252-6099**; E-ISSN: **2721-2483**
DOI: https://doi.org/10.54437/juw
Journal Page: https://pondokjurnal.uwj.ac.id/index.php/UrwatulWutsqo

The Accuracy of "Nahwu AI" in Arabic Syntactic Analysis (*Tarkib*): An Evaluative Study Based on Ibn Malik's Perspective

Walidatur Rohmah[1], Isna Finurika[2]
Universitas Sunan Drajat Lamongan, Indonesia[12]
walidaa064@gmail.com,[1] isnafinurika91@gmail.com[2]

Abstract: Advances in artificial intelligence (AI) for Arabic language learning have encouraged the development of automated syntactic analysis platforms, including Nahwu AI. However, the extent to which its analytical output conforms to the principles of classical Arabic grammar remains insufficiently evaluated. This study aims to assess the conformity of Nahwu AI's syntactic analysis with the grammatical rules of Alfiyyah Ibn Malik and to identify the platform's strengths and limitations. Employing a qualitative evaluative research design, the study adopted a critical-comparative analysis by comparing the output generated by Nahwu AI with the grammatical principles presented in Alfiyyah Ibn Malik and Syarh Ibn Aqil as normative references. The dataset comprised 30 Arabic sentences selected through purposive sampling and analyzed using the Nahwu AI platform. The findings indicate that 20 sentences (66.7%) were classified as accurate, 5 (16.7%) as moderately accurate, and 5 (16.7%) as inaccurate. The highest accuracy was achieved in tawabi' and idhafah constructions (80%), followed by jumlah fi'liyyah and jumlah syarthiyyah (75%), whereas jumlah ismiyyah showed the lowest accuracy (37.5%). Most errors involved identifying verbal sentence structures, khabar muqaddam, syibh al-jumlah predicates, mudhaf–mudhaf ilaih relations, and isim syarth functions. These findings suggest that Nahwu AI effectively recognizes basic syntactic structures but remains limited in handling grammatically complex constructions requiring deeper linguistic reasoning. This study contributes a normative evaluation framework that bridges AI-based syntactic analysis with the principles of classical Nahwu, providing implications for the development of more linguistically reliable Arabic learning technologies.

Keywords: Artificial Intelligence, Nahwu AI, Arabic Syntax, Alfiyyah Ibn Malik, Classical Nahwu.

Abstrak: *Perkembangan kecerdasan buatan (Artificial Intelligence/AI) dalam pembelajaran bahasa Arab telah mendorong lahirnya berbagai platform analisis sintaksis otomatis, salah satunya Nahwu AI. Namun, tingkat kesesuaian hasil analisis sintaksis yang dihasilkan sistem tersebut dengan kaidah nahwu klasik masih memerlukan evaluasi normatif. Penelitian ini bertujuan mengevaluasi kesesuaian hasil analisis sintaksis Nahwu AI dengan kaidah yang terdapat dalam Alfiyyah Ibn Malik serta mengidentifikasi keunggulan dan keterbatasan sistem tersebut. Penelitian menggunakan pendekatan kualitatif dengan jenis penelitian evaluatif melalui metode analisis kritis-komparatif.*

Hasil analisis Nahwu AI dibandingkan dengan kaidah nahwu yang terdapat dalam Alfiyyah Ibn Malik dan Syarh Ibn Aqil sebagai standar normatif. Data penelitian berupa 30 kalimat bahasa Arab yang dipilih secara purposive dan dianalisis menggunakan platform Nahwu AI. Hasil penelitian menunjukkan bahwa 20 kalimat (66,7%) tergolong akurat, 5 kalimat (16,7%) cukup akurat, dan 5 kalimat (16,7%) tidak akurat. Tingkat akurasi tertinggi ditemukan pada konstruksi tawabi' dan idhafah (80%), diikuti jumlah fi'liyyah dan jumlah syarthiyyah (75%), sedangkan akurasi terendah terdapat pada jumlah ismiyyah (37,5%). Kesalahan yang paling sering ditemukan berkaitan dengan identifikasi jumlah fi'liyyah, khabar muqaddam, syibh al-jumlah, hubungan mudhaf–mudhaf ilaih, dan fungsi isim syarth. Temuan ini menunjukkan bahwa Nahwu AI cukup efektif dalam mengenali struktur sintaksis dasar, tetapi masih memiliki keterbatasan pada konstruksi gramatikal yang kompleks. Penelitian ini memberikan kontribusi berupa kerangka evaluasi normatif yang menghubungkan analisis sintaksis berbasis AI dengan prinsip-prinsip nahwu klasik sebagai landasan pengembangan teknologi pembelajaran bahasa Arab yang lebih andal.

Kata Kunci: Kecerdasan Buatan, Nahwu AI, Sintaksis Bahasa Arab, Alfiyyah Ibn Malik, Nahwu Klasik.

