



Arabiyat : Jurnal Pendidikan Bahasa Arab dan Kebahasaaraban

Vol. 11 No. 2, December 2024, 160-174

P-ISSN: 2356-153X; E-ISSN: 2442-9473

doi: <http://dx.doi.org/10.15408/a.v11i2.42671>



Exploring AI Capabilities in Arabic Grammar: Comparative Analysis of ChatGPT and Gemini

Amanda Resti Maulidiya, Maman Abdurrahman, Nalahuddin Saleh

Universitas Pendidikan Indonesia, Indonesia

Corresponding E-mail: amandarestii@upi.edu

Abstract

This study aims to compare the performance of ChatGPT and Gemini in analyzing *i'rāb Marfū'āt al-Asmā'*, a critical aspect of Arabic grammar for determining the grammatical roles of words in sentences. The analysis uses 11 examples from the book *Mulakhas Qawā'id al-Lughah al-'Arabiyyah* by Fuad Ni'mah, focusing on identifying grammatical components such as *muftada'*, *khabar*, *isim asbbah*, *isim kāda*, *khabar inna*, *fā'il*, *nā'ib al-fā'il*, *na'at*, *'atf*, *tawkid*, and *badal*. The study employs the Mann-Whitney test to assess statistical significance and the Cosine Similarity Index (CSI) to measure semantic similarity. Results show that ChatGPT outperforms Gemini with a significant value of 0.019, while the CSI score of 0.800 indicates high similarity between the models' outputs. ChatGPT excels in providing detailed and accurate analyses, while Gemini is more suitable for concise answers but may lack precision. These findings highlight the unique strengths of each model and underscore the necessity of manual correction to ensure the accuracy and relevance of results, particularly in technology-based Arabic grammar learning.

Keywords: *ChatGPT, Gemini, Generative AI, I'rāb analyzing*

Introduction

Arabic has its challenges in linguistic analysis, especially sentiment analysis, due to its complex linguistic nature. With morphology involving a lot of inflection and derivation, Arabic can disguise the polarity of the underlying sentiment (Aladeemy et al., n.d., 2024). In a linguistic context, syntax, known in Arabic as the Science of *Nahwu* or *i'rāb*, is the process of determining the role of each word in a sentence. This is done by adding diacritic marks at the end of the word along with the right reason, a technique known as final analysis. This aspect is important in Arabic because mistakes in *i'rāb* can lead to misunderstanding of the meaning of sentences, especially considering that Arabic is the language of the Qur'an, so a correct understanding is necessary (Mannaa

et al., 2022). The science of Nahwu itself is one of the sciences that is considered challenging and tiring to learn (Marsiah et al., 2021). Reading a good Arabic text requires the ability to determine *harakat* or *syakl*, which basically depends on understanding the context of the sentence. According to (Mualif, 2019), reading with a complete *harakat* is very important because it helps the reader recognize the position of the word in the sentence. However, giving *harakat* accurately, especially at the end of a word, is still a challenge for many Arabic learners. Difficulties in performing *i'râb* result in errors in identifying the type of sentence and the position of the word or phrase in a sentence (Maulida & Marsiah, 2023). In this context, advances in artificial intelligence technology offer potential solutions to the challenges faced in sentiment analysis and understanding of Arabic grammar.

The development of artificial intelligence (AI) technology, particularly in the field of education and linguistic analysis, has made rapid progress, creating advanced language models capable of handling complex natural language processing (NLP) tasks. Through the application of deep learning techniques, AI helps researchers decipher the unique properties of the Arabic language, especially in the *i'râb* aspect that involves morphological and syntactic analysis (Ali, 2016; Kishan & Reiter, 2021; Khalatia & Al-Romanyb, 2020). Tools like ChatGPT offer an innovative approach to grammar analysis that was previously considered complicated, providing personalized solutions in learning tailored to students' needs (Musthafa, 2024). The use of AI in learning creates a more interactive and responsive experience, increasing student engagement by providing instant feedback as well as varied sentence examples for a deeper understanding of grammar (Yatri et al., 2023). Research on AI models such as ChatGPT and Gemini also highlights their abilities and limitations in different contexts, providing insight into their suitability for specific tasks in understanding important aspects of the Arabic language, such as *Marfû'ât al-Asmâ'*, which is often a challenge in learning (Khidhir, 2024). Although these tools have been implemented, there are still gaps in accuracy and effectiveness in identifying complex structures (Berliani et al., 2024).

ChatGPT's capabilities, designed to provide deeper responses in natural conversational interactions, allow for complex analysis in technical topics such as *i'râb*. Its sensitivity to cultural and linguistic contexts makes it more suitable for academic environments requiring detailed explanations (Wijonarko et al., 2023; Ramli, 2023). Conversely, Gemini offers more concise responses and is often suitable for users who need quick and immediate understanding (Gomez-Cabello et al., 2024). This model may not be as flexible as ChatGPT in deep conversations, but it can be a great choice for those looking for practical solutions without needing too much detail (Suryono et al., 2023). Thus, the selection of the ideal model depends on the purpose of the analysis and the expected level of complexity, whether it is for more detailed needs or simply a basic understanding.