Corresponding Author:
Walidatur Rohmah
Universitas Sunan Drajat Lamongan, Indonesia; walidaa064@gmail.com

Introduction

The teaching of Arabic in Indonesia, particularly the study of Nahwu, which governs sentence syntax (tarkib), continues to face structural challenges due to the complexity of grammatical rules, dialectal variations, and conventional teaching approaches that lack interactivity (Sofa et al., 2021). Advances in artificial intelligence (AI) have led to innovations such as the Nahwu AI Platform from Alma Ata University in Yogyakarta, which uses natural language processing (NLP) and grammatical algorithms to automatically analyze Arabic sentence structure and provide personalized feedback (Rohma & Machfudi, 2026; Satrio, 2025). However, a crucial challenge arises from the accuracy of AI output in mapping complex syntactic relationships such as dynamic i'rab, semantic dependencies, and classical nahwu rules, which often deviate from traditional normative logic (Hadi & Qohar, 2024). This phenomenon raises a fundamental epistemological question: to what extent can probabilistic AI systems represent the deterministic deductions of nahwu, particularly in the context of learning within Muslim academic circles that uphold the authority of historical grammar?

In the tradition of Arabic-Islamic linguistics, nahwu occupies a central position as the foundation for understanding the texts of the Qur'an and the Hadith (Sofa & Erviana, 2025), with Alfiyyah Ibn Malik as one of the key references that systematically summarizes various rules of Arabic syntax (Khairunnisa et al., 2024). This book has been widely used in the study of Arabic grammar and has become one of the primary references for syntactic functions, grammatical relations, and the

i'rab of various Arabic constructions (Mudhofar & Badi'ah, 2019). In addition, Alfiyyah Ibn Malik is known for successfully systematizing the rules of nahwu that had developed within the tradition of Arabic grammar into a concise, structured format that has been widely used as a pedagogical standard across generations (Khotimah, 2025). Its normative and systematic nature makes the Alfiyyah a relevant framework for evaluating the accuracy of the identification of grammatical functions and syntactic relations generated by artificial intelligence-based language analysis systems (Nulhakim & Ilmi, 2024). Therefore, this research uses Alfiyyah Ibn Malik and its commentary as a framework of normative reference (ground-truth) to evaluate the level of accuracy of the syntactic analysis output generated by the Nahwu AI Platform. The selection of this framework is based on its strong standing in the classical nahwu tradition and the systematic and normative nature of its rules, making it relevant as a reference for evaluating the accuracy of the identification of grammatical functions and syntactic relations generated by the AI system (Mubin et al., 2024; Toukhi & Al-rajhi, 2023). This research is not intended to compare Nahwu AI with human annotators or contemporary Arabic linguistic approaches, but rather to evaluate its degree of conformity with a specific normative framework represented by Alfiyyah Ibn Malik and its commentaries. Thus, this research uses the Alfiyyah Ibn Malik as a normative framework for evaluating the extent to which the results of analyses produced by data-driven AI systems remain consistent with the principles of syntax formulated in the classical nahwu tradition (Baroroh & Rahmawati, 2020; Cahyati et al., 2025; Elitahari & Naifah, 2026).

.Previous studies have explored the application of AI in Arabic grammatical analysis, but most have focused primarily on the technical evaluation of the systems. Muttaqin Yasis (2020) developed a rule-based NLP system for predicting shakl and word structure with a relevance of 70.21%, but without normative grammatical validation (Yasis, 2020). Rishanda et al. (2025) highlight the potential of AI for self-directed Nahwu learning, but overlook the affective aspects and normative rules (Rishanda et al., 2025). Rahman et al. (2024) qualitatively confirmed the relevance of classical syntax through Nahwu Wadhih, but have not yet integrated it with computational analysis (Rahman et al., 2024). Thus, there remains a research gap in the form of the lack of a systematic evaluation of AI syntactic analysis results based on classical nahwu rules as a normative standard. Most previous studies have relied on technical performance metrics such as accuracy, precision, recall, and F1-score, while the alignment of AI output with classical grammatical authorities has rarely been examined. This research aims to fill that gap by using the Alfiyyah Ibn Malik as a normative evaluation framework.

This research offers an evaluative approach that uses Alfiyyah Ibn Malik and Syarah Ibn Aqil as a reference framework (ground-truth) to assess the accuracy of the syntactic analysis (tarkib) output generated by the Nahwu AI Platform. The

novelty of this research lies in its effort to bridge modern computational approaches with classical nahwu rules through a systematic evaluation of the accuracy of syntactic functions and tarkib relations based on Ibn Malik's perspective. The purpose of this research is to evaluate the level of normative accuracy of the syntactic analysis output from the Nahwu AI system for several Arabic language structures, including nominal sentences, verbal sentences, tawabi' and idhafah constructions, and conditional sentences. In addition to measuring accuracy, this research also aims to identify the strengths and limitations of the system in analyzing syntactic functions and tarkib relations from the perspective of classical Nahwu. Theoretically, this research contributes to the development of a dialogue between the classical Nahwu tradition and computational linguistics, while practically, it provides insights for the utilization and development of AI-based Nahwu systems in Arabic language learning.