Previous studies have explored the role of AI in Arabic language learning, including grammar teaching and morphological analysis. Based on findings (Berliani et al., 2024), ChatGPT has great potential in helping students learn Arabic by providing quick responses and in-depth grammar analysis. However, its use still faces challenges, such as student dependency and lack of critical understanding. Research (Qadir, 2023)

states that ChatGPT can provide a more personalized learning experience through tailored feedback, while Khidhir (2024) shows that natural language processing (NLP) in Arabic requires greater attention due to the complexity of sentence structure and its rich morphology. While (Muawanah & Rihi, 2018) demonstrated the effectiveness of AI in analyzing Arabic *i'râb*, supporting the relevance of using AI for better understanding of Arabic. However, while these findings are relevant, most of the research focuses on applications in education and has not specifically compared AI models in the context of *Marfû'ât al-Asmâ'* analysis.

Some other studies, including (Aljanabi, 2024), evaluate the comparison of ChatGPT and Cloude, showing that Cloude is superior in addressing Arabic language morphology and parsing accuracy. Research (Elkhatat et al., 2023) and (Weber-Wulff et al., 2023) shows that AI models still face challenges in analyzing complex texts that require accuracy in the nuances of language. (Pressman et al., 2024) and (Rane et al., 2024) also compared ChatGPT with Gemini in medical and educational contexts, but the study showed gaps in specific Arabic linguistic analysis, including *i'râb* analysis. There have not been many studies that specifically explore the ability of ChatGPT and Gemini to overcome linguistic challenges *Marfû'ât al-Asmâ'*. This indicates the need for further in-depth research to strengthen the application of AI in complex grammatical analysis of the Arabic language.

This study aims to conduct a comparative study between ChatGPT and Gemini in analyzing *Marfû'ât al-Asmâ'*, a key element in Arabic grammar that functions to show the grammatical role of a word in sentence structure (Anwar & Ahyarudin, 2023). By evaluating the capabilities of these two AI platforms, this study is expected to provide significant insights into the effectiveness of AI technology in helping deeper grammatical understanding for Arabic language learners, especially in understanding and applying *Marfû'ât al-Asmâ'* (Nurmala et al., 2022). This analysis will refer to Fuad Ni'mah's book *Mulakhasb Qawâ'id al-Lughah al-'Arabiyah* as the main source of examples of *marfû'âtul asma'* cases, to provide more accurate guidance in testing the ability of the two platforms to identify complex grammatical structures. In addition, this research is also expected to make a significant contribution to the development of Arabic language learning, which focuses on teaching *nahwu* and *sorogan* as a classical method in Arabic grammatical education (Rahmat & Abdurrahman, 2017). Thus, the results of this research will provide an innovative foundation for AI-based Arabic language learning and offer solutions to more adaptive and technology-based learning challenges in the digital era.

Method

This study uses a quantitative method with the Mann-Whitney and Cosine Similarity Index (CSI) test approaches to compare *Marfû'ât al-Asmâ'* analysis ability between ChatGPT and Gemini. The study aims to evaluate the results of the answers of the two models in answering questions related to *Marfû'ât al-Asmâ'*, by referring to the book *Mulakhasb Qawâ'id al-Lughah al-'Arabiyah* as the main reference. The data analyzed included 11 examples designed based on the book to measure the ability of both models to analyze aspects of *i'râb marfû'âtul asma'*, namely *mubtada'*, *khobar*, *isim*

ashbah, *isim kâda*, *kbabar inna*, *fâ'il*, *nâ'ib al-fâ'il*, *na'at*, *'atf*, *tawkid*, and *badal*. The study applies the same commands to every interaction with ChatGPT and Gemini to ensure consistency of analysis. The purposive sampling technique was used in data collection, with relevant criteria according to the research objectives, so that the results obtained could provide a valid comparative picture of the ability of both models to analyze *i'râb* in Arabic.

Mann-Whitney test was chosen because it was suitable for comparing two independent samples that did not have to be normally distributed, especially on the ordinal scale resulting from the Likert scale assessment (Milenović, 2011). The Likert scale will be used to measure the quality and effectiveness of the answers from both models, where respondents rate within a certain range the answers generated by ChatGPT and Gemini (Siregar et al., 2023).

The test is carried out with the following hypotheses:

- H_0 : There is no significant difference in the analysis ability of *i'râb Marfû'ât al-Asmâ'* between ChatGPT and Gemini.
- H_1 : There is a significant difference in the analytical ability of *i'râb Marfû'ât al-Asmâ'* between ChatGPT and Gemini.

Moreover, the Cosine Similarity Index (CSI) will be used to measure the level of semantic similarity between the AI model's answers and the references in *Mulakhasb's* book. CSI is useful in determining the degree of similarity between the text generated by the model and the standard text, allowing for an objective assessment of the accuracy and relevance of the answers (Rahutomo et al., n.d., 2012; Lahitani et al., 2016).

The Mann-Whitney test is performed by calculating the ratings for each answer from the two models and determining whether there is a significant difference in the distribution of those ratings. The formula used in this test is:

$$U = n_x n_y + \frac{n_x (n_x + 1)}{2} - R_1$$

where n_x and n_y is the sample size of each group, and R_1 is the number of ratings of the first sample.