Method

This research is a qualitative evaluative research (Patton, 2015) using a critical-comparative analytical approach, aimed at assessing the normative conformity of the syntactic analysis (tarkib) output from the Nahwu AI Platform with the classical rules of nahwu in Alfiyyah Ibn Malik (Creswell et al., 2007). This approach was chosen because it allows for an in-depth evaluation of the epistemological representations of traditional nahwu within a probabilistic computational system, without relying solely on quantitative statistical metrics (Ash-Shiddieqy & Hermina, 2025). Its epistemological framework is rooted in the Arab-Islamic tradition of prescriptive grammar, which acknowledges the fundamental difference between the deterministic deduction of the Alfiyyah and the modern natural language processing (NLP) approach, thereby emphasizing normative validation over universal technical performance (Rofiqoh, 2022).

The data consisted of 30 authentic Arabic sentences as units of analysis, selected through purposive sampling based on the following explicit criteria: (1) representation of the main syntactic structures that are often problematic in the study of nahwu (nominal sentence, verbal sentence, sentences with tawabi' and idhafah, conditional sentences) (Sulaikho et al., 2025), (2) variations in the level of i'rab complexity (standard, ambiguity without harakat, complex khabar) (Salafuddin & Suharto, 2022), and (3) relevance to the rules of nahwu in Alfiyyah Ibn Malik, which include syntactic functions, syntactic relations, and i'rab categories the focus of this research evaluation. The corpus size was intentionally limited because the study prioritizes depth of normative analysis rather than corpus-scale evaluation.

Each sentence was subjected to detailed grammatical examination against the rules of *Alfiyyah Ibn Malik* and its commentaries, making a smaller but representative

dataset more appropriate for the evaluative objectives of the study. The data sources consisted of pedagogical sentence examples from classical grammar texts (Alfiyyah Ibn Malik and its commentary) as well as controlled authentic variations for learning scenario simulations, limiting the scope to ensure control and focus on representative cases. The Alfiyyah Ibn Malik, along with authoritative commentaries such as the Syarh of Ibn Aqil, serves as a normative benchmark (ground-truth), with the following operational rules: syntactic functions (mubtada', khabar, fi'il, fa'il, maf'ul bih), syntactic relations (dependency hierarchy), and i'rab cases (rafa', nashab, jarr, jazm).

Data collection was conducted systematically through digital documentation: each sentence was entered into the Nahwu AI Platform to generate automated syntactic analysis (i'rab parsing, tarkib structure) (Muharom, 2025), which was then documented in a comparative table format (input sentence, AI output, identified syntactic elements). This procedure ensures reproducibility, with the AI output re-verified for platform consistency.

The data analysis employs comparative content analysis (Sofwatillah et al., 2024; Wekke, 2019), and is divided into the following sequential stages(1) Normative codification: Relevant rules are extracted from Ibn Malik's Alfiyyah and its commentary for each sentence analyzed, (2) Qualitative comparison: AI output is compared with codified rules using a comparative table to identify the degree of conformity (accurate: complete identification), partial deviation (fairly accurate: the main structure is correct but the i'rab details are incorrect), or nonconformity (inaccurate: violation of syntactic relations), (3) Handling ambiguity: If a rule is found that allows for more than one interpretation, triangulation is performed using classical commentaries as the basis for clarification, with unresolved cases marked as separate findings, (4) Critical synthesis: Identification of patterns of strength (e.g., detection of simple structures) and limitations (e.g., semantic ambiguity), presented in a descriptive-analytical manner with grammatical reasoning. The validity of the findings is reinforced through triangulation of sources by comparing the rules in Alfiyyah Ibn Malik and several relevant authoritative commentaries, as well as through an audit trail documenting the entire analysis process. In cases where more than one interpretation is possible, the determination of the result is based on the strongest argumentative justification according to the classical references used.

Although this research employs an evaluative qualitative approach, the comparison results are also presented as descriptive percentages to provide an overview of trends in the system's accuracy. These percentages are not used for hypothesis testing or statistical inference, but rather as a descriptive tool for mapping the distribution of evaluation results. The accuracy rate is calculated using the formula:

$$\text{Accuracy} = \left(\frac{\sum \text{Accurately Classified Sentences}}{\sum \text{Total Sentence Samples}}\right) \times 100\%$$

Before the analysis was conducted, the researchers established three evaluation categories based on the alignment of the Nahwu AI's output with the rules of Ibn Malik, namely: (1) Accurate, if the system successfully identifies the syntactic structure, grammatical function, and i'rab case accurately, (2) Fairly accurate, if the system is able to recognize the main structure of a sentence but there are still inaccuracies in some grammatical functions or syntactic relationships, and (3) Inaccurate, if the system fails to identify the syntactic structure correctly or produces an analysis that contradicts the reference rules. The results of this categorization are then summarized in the form of percentage distributions to illustrate the performance trends of Nahwu AI for each syntactic structure analyzed.

The scope of this research is limited to an exclusive focus on the normative appropriateness of Ibn Malik's perspective, rather than the universal validation of AI systems or probabilistic metrics. The results cannot be generalized to modern dialects or non-classical corpora; rather, they provide critical insights for the pedagogical optimization of digital nahwu. This approach aligns with evaluating tarkib accuracy, ensuring a significant epistemological contribution to Arabic language education.

Results And Discussion

Results

This research analyzes 30 Arabic sentences processed through the Nahwu AI Platform. The corpus was compiled using purposive sampling to ensure representation of various syntactic structures (tarkib) relevant to classical nahwu rules. The distribution of sentence types includes: (1) verbal sentences (8 sentences), (2) nominal sentences (8 sentences), (3) sentences with tawabi' and idhafah (10 sentences), and (4) conditional sentences (4 sentences). This selection was based on the syntactic complexities frequently encountered in the study of nahwu, particularly regarding the identification of syntactic functions (mubtada', khabar, fi'l, fa'il, maf'ul bih) and i'rab relationships (rafa', nashab, jarr, jazm).

The examples presented in this results section are representative, not selective, and were chosen to illustrate the general patterns within each category of findings. Each example is accompanied by a quotation from the Alfiyyah Ibn Malik as the normative standard for evaluation, as well as a systematic explanation of the grammatical reasoning process.

Table 1.1 Distribution of AI Nahwu Analysis Results Based on Syntactic Structure (N=30)

Structure	Total	Accurate	Fairly Accurate	Inaccurate	Accurate %
verbal sentences	8	6	1	1	75%
nominal sentences	8	3	3	2	37,5%
sentences with tawabi' and idhafah	10	8	1	1	80%
conditional sentences	4	3	0	1	75%
Total	30	20	5	5	66,7%

The table above shows that of the 30 sentences analyzed, 20 sentences (66.7%) were classified accurately, meaning that Nahwu AI's output fully conforms to the rules of Ibn Malik's in identifying syntactic functions and i'rab relations. A total of 5 sentences (16.7%) fell into the "fairly accurate" category, where the system was able to identify the main structure of the sentence but still exhibited inaccuracies in certain details of grammatical classification. Meanwhile, 5 sentences (16.7%) were classified as "inaccurate," indicating a fundamental discrepancy between the AI's analysis results and the Nahwu rules used as the evaluation framework.

Further analysis of the data categorized as "fairly accurate" and "inaccurate" reveals that Nahwu AI's inaccuracies follow several recurring patterns. The most prevalent errors relate to the identification of the khabar function in nominal sentence. Recurring error patterns include failure to recognize khabar in the form of verbal sentence, muqaddam khabar, and khabar realized in the form of syibh al-jumlah (jar–majrur). In addition, two cases were found in the verbal sentence structure: one case where the system was able to identify the fa'il element but was less accurate in interpreting the overall sentence structure, and another case where the fa'il element was actually identified as the maf'ul bih. In idhafah constructions, one instance of an error in determining the mudhaf–mudhaf ilaih relationship was found, while in syarthiyyah structures, one instance of the system's inability to recognize the function of the isim syarth was found, resulting in the element man (من) being interpreted as an isim maushul and the relationship between the fi'il syarth and the jawab syarth not being accurately identified. These findings indicate that Nahwu AI tends to be more successful at recognizing basic syntactic functions, but still faces challenges when dealing with structures involving multi-level grammatical relations, variations in the form of the predicate, and syntactic constructions that require more complex contextual analysis.

Table 1.2 Representative Examples by Category

No	Sentence	AI Analysis	The Rules of the Alfiyyah	Category	Reason
1.	فهم الطالب الدرس	فعل الماضي + فاعل + مفعول به	Verse 225: وَبَعْدَ فِعْلٍ فَاعِلٌ فَإِنْ ظَهَرْ ... فَهْوَ وَإِلَّا فَضَمِيرٌ اسْتَتَرْ Verse 237: وَالْأَصْلُ فِي الْفَاعِلِ أَنْ يَتَّصِلَا ... وَالْأَصْلُ فِي الْمَفْعُولِ أَنْ يَنْفَصِلَا	Accurate	Accurate identification
2.	الطالبة مؤدبة	مبتدأ + خبر	Verse 113: مُبْتَدَأٌ زَيْدٌ وَعَاذِرٌ خَبَرْ ... إِنْ قُلْتَ زَيْدٌ عَاذِرٌ مَنِ اعْتَذَرْ Verse 114: وَأَوَّلٌ مُبْتَدَأٌ وَالثَّانِي ... فَاعِلٌ اغْنَى فِي أَسَارٍ ذَانِ	Accurate	In accordance with the rules
3.	بسم الله الرحمن الرحيم	حرف الجر + اسم + اسم المجرور + الصفة الأولى (نعت) + الصفة الثانية (نعت)	Verse385: وَالثَّانِيَ اجْرُرْ وَانْوِ مِنْ أَوْ فِي إِذَا ... لَمْ يَصْلُحِ الَّا ذَاكَ وَاللَّامَ خُذَا Verse 507: فَالنَّعْتُ تَابِعٌ مُتِمٌّ مَا سَبَقْ ... بِوَسْمِهِ أَوْ وَسْمِ مَا بِهِ اعْتَلَقْ	Fairly Accurate	The AI successfully identified the na'at–man'ut relationship, but was incorrect in identifying the word الله as the mudhaf ilaih
4.	زيد قام ابوه	مبتدأ + فعل الماضي + فاعل + الضمير	Verse 119: وَمُفْرَدًا يَأْتِي وَيَأْتِي جُمْلَهْ ... حَاوِيَةً مَعْنَى الَّذِي سِيقَتْ لَهْ	Fairly Accurate	The AI correctly identifies the mubtada', but cannot accurately identify the khabar in the form of a verbal sentences.
5	من جدّ وجد	اسم الموصول + فعل الماضي + فعل الماضي	Verse 696: وَاجْزِمْ بِـإِنْ وَمَنْ وَمَا وَمَهْمَا ... أَيٍّ مَتَى أَيَّانَ أَيْنَ إِذْمَا Verse 698: فِعْلَيْنِ يَقْتَضِينَ شَرْطٌ قُدِّمَا ... يَتْلُو الْجَزَاءُ وَجَوَابًا وُسِمَا	Inaccurate	The AI interprets "من" as a relative noun, and thus fails to identify the conditional clause structure consisting of a conditional verb and a conditional response.