Meanwhile, Cosine Similarity is calculated using the formula:

$$\text{Similarity} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i \cdot \sum_{i=1}^n B_i}}$$

where A and B is a vector that represents the frequency or weight of words in the reference text and the resulting text of the model AB (Rahutomo et al., n.d., 2012; Oti et al., 2021).

The assessment of the answers given by the two models (ChatGPT and Gemini) in this study was carried out using the Likert scale which is designed to measure the level of conformity of answers with references in the book *Mulakhasb*. This scale consists of five levels of assessment, each of which describes the level of accuracy and

suitability of the answers to the criteria of *Marfû'ât al-Asmâ'*. A score of 5, which indicates a very correct/very precise answer, is given to an answer that is fully in accordance with the reference, including the conformity in position, the status of *mu'rab/mabni*, the final letter, the position, and its grammatical role. A score of 4, which means correct/correct, is given for answers that correspond to the reference even if there are small insignificant differences, such as differences in the last letters. A score of 3, which means that it is quite correct, is given if the answer has a general agreement with the reference, but there are some errors or shortcomings in the details, such as inconsistencies in the status of *mu'rab* and *mabni*. A score of 2 indicates incorrect/incorrect if the answer does not fully match the reference, in terms of its grammatical position/position even though some parts are still acceptable. Finally, a score of 1, which means wrong/inappropriate, is given if the answer does not match the reference and contains significant errors in terms of position, grammatical status, position, and grammatical role.

By using this approach, the study is expected to provide insight into the performance of the two models in the context of Arabic language learning, especially in the analysis of *Marfû'ât al-Asmâ'*.

Result and Discussion

Results of *I'râb ChatGPT Analysis*

The following are presented and the results of the *i'râb* analysis obtained through the Gemini model, as listed in Table 1.

Table 1. *I'râb ChatGPT Results*

قيمة	ChatGPT	الجملة
5	- الدَّهْبُ: مبتدأ مرفوع و علامة رفعه الضمة الظاهرة على آخره - مَعْدِنٌ: خبر مرفوع وعلامة رفعه الضمة الظاهرة على آخره.	الدَّهْبُ مَعْدِنٌ (مبتدع)
5	- الشَّمْسُ: مبتدأ مرفوع و علامة رفعه الضمة الظاهرة على آخره. - أَشْرَقَتْ: فعل ماض على الفتح، و التاء التأنيث الساكنة، لامحل لها من الإعراب، و الجملة الفعلية "أشرفت" في محل رفع خبر للمبتدأ "الشمس".	الشَّمْسُ أَشْرَقَتْ (خبر)
5	- أَصْبَحْتُ: فعل ناقص مبني على السكون لاتصاله بضمير الرفع المتحرك، و التاء ضمير متصل مبني في محل رفع اسم "أصبح". - مَتَفَانًا: خبر "أصبح" و علامة نصبه الفتحة الظاهرة على آخره.	أَصْبَحْتُ مَتَفَانًا (إسم أصبح)
5	- كَادَتْ: فعل الماض ناقص (من أفعال المقاربة) مبني على الفتح، و التاء للتأنيث الساكنة، لامحل لها من الإعراب. - الشَّمْسُ: اسم كاد مرفوع، و علامة رفعه الضمة الظاهرة على آخره. - تَشَرَّقُ: فعل المضارع مرفوع، و علامة رفعه الضمة الظاهرة على آخره، و هو في محل نصب خبر كاد.	كَادَتْ الشَّمْسُ تَشَرَّقُ (إسم كاد)
5	- إِنَّ: حرف توكيد ونصب مبني على الفتح لا محل له من الإعراب. - مِصْبَاحٌ: اسم إن منصوب و علامة نصبه الفتحة الظاهرة على آخره.	إِنَّ مِصْبَاحَ ضَوْؤُهُ شَدِيدٌ (خبر إن)