In the phrase بسم الله الرحمن الرحيم, the system successfully recognized the na'at–man'ut relationship between the terms الرحمن and الرحيم, but failed to identify the term الله as the mudhaf ilaih. Based on the rules of idhafah in Alfiyyah Ibn Malik, the word اسم functions as the mudhaf and the word الله functions as the mudhaf ilaih, which must be in the majrur case. This inaccuracy in recognizing this relationship resulted in the analysis being categorized as moderately accurate because the main structure of the sentence could still be identified.

In the sentence زيد قام أبوه, the system successfully identified زيد as the mubtada'. However, the AI did not recognize that قام أبوه forms a verbal sentence that functions as the khabar. This finding indicates that the system is capable of recognizing basic syntactic elements but encounters difficulties when the khabar is realized in the form of a verbal clause, as explained in the verse of the Alfiyyah regarding the khabar that can take the form of a phrase.

Unlike the two previous cases, the phrase من جدّ وجد reveals a more fundamental inconsistency. The system identifies "من" as a relative noun (isim maushul), whereas according to the rules of conditional clauses in Alfiyyah Ibn Malik, this phrase functions as a definite conditional noun (isim syarth jazim) that requires a conditional verb (fi'il syarth) and a conditional response (jawab syarth). As a result, the conditional relationship between "جدّ" and "وجد" is not identified, so the analysis is classified as inaccurate.

Based on a comprehensive analysis of 30 sentences, several patterns of findings can be identified. The strength of Nahwu AI lies in its ability to recognize basic syntactic structures, particularly in tawabi' and idhafah constructions, which achieved the highest accuracy rate (80%). In addition, the system also demonstrated relatively good performance in analyzing simple verbal sentence, with an accuracy rate of 75%. In most of the data categorized as accurate, the system was able to identify key syntactic functions such as fi'il, fa'il, maf'ul bih, mubtada', and khobar in accordance with the Nahwu rules used as the evaluation framework.

On the other hand, several limitations were also identified. The lowest accuracy rate was found in the nominal sentence structure (37.5%), indicating that the system still has difficulty identifying more complex variations in the function of the predicate. These findings are consistent with the discovery of recurring errors in predicates in the form of verbal sentences, muqaddam predicates, and syibh al-jumlah (jar majrur). In addition, errors were also found in determining the mudhaf-mudhaf ilaih relationship in idhafah constructions and in identifying the function of the isim syarat in syarthiyyah structures. These findings indicate that the system's performance tends to decline when dealing with syntactic structures that involve multi-level grammatical relations and require more complex syntactic reasoning compared to more explicit structures.

The analysis results were interpreted systematically, with direct reference to the rules of Ibn Malik's as the normative standard. In cases of ambiguity, triangulation of authoritative commentaries (particularly Syarh Ibn Aqil) was used to ensure consistency in grammatical reasoning. This process confirms that the evaluation was based on the application of the Alfiyyah's framework of rules, rather than on subjective judgment.

Overall, a normative evaluation of the Nahwu AI Platform based on the Alfiyyah Ibn Malik framework shows that the system achieved an accuracy rate of 66.7%, with 20 of the 30 analyzed sentences classified as accurate. A total of 5 sentences (16.7%) were classified as fairly accurate, and 5 sentences (16.7%) were classified as inaccurate. The best performance was demonstrated in the analysis of tawabi' and idhafah constructions, with an accuracy rate of 80%, while the greatest challenge was found in the analysis of nominal sentence, which achieved an accuracy rate of only 37.5%. These results indicate that Nahwu AI has a fairly good ability to recognize basic syntactic structures, but still requires further development to handle variations in the function of the khabar, idhafah relations, and more complex syarthiyyah constructions.

It should be emphasized that the results represent a normative evaluation based on Ibn Malik's perspective, not a universal validation of the overall performance of AI systems. Claims of generalizability are limited to the scope of the corpus and the evaluation framework used in this research.

Discussion

This research is a normative evaluation that uses Ibn Malik's perspective as its normative framework of reference. Therefore, the findings are not intended as a universal validation of the performance of Nahwu AI, but rather as an assessment of the degree to which the system's syntactic analysis output aligns with classical Nahwu rules used as the evaluation standard. Consequently, the interpretation of the research results is limited to the normative framework employed and cannot be directly generalized to other approaches in Arabic linguistics.