	- ضَوْؤُهُ: مبتدأ مرفوع و علامة رفعه الضمة الظاهرة على آخره، و هو مضاف. الهاء: ضمير متصل مبني في محل جر مضاف إليه. - شديد: خبر المبتدأ "ضوء" مرفوع و علامة رفعه الضمة الظاهرة على آخره. الجملة الاسمية "ضوءه شديد" في محل رفع خبر "إن".	
5	- نَجَّحَ: فعل ماض مبني على الفتح - هذا: اسم إشارة في محل رفع فاعل، و علامة رفعه الضمة المقدرة. - الطالب: بدل أو عطف بيان مرفوع، و علامة رفعه الضمة الظاهرة على آخره.	نَجَّحَ هَذَا الطَّالِبُ (فاعل)
5	- أقبِل: فعل ماض مبني للمجهول مبني على الفتح و الضمير المتصل "أل" في محل رفع نائب فاعل. - إقبال: نائب فاعل مرفوع و علامة رفعه الضمة الظاهرة على آخره. - شديد: نعت (صفة) ل "إقبال" مرفوع و علامة رفعه الضمة الظاهرة على آخره.	أَقْبَلَ إِقْبَالَ شَدِيدٌ (نائب الفاعل)
5	- هذا: اسم إشارة مبني في محل رفع مبتدأ. - عمل: خبر المبتدأ "هذا" و علامة رفعه الضمة الظاهرة على آخره. - يفيد: فعل مضارع و علامة رفعه الضمة الظاهرة على آخره، و الفاعل ضمير مستتر تقديره (هو) يعود على "عمل"، و جملة الفعلية "يفيد" في محل رفع نعت ل "عمل".	هَذَا عَمَلٌ يَفِيدُ (نعت)
5	- نجحت: فعل الماض مبني على الفتح و التاء تاء التانيث الساكنة. - سعاد: فاعل مرفوع، و علامة رفعه الضمة الظاهرة على آخره. - أختها: معطوف على سعاد مرفوع، و علامة رفعه الضمة الظاهرة على آخره، و أخت مضاف، و ها ضمير متصل مبني في محل جر مضاف إليه.	نَجَّحَتْ سَعَادٌ وَ أُخْتُهَا (عطف)
1	- حضر: فعل ماض مبني على الفتح. - القوم: فاعل مرفوع و علامة رفعه الضمة الظاهرة على آخره. - عامتهم: مضاف إليه مجرور بالكسرة الظاهرة على آخره و هي نعت (صفة) ل "القوم" و الهاء ضمير متصل في محل جر مضاف إليه.	حَضَرَ القَوْمَ عَامَتِهِمْ (توكيد)
5	- طبع: فعل ماض مبني للمجهول مبني على الفتح - الكتاب: نائب فاعل مرفوع و علامة رفعه الضمة الظاهرة على آخره. - جزء: بدل من "الكتاب" مرفوع و علامة رفعه الضمة الظاهرة على آخره. و الهاء ضمير متصل في محل جر مضاف إليه. - الأول: نعت (صفة) ل "جزء" مرفوع و علامة رفعه الضمة الظاهرة على آخره.	طَبَعَ الكِتَابَ جُزْؤُهُ الأَوَّلُ (بدل)

a) Sentence: الذَّهَبُ مَعْدِنٌ

ChatGPT provides an explanation that the word "الذَّهَبُ" is *mubtada'* which is *marfu'* with the sign *i'rāb* in the form of *dhammah* that appears at the end of the word, while "مَعْدِنٌ" is the *khabar* that *marfu'* with the sign *i'rāb* in the form of *dhammah* at the end of the word. This answer is completely in accordance with the references in the book of *Mulakhasb*, including the position of the word, the status of the *i'rāb*, and the conformity of the final letter, resulting in a score of 5 (Very True).

b) Sentence: الشمسُ أشرقَت

ChatGPT's explanation of the word "الشمسُ" as *mubtada' marfu'* with the sign *dhammah* at the end, and "أشرقَت" as the number of *jumlah fi'liyyah* becomes the *kbabar* of "الشمسُ", according to the reference to the book of *Mulakhash*. This accuracy includes aspects of grammatical position and relationships between elements in a sentence, so it is given a score of 5 (Very True).

c) Sentence: أصبحت متفائلاً

ChatGPT explains that "أصبحت" is *fi'il nāqish* and *dhamir* as its *isim*, and "متفائلاً" is a *manshub kbabar* with a *fathah* sign that appears at the end of the word. This explanation is in accordance with the reference, covering the grammatical status of the word and the role of its position. Since no errors were found in this analysis, the value given was 5 (Very True).

d) Sentence: كاذب الشمسُ تشرقُ

ChatGPT identified "كاذب" as *fi'il nāqish* from the *af'al muqārabah* category "الشمسُ" as *ism marfu'*, and "تشرقُ" as *fi'il muḍari'* in the *kbabar* position. According to the reference book, all the words in the *i'rāb* are correct, so they are given a grade of 5 (Very True).

e) Sentence: إنَّ مصباحَ ضوءه شديدٌ

In this sentence, ChatGPT explains that "إنَّ" is a *tawkid* letter that requires *isim manshub*, which is "مصباح", while "ضوءه شديدٌ" is *kbabar* in the form of *jumlah ismiyyah*. This explanation is complete and includes the exact position and status of the *i'rāb*, so it is given a score of 5 (Very True).

f) Sentence: نَجَحَ هَذَا الطَّالِبُ

ChatGPT explains that "نَجَحَ" is *fi'il madhi*, "هَذَا" is *isim isyārah* in the position of *rafa'* occupying the position of *fa'il*, and "الطَّالِبُ" is *badal* that *marfu'*. This analysis accurately explains the functions and *i'rāb* signs of each element. Therefore, the value given is 5 (Very True).

g) Sentence: أُقْبِلَ إِبْرَاهِيمُ شَدِيدًا

ChatGPT provides an analysis that "أُقْبِلَ" is *fi'il mādhī mabnī lil majbūl* "إِبْرَاهِيمُ" is *na'ib al-fa'il*, and "شَدِيدًا" *shifat* that is *marfu'*. All words are explained according to the book of *Mulakhash*, so it deserves to be given a grade of 5 (Very True).