Based on a synthesis of the research findings, several significant patterns can be identified regarding the performance of the Nahwu AI system in analyzing Arabic syntactic structures. The system's main strength lies in its ability to analyze tawabi' and idhafah constructions with an accuracy rate of up to 80%, indicating that the platform is relatively reliable in processing attributive and possessive relationships in Arabic. In addition, the system also demonstrates adequate performance in analyzing simple verbal constructions with an accuracy rate of 75%. These findings suggest that Nahwu AI has relatively good capabilities in recognizing basic syntactic functions such as fi'il, fa'il, and maf'ul bih when sentence

structures are presented explicitly and do not involve complex grammatical relationships.

On the other hand, the system's main limitation was identified in the analysis of nominal sentences, which achieved an accuracy rate of only 37.5%. The low accuracy in this category indicates that the system still struggles to identify more complex variations in the function of the khabar, particularly the khabar in the form of a verbal clause, a muqaddam khabar, and a syibh al-jumlah. These findings suggest that Nahwu AI finds it easier to recognize linear and explicit syntactic structures than those involving non-linear grammatical relationships.

From the perspective of classical grammar, this finding can be understood through the flexibility of sentence structure recognized in Ibn Malik's *Alfiyyah*. In the chapter on the Subject (Mubtada') and Predicate (Khabar), it is stated (Saifuddin, 2016)

وَالْأَصْلُ فِي الْأَخْبَارِ أَنْ تُؤَخَّرَا ۞ وَجَوَّزُوا التَّقْدِيمَ إِذْ لَا ضَرَرَا

This verse indicates that the khabar is fundamentally placed after the mubtada', but it may be placed before it as long as it does not cause ambiguity in meaning. Such flexibility requires the ability to recognize syntactic relationships that do not always directly follow word order. Ibn Aqil explains that in some constructions, the grammatical relationship between sentence components must be understood through their semantic and syntactic functions, rather than solely based on the position of the elements within the sentence (ibn aqil, 1980). This condition appears to remain a challenge for AI systems that operate based on data patterns and probabilities.

In addition, errors were also found in the identification of the mudhaf–mudhaf ilaih relationship in idhafah constructions and in the recognition of the function of the isim syarth in syarthiyyah structures. These findings indicate that the performance of Nahwu AI tends to decline when dealing with structures that require multi-level syntactic reasoning and more complex grammatical relationships.

The research findings indicate a difference in epistemological paradigms between classical nahwu and AI systems that operate probabilistically based on data. Therefore, discrepancies between the output of AI-based nahwu and the rules of the Alfiyyah should not always be interpreted as functional errors, but rather as a consequence of the differing analytical frameworks employed. In this context, the study seeks to foster a dialogue between the classical nahwu tradition and computational linguistics, rather than positioning one as a superior approach over the other.

The Alfiyyah Ibn Malik was chosen as the evaluative framework due to its authoritative status in the tradition of nahwu and its structured system of rules. However, this framework represents one of the normative traditions in Arabic

grammar; therefore, the results of this research are not intended to establish the Alfiyyah as an absolute standard, but rather as a specific reference for assessing the level of conformity of AI-generated nahwu output. The selection of Alfiyyah Ibn Malik does not imply that it represents the only valid framework for Arabic grammatical analysis; rather, it serves as a historically influential and systematically codified reference within the classical nahwu tradition.

In practice, Nahwu AI can be used as a learning aid for relatively simple structures, such as tawabi', idhafah, and verbal sentence. However, for structures involving complex syntactic relationships, particularly variations in the khabar within nominal sentence and certain syarthiyyah constructions, the analysis results still require verification and teacher guidance. These findings indicate that Nahwu AI is better positioned as a learning aid rather than a replacement for independent nahwu analysis.

Like other qualitative studies, this research has several limitations that must be transparently acknowledged. The limited corpus size (only 30 sentences) does not allow for statistical generalizations regarding the system's overall performance, given that the variety of syntactic structures in Arabic is far broader than the sample analyzed. Furthermore, the evaluative nature of this qualitative research makes the findings more exploratory and descriptive in nature; therefore, they are not intended to yield inferential generalizations (Labuschagne, 2003).

Based on the findings, several recommendations can be proposed for the future development of the system. First, a specialized module is needed to improve the accuracy of analyzing sentence variations, particularly sentences in the form of verbal sentence, muqaddam clauses, and syibh al-jumlah. Second, the syntactic analysis mechanism for conditional structures needs to be refined, including the identification of the conditional noun (isim syarth) and the relationship between the conditional verb (fi'il syarth) and the conditional response (jawab syarth). Third, validation using a more comprehensive corpus is needed to strengthen the empirical findings before broader generalizations can be made. These suggestions are proportionate and based on empirical evidence from this research, rather than being grand architectural claims without a foundation.