h) Sentence: هذا عملٌ يفيدُ

In this sentence, ChatGPT identifies "هذا" as *mubtada'*, "عملٌ" as *kbabar*, and "يفيدُ" as *fi'il mudhari'* in the *na'at* position. A complete and appropriate explanation provides the basis for a score of 5 (Very True).

i) Sentence: نَجَحَتْ سَعَادٌ وَأُخْتُهَا

ChatGPT mentions that "نَجَحَتْ" is *fi'il madhi*, "سَعَادٌ" is *fa'il*, and "أُخْتُهَا" is *ma'thub*. This

explanation is in accordance with the grammatical aspect analyzed, so that it gets a score of 5 (Very True).

j) Sentence: حَضَرَ الْقَوْمَ عَامَّتَهُمْ

ChatGPT provides an inaccurate analysis in analyzing the position of "عَامَّتَهُمْ" which should be *tamkîd ma'nawi*, but ChatGPT explains it as *na'at* in the jar position because of *idāfab*. So it was given a score of 1 (False)

k) Sentence: طَعَّ الْكِتَابُ جُزْؤُهُ الْأَوَّلُ

ChatGPT's explanation of "طَعَّ" as *fi'il madhi mabni li al-majbul*, "الْكِتَابُ" as *na'ib al-fa'il*, and "جُزْؤُهُ الْأَوَّلُ" as *badal* and *shifat*, shows the rigor of grammatical analysis. Therefore, the value given is 5 (Very True).

Results of I'rāb Gemini Analysis

The following is presented the results of the *i'rāb* analysis obtained through the Gemini model, as listed in Table 2 below.

Table 2. I'rāb Gemini Results

قيِّمة	Gemini	الجملة
5	- المبتدأ: الذَّهَبُ (اسم ظاهر مرفوع وعلامة رفهه الضمة الظاهرة على آخره) يدل على الشيء الذي يتم وصفه أو الحديث عنه. - الخبر: مَعْدِنٌ ((اسم ظاهر مرفوع وعلامة رفهه الضمة الظاهرة على آخره). يصف المبتدأ ويوضح ماهيته.	الذَّهَبُ مَعْدِنٌ (مبتدأ)
2	- الشمسُ: مبتدأ مرفوع و علامة رفهه الضمة الظاهرة على آخره. - أَشْرَقَتْ :: فعل ماض مبني على الفتح، و التاء تاء التانيث الساكنة، و الفاعل ضمير مستتر تقديره هي.	الشمسُ أَشْرَقَتْ (خبر)
2	- أَصْبَحْتُ: فعل ماض مبني على الفتح، و التاء تاء التانيث الساكنة، و الفاعل ضمير مستتر وجوبا تقديره أنا. - مَتَفَانًا: خبر كان مرفوع و علامة رفهه الضمة الظاهرة على آخره.	أَصْبَحْتُ مَتَفَانًا (إسم أصبح)
1	- كَادَتْ: فعل ماض ناقص مبني على الفتح و التاء تاء التانيث الساكنة. - الشَّمْسُ: فاعل مرفوع و علامة رفهه الضمة الظاهرة على آخره. - تَشْرُقُ: فعل مضارع مرفوع و علامة رفهه الضمة الظاهرة على آخره.	كَادَتْ الشَّمْسُ تَشْرُقُ (إسم كاد)
1	- إِنَّ: حرف نصب يدل على الجملة الاسمية ليجعلها جملة إن. - مِصْبَاحٌ: اسم إن منصوب و علامة نصبه الفتحة الظاهرة على آخره. و هو المبتدأ المقدر محذوف و جملة إن خبر له. - ضَوْؤُهُ: صفة لمصباح مرفوع محذوفة تقديرها (هو) و جرت و علامة جرهما	إِنَّ مِصْبَاحَ ضَوْؤُهُ شَدِيدٌ (خبر إن)

	الكسرة الظاهرة على آخره. و الهاء ضمير متصل مبني في محل جر مضاف إليه. - شديداً: خبر الصفة المرفوعة المحذوفة (هو).	
2	- نَجَحَ: فعل ماض مبني على الفتح، و الفاعله ضمير مستتر تقديره (هو) و جاء مرفوعاً. - هذا: اسم إشارة مبني على السكون في محل رفع مبتدأ. - الطالب.: خبر المبتدأ مرفوع و علامة رفعه الضمة الظاهرة على آخره.	نَجَحَ هَذَا الطَّالِبُ (فاعل)
2	- أقبِل: فعل ماض مبني للمفعول به، و الفاعل ضمير مستتر تقديره (هو) أي الإقبال. - إقبال: مفعول به مرفوع و علامة رفعه الضمة الظاهرة على آخره. - شديد: خبر مرفوع و علامة رفعه الضمة الظاهرة على آخره.	أَقْبَلَ إِبْقَالَ شَدِيدًا (نائب الفاعل)
2	- هذا: اسم إشارة مبني على السكون في محل رفع مبتدأ. - عمل: خبر المبتدأ مرفوع و علامة رفعه الضمة الظاهرة على آخره. - يفيد: فعل مضارع و علامة رفعه الضمة الظاهرة على آخره، و الفاعل ضمير مستتر تقديره (هو) و الهاء تعود على العمل.	هَذَا عَمَلٌ يَفِيدُ (نعت)
5	- نجحت: فعل ماض مبني على الفتح و التاء التانيث الساكنة. - سعاد: فاعل مرفوع و علامة رفعه الضمة الظاهرة على آخره. - و: حرف عطف - أختها: معطوف على سعاد، مرفوع و علامة رفعه الضمة المقدرة على الألف لأنه ممنوع من الصرف، و الهاء ضمير متصل مبني على السكون في محل جر مضاف إليه.	نَجَحْتَ سَعَادٌ وَأُخْتِيَا (عطف)
5	- حضر: فعل ماض مبني على الفتح و الفاعل ضمير مستتر تقديره (هم). - القوم: فاعل مرفوع و علامة رفعه الضمة الظاهرة على آخره. - عامتهم: توكيد معنوي مرفوع و علامة رفعه الضمة الظاهرة على آخره، و الهاء ضمير متصل في محل جر مضاف إليه.	حَضَرَ الْقَوْمُ عَامَتَهُمْ (توكيد)
5	- طبع: فعل ماض مبني للمفعول، و الفاعل ضمير مستتر تقديره (هو). - الكتاب: مفعول به مرفوع و علامة رفعه الضمة الظاهرة على آخره. - جزؤه: بدل مرفوع و علامة رفعه الضمة الظاهرة على آخره و الهاء ضمير متصل في محل جر مضاف إليه. - الأول: نعت مرفوع و علامة رفعه الضمة الظاهرة على آخره.	طَبَعَ الْكِتَابُ جُزْؤُهُ الْأَوَّلُ (بدل)

a) Sentence: الدَّهْبُ مَعْدِنٌ

Gemini analyzes this sentence very precisely. *Mubtada'* (الدَّهْبُ) is identified as *isim z̤habir marfu'* with the sign *rafa'* in the form of *dhammah z̤habirah*, which is fully in accordance with the reference in the book of *Mulakhash*. *Khabar* (معدن) is also correctly classified as *isim z̤habir marfu'* with the same *rafa'* sign. This analysis reflects an accurate understanding of the structure of the *i'rab* in the book of *Mulakhash*, so it deserves a score of 5 (Very True).

b) Sentence: الشمسُ أشرقت

Gemini identifies الشمسُ as *mubtada' marfu'* with the sign *rafa' dhammah z̤habirah*, which is correct. However, there is an error in the analysis of أَشْرَقَتْ, where it is not explained that the *fi'il madhi* occupies a position or position as *khabar* derived from the number of *jumlah fi'liyyah*. This causes inconsistencies with the book of *Mulakhash*, so it is given a score of 2 (Less True).

c) Sentence: أصبح متفائلاً

Gemini provides an improper analysis for this sentence. *Isim dhamir* should have the position of *isim* of أصبح. Gemini showed a basic understanding but failed to present the details of the *i'rab*, this obtaining a score of 2 (Less True).

d) Sentence: كادت الشمس تشرق

Gemini's analysis of this sentence does not correspond to the reference. *Kāda wa Akhwātuba* (كادت) is analyzed without regard to the role of *isim kāda* (الشمس), which should be *isim kāda* but is in *i'rāb* as *fā'il*, thus getting a score of 1 (False).

e) Sentence: إن مصباح ضوءه شديد

Gemini difficulties in *i'rāb khabar inna* derived from the number of *jumlah ismiyyah*. Gemini does not explain that the sentence ضوءه شديد is the number of *jumlah ismiyyah* that occupies the position of *khabar inna*. This is what causes Gemini to get a score of 2 (Less True).

f) Sentence: نجح هذا الطالب

Gemini analyzes *isim isyarah* (هذا) as *isim marfu'* with *dhammah z̤habirah*, which occupies the position of *rafa' mubtada*. This is wrong because *isim isyarah* (هذا) should occupy the position of *rafa' fa'il*. So it was given a score of 2 (Less True).

g) Sentence: أقبل إقبالاً شديداً

Gemini identified this structure with a mistake in the status of *na'ib al-fa'il* (إقبال). Although the *rafa'* sign is correctly presented, the grammatical position analysis does not conform to the reference from the book of *Mulakhash*. Therefore, the score given is 2 (Less True).

h) Sentence: هذا عملاً يفيد

Gemini provides an inaccurate analysis in analyzing the position of the number of *fi'liyyah* يفيد that returns to the previous word. Gemini does not explain that يفيد is the

na'at of *عَمَلٌ* according to the reference to the book of *Mulakhash*. Therefore, the score given is 2 (Less True).

i) Sentence: نَجَحَتْ سَعَادٌ وَأُخْتُهَا

Gemini's analysis fits the reference perfectly. *Fa'il* (سَعَادٌ) and *ma'thub* (أُخْتُهَا) are each classified as *isim marfu'* with the correct sign. No errors or omissions were found, so the score given was 5 (Very True).

j) Sentence: حَضَرَ الْقَوْمَ عَامَتَهُم

Gemini shows an excellent understanding of *fā'il* (الْقَوْمَ) and *tawkid ma'nawi* (عامتهم) as *isim marfu'*. The word was analyzed correctly, thus getting a 5 (Very True) score.

k) Sentence: طَبَعَ الْكِتَابُ جُزْؤُهُ الْأَوَّلُ

Gemini provides a very accurate analysis of this sentence. *Badal* (جُزْؤُهُ) is clearly explained and in accordance with the references in the book of *Mulakhash*. No flaws were found, so it was given a score of 5 (Very True).