Conceptually, this research demonstrates that the evaluation of artificial intelligence-based language technologies need not be positioned in opposition to the classical nahwu tradition. Rather, the findings of this research open up a space for dialogue between modern computational systems and the intellectual legacy of Arabic nahwu through the use of the rules of Ibn Malik's Alfiyyah as a normative evaluative framework. Thus, the main contribution of this research lies not in establishing the superiority of one approach over another, but in the effort to bring together the tradition of classical grammar and artificial intelligence technology in

the development of a more adaptive and science-based approach to Arabic language learning.

Conclusion

Based on the results of an evaluation of the syntactic analysis output from the Nahwu AI Platform using Ibn Malik's Alfiyyah as a normative framework, it can be concluded that: (1) Nahwu AI demonstrates a fairly good level of accuracy in analyzing several Arabic syntactic structures, with an overall accuracy rate of 66.7%. The highest accuracy was found in tawabi' and idhafah constructions as well as simple verbal sentences, indicating the system's ability to recognize basic syntactic functions and relatively explicit grammatical relationships. (2) The system's main limitation lies in the analysis of structures involving more complex syntactic relations, particularly nominal with variations in predicate realization, certain mudhaf–mudhaf ilaih relationships, and conditional constructions. These findings indicate that the system's performance tends to decline when dealing with structures that require more in-depth grammatical reasoning, and (3) from the perspective of classical grammar, Nahwu AI is best positioned as a learning tool that can support the understanding of basic syntactic functions, but still requires verification and teacher guidance for more complex structures (Labuschagne, 2003).

This research contributes to the development of a dialogue between the classical nahwu tradition and computational linguistics by using Ibn Malik's Alfiyyah as a normative framework for evaluating artificial intelligence-based syntactic analysis systems. The research findings also provide an empirical overview of the strengths and limitations of AI-based nahwu, which can serve as input for the development of AI-based Arabic language learning technologies.

This research has limitations in terms of its relatively small corpus and the scope of its evaluation, which focuses solely on Ibn Malik's normative perspective. Therefore, future research could utilize a broader and more diverse corpus and combine the normative evaluation framework with other linguistic approaches to gain a more comprehensive understanding of the performance of artificial intelligence-based Arabic syntactic analysis systems.

Reference

Ash-Shiddieqy, M. I., & Hermina, D. (2025). Desain Penelitian Evaluatif. *Psikosospen : Jurnal Psikososial Dan Pendidikan*, *1*(2), 492–500.

Baroroh, R. U., & Rahmawati, F. N. (2020). Metode-Metode Dalam Pembelajaran Keterampilan Bahasa Arab Reseptif. *Urwatul Wutsqo: Jurnal Studi Kependidikan Dan Keislaman*, *9*(2), Article 2. https://doi.org/10.54437/urwatulwutsqo.v9i2.181

Cahyati, L., Abdurahman, M., & Saleh, N. (2025). Analysis of Learning Obstacles in the Damair Rules Material in Arabic Language Learning. *Urwatul Wutsqo:*

Jurnal Studi Kependidikan Dan Keislaman, *14*(2), 417–434. https://doi.org/10.54437/urwatulwutsqo.v14i2.2161

Creswell, J. W., Hanson, W. E., Clark Plano, V. L., & Morales, A. (2007). Desain Penelitian Kualitatif: Seleksi dan Pelaksanaan. *Psikolog Konseling*, *35*(2), 236–264. https://doi.org/10.1177/0011000006287390

Elitahari, J., & Naifah, N. (2026). Linguistic Feasibility Analysis of Grade V Arabic Textbook by Nur Hidayah Press BSKAP Perspective. *Urwatul Wutsqo: Jurnal Studi Kependidikan Dan Keislaman*, *15*(01), 254–271. https://doi.org/10.54437/urwatulwutsqo.v15i01.2890

Hadi, M., & Qohar, H. A. (2024). Peran Artificial Intelligence dalam Meningkatkan Pembelajaran Interaktif Bahasa Arab. *Ranah Reserch*, *6*(6), 3010–3023.

ibn aqil, abdullah bin abdurrahman al-aqily. (1980). *Syarh ibn aqil ala alfiyah ibn malik* (ke-20). Dar At-Turast.

Khairunnisa, A., Pribadi, M., & Sugiyono, S. (2024). Dasar Pemikiran Ibnu Malik Serta Representasi Pemikiran Basrah Dalam Kitab Alfiyahnya. *Al Mi'yar*, *7*(1), 481–500.

Khotimah, K. (2025). Peran dan kontribusi ibn malik dalam pengembangan tata bahasa arab klasik. *Maliki Interdisciplinary Journal (MIJ)*, *3*, 895–901.

Labuschagne, A. (2003). *Qualitative Research – Airy Fairy or Fundamental? Qualitative Research – Airy Fairy or Fundamental ? 8*(1), 100–103.

Mubin, K., Purnama, B. B., & Folliero, D. (2024). Linguistic Harmonization: Dissecting Alfiyah Ibn Malik with Chomsky ' s Transformational Generative Theory. *Alibbaa':Jurnal Pendidikan Bahasa Arab*, *5*(2).