Analysis Based on the Mann-Whitney Test

The Mann-Whitney test was used to compare the difference in score distribution between two groups, namely GPT and Gemini. Based on the Ranks table:

Table 3. Mann-Whitney

		Ranks		
	Kelompok	N	Mean Rank	Sum Of Ranks
Skor	ChatGPT	11	14,27	157,00
	GEMINI	11	8,73	96,00
Total		22		

The ChatGPT group has a mean rank of 14.27, which is higher than the Gemini group (8.73). This shows that GPT scores tend to be higher than Gemini scores on average.

Statistical Test Results

The following table summarizes the results of the Mann-Whitney statistical test:

Tabel 4. Tabel Hasil Uji Statistik

Test Statistics ^a	
	skor
Mann-Whitney U	30,000

Wilcoxon W	96,000
Z	-2,345
Asymp. Sig. (2-tailed)	0,019
Exact Sig. [2*(1-tailed Sig.)]	.047 ^b
a. Grouping Variable: kelompok	
b. Not corrected for ties.	

The Asymp. Sig. (2-tailed) *value* of 0.019 was smaller than 0.05, indicating that there was a significant difference between the scores of the two groups. Thus, there is evidence that the difference in performance between GPT and Gemini is statistically significant. So the hypothesis, that there is a significant difference in the analytical ability of *i'rab marfû'atul asma'* between ChatGPT and Gemini, is accepted H_1 .

Results Based on Cost Similarity Index (CSI)

The Cost *Similarity Index* (CSI) result of 0.800 shows a high level of similarity between GPT and Gemini. This value is close to 1, which indicates that despite the difference in performance, the answers produced by the two are almost similar in some examples.

The comparative study between ChatGPT and Gemini was conducted using the Mann-Whitney statistical test approach and the Cost Similarity Index (CSI) to measure significant differences in the performance of the two systems in analyzing *i'rab*. Based on the results of the Mann-Whitney test, there are significant differences in several aspects of the analysis, especially in terms of accuracy and speed, with a significance value of 0.019. This suggests that the hypothesis, that there is a significant difference in the analytical capabilities of *Marfû'ât al-Asmâ'* between ChatGPT and Gemini, is accepted H_1 .

Meanwhile, the results of the CSI measurement show a fairly high level of similarity between the two systems with the reference standard, with a score of 0.800. This suggests that the two systems have sufficient performance similarities, although there are variations in some case studies, especially in the context of complex sentence analysis.

Conclusion

Based on the discussion above, this study shows that artificial intelligence technology, especially language models such as ChatGPT and Gemini, has significant potential in supporting Arabic grammar analysis, especially in the aspect of *Marfû'ât al-Asmâ'*. Based on the results of statistical tests using the Mann-Whitney method, there are significant differences between the two models, where ChatGPT proves to be superior in terms of accuracy and depth of analysis. ChatGPT provides a more complete and precise response in explaining complex grammatical structures, making it more

suitable for deep learning. Meanwhile, Gemini shows excellence in providing quick and concise responses, but tends to be less accurate in handling more complex sentences.

The level of answer similarity between the two models is also high, as shown by the Cosine Similarity Index (CSI) result of 0.800. This shows that despite the difference in performance, the answers produced by the two models match fairly well, according to the references in some questions. This research provides important insights into how artificial intelligence can be effectively utilized for Arabic grammar learning, with the selection of models tailored to the needs and complexity of the desired analysis. Thus, this study not only strengthens the role of artificial intelligence in Arabic language learning but also underscores the importance of further research to improve the ability of AI models to understand and analyze complex grammatical structures of Arabic.

REFERENCES

- Aladeemy, A. A., Alzahrani, A., Algarni, M. H., Alsubari, S. N., Aldhyani, T. H. H., Deshmukh, S. N., Khalaf, O. I., Wong, W.-K., & Aqburi, S. (n.d.). Advancements and Challenges in Arabic Sentiment Analysis: A Decade of Methodologies, Applications, and Resource Development. *Heliyon*.
- Ali, M. A. (2016). Artificial intelligence and natural language processing: The Arabic corpora in online translation software. *International Journal of Advanced and Applied Sciences*, 3(9).
- Aljanabi, M. (2024). Assessing the Arabic Parsing Capabilities of ChatGPT and Cloude: An Expert-Based Comparative Study. *Mesopotamian Journal of Arabic Language Studies*, 2024.
- Anwar, M. R., & Ahyarudin, H. A. (2023). AI-Powered Arabic Language Education in the Era of Society 5.0. *LAIC Transactions on Sustainable Digital Innovation (ITSDI)*, 5(1).
- Berliani, D., Azhari, I., & Taufiqurrochman, R. (2024). Exploring The Potential Of Chatgpt As A Technology-Based Tool For Arabic Language Learning: Benefits And Risks. *Kitaba*, 2(2).
- Elkhatat, A. M., Elsaid, K., & Almeer, S. (2023). Evaluating the efficacy of AI content detection tools in differentiating between human and AI-generated text. *International Journal for Educational Integrity*, 19(1).
- Gomez-Cabello, C. A., Borna, S., Pressman, S. M., Haider, S. A., & Forte, A. J. (2024). Large Language Models for Intraoperative Decision Support in Plastic Surgery: A Comparison between ChatGPT-4 and Gemini. *Medicina*, 60(6).
- Khalatia, M., & Al-Romanyb, T. A. H. (2020). Artificial intelligence development and challenges (Arabic language as a model). *Artificial Intelligence*, 13(5).
- Khidhir, A. M. (2024). An AI model for Parsing the Text of Holy Quran Sentences. *Mesopotamian Journal of Quran Studies*, 2024.