Mudhofar, & Badi'ah, D. (2019). Nilai-nilai Pendidikan Karakter Religius dalam Nadhm Alfiyyah Ibn Mālik fī al—Nahw wa al-Sharf. *Jurnal Penelitian Ilmiah Intaj*, *03*(01), 1–33.

Muharom, A. (2025). Inovasi Pembelajaran Bahasa Arab Berbasis Artificial Intelligence: Analisis Platform NahwuAI dan Implikasinya Terhadap Tradisi Nahwu-Sharaf Pesantren. *Proceeding of International Conference on Islamic Education (ICIED) BT - Impactful Education: Integrating Technology and Ecotheological Values for a Sustainable Future*, *10*(1), 906–914. https://doi.org/10.18860/icied.v10i1.3696

Nulhakim, L., & Ilmi, V. W. (2024). Naqd Abī Ḥayyān al-Naḥwī li Ibn Mālik: Dirāsah Muqāranah bayna al-Fikr al-Naḥwī li Abī Ḥayyān wa al-Fikr al-Naḥwī li Ibn Mālik. *Al-Zahra: Journal for Islamic and Arabic Studies*, *21*(2). https://doi.org/10.15408/zr.v21i2.40979

Patton, M. Q. (2015). *Qualitative Research & Evaluation Methods: Integrating Theory and Practice* (4th Edition). SAGE Publications, Inc.

Rahman, A., Haniefa, R., & Kafidhoh, S. (2024). Analisis Materi Sintaksis Dalam Kitab Nahwu Wadhih Serta Relevansinya Terhadap Pembelajaran Bahasa Arab. *Jurnal Ilmiah Mahasiswa Pendidikan Bahasa Arab Sekolah Tinggi Agama Islam Nurul Iman*, *2*(1), 35–50. https://doi.org/https://jim.stai-nuruliman.ac.id/index.php/jimpba

Rishanda, A., Koderi, K., … F. G.-A. M. J. I., & 2025, undefined. (2025). Artificial Intelligence Dalam Pembelajaran Nahwu Secara MandiriNANo ranking

found for 'Al Mi'yar: Jurnal Ilmiah Pembelajaran Bahasa Arab dan Kebahasaaraban'. *Jurnal.Stiq-Amuntai.Ac.Id*, *8*(1), 27.

Rofiqoh, E. (2022). خصائص اللغة العربية. *Lisan An Nathiq: Jurnal Bahasa Dan Pendidikan Bahasa Arab*, *4*(1), 34–46.

Rohma, W. H., & Machfudi, I. (2026). Optimizing Artificial Intelligence (AI) as a Learning Media for Islamic Religious Education in the Society Era 5. 0. *Urwatul Wutsqo: Jurnal Studi Kependidikan Dan Keislaman*, *15*(1), 119–134.

Saifuddin, M. (2016). *Kajian dan Analisis Alfiyah*.

Salafuddin, & Suharto, T. (2022). Tathbiqu Kitabi "Taqribu An Nahwi Wa As Shorfi" Fi Ta'limi An Nahwi Wa As Shorfi Bi Ma'hadi Barakah Al 'Ulum Kajen –. *Rayah Al-Islam*, *6*(2), 1–10.

Satrio, S. (2025). Integrasi Artificial Intelligence dalam Pembelajaran Bahasa Arab: Peluang, Tantangan, dan Inovasi Pedagogis di Era Digital. *RIGGS: Journal of Artificial Intelligence and Digital Business*, *4*(2), 5907–5914.

Sofa, A. R., Aziz, A., & Ichsan, M. (2021). Jurnal Inovasi Penelitian. *Jurna Inovasi Penelitian*, *1*(9), 1761–1774.

Sofa, A. R., & Erviana, I. (2025). *Program Pengabdian Kemasyarakatan: Optimalisasi Pembelajaran Nahwu melalui Kitab Al Miftah di Pesantren Motivator Qur ' an Darussalam Klaseman*. *3*.

Sofwatillah, Risnita, Jailani, M. S., & Saksitha, D. A. (2024). Tehnik Analisis Data Kuantitatif Dan Kualitatif Dalam Penelitian Ilmiah. *Journal Genta Mulia*, *15*(2), 79–91.

Sulaikho, S., Hamid, M. A., & Machmudah, U. (2025). *Interpretation of I ' rāb Rules in Naẓam Alfiyah Based on Jumlah Ismiyah and Jumlah Fi ' liyah Interpretasi Kaidah I ' rāb dalam Naẓam Alfiyah Berdasarkan Jumlah Ismiyah dan Jumlah Fi ' liyah*. *09*(01), 212–224. https://doi.org/10.30762/asalibuna.v9i01.5338

Toukhi, S. S., & Al-rajhi, H. A. (2023). Ibn Malik's Approach in Presenting Syntactic and Morphological Judgements; A study in Alfiyah Received: *Journal of Arabic Language and Literature (JALSL)*, *2*(3), 9–20.

Wekke, I. S. (2019). *Study Naskah Bahasa Arab: Teori,Konstruksi,Praktek*.

Yasis, M. M. (2020). *Implementasi Natural Language Processing dalam Penggunaan Ilmu Nahwu pada Bahasa Arab*. Universitas Komputer Indonesia.