- Kishan, A. U., & Reiter, R. E. (2021). Caught Between a Rock and a Hard Place. *International Journal of Radiation Oncology, Biology, Physics*, 109(4).
- Lahitani, A. R., Permanasari, A. E., & Setiawan, N. A. (2016). Cosine similarity to determine similarity measure: Study case in online essay assessment. *2016 4th International Conference on Cyber and IT Service Management*.
- Mannaa, Z. M., Azmi, A. M., & Aboalsamh, H. A. (2022). Computer-assisted i 'raab of Arabic sentences for teaching grammar to students. *Journal of King Saud University-Computer and Information Sciences*, 34(10).
- Marsiah, M., Mubarak, M. R., & Audina, N. A. (2021). The students' perspective towards youtube as the replacement of lecturer in nahwu learning. *Jurnal Al Bayan: Jurnal Jurusan Pendidikan Bahasa Arab*, 13(2).
- Maulida, M., & Marsiah, M. (2023). Mapping Kesulitan Dalam Membaca dan Mengharokati Teks Bahasa Arab Bagi Mahasiswa Pendidikan Bahasa Arab. *Journal Of Foreign Language Learning and Teaching*, 3(2).
- Milenović, Ž. (2011). Application Of Mann-Whitney U Test In Research Of Professional Training Of Primary School Teachers. *Metodički Obzori/Methodological Horizons*, 6(1).
- Mualif, A. (2019). Metodologi pembelajaran ilmu nahwu dalam pendidikan bahasa Arab. *AL-HIKMAH (Jurnal Pendidikan Dan Pendidikan Agama Islam)*, 1(1).
- Muawanah, R., & Rifa'i, A. F. (2018). Analisis Kedudukan Prab Kalimah Bahasa Arab Pada Kitab Al-Imrithi Menggunakan Algoritma Breadth First Search (BFS). *JISKA (Jurnal Informatika Sunan Kalijaga)*, 3(1).
- Musthafa, F. A. D. (2024). Penggunaan Artificial Intelligence (AI) dalam Pembelajaran: Fenomena Transformasi Otoritas Pengetahuan di Kalangan Mahasiswa. *Journal of Contemporary Islamic Education*, 4(1).
- Nurmala, M., Sopian, A., Saleh, N., & Syihabuddin, S. (2022). Why is Masdar Important? An Investigating of Masdar and its Translation. *Asuniyat: Jurnal Penelitian Bahasa, Sastra, Dan Budaya Arab*, 5(1).
- Oti, E. U., Olusola, M. O., & Esemokumo, P. A. (2021). *Statistical Analysis of the Median Test and the Mann-Whitney U Test*. 7(9).
- Pressman, S. M., Borna, S., Gomez-Cabello, C. A., Haider, S. A., & Forte, A. J. (2024). AI in Hand Surgery: Assessing Large Language Models in the Classification and Management of Hand Injuries. *Journal of Clinical Medicine*, 13(10).
- Qadir, J. (2023). *Engineering education in the era of ChatGPT: Promise and pitfalls of generative AI for education*.
- Rahmat, A. S., & Abdurrahman, M. (2017). Analysis of Nahwu Learning Progress in Indonesian Universities Based on the Character of Value Concerns. *Tawarikh*, 8(2).
- Rahutomo, F., Kitasuka, T., & Aritsugi, M. (n.d.). *Semantic Cosine Similarity*.

- Ramli, M. (2023). Mengeksplorasi tantangan etika dalam penggunaan Chat GPT sebagai alat bantu penulisan ilmiah: Pendekatan terhadap integritas akademik. *TA'DIBAN: Journal of Islamic Education*, 4(1).
- Rane, N., Choudhary, S., & Rane, J. (2024). Gemini or ChatGPT? Capability, Performance, and Selection of Cutting-edge Generative Artificial Intelligence (AI) in Business Management. *Capability, Performance, and Selection of Cutting-Edge Generative Artificial Intelligence (AI) in Business Management (February 19, 2024)*.
- Siregar, F. H., Hasmayni, B., & Lubis, A. H. (2023). The analysis of chat gpt usage impact on learning motivation among scout students. *International Journal of Research and Review*, 10(7).
- Suryono, M. N. R. N., Bhagaskara, R. E., Pratama, M. A., & Pratama, A. (2023). *Analisis Pengaruh ChatGPT Terhadap Produktivitas Mahasiswa*. 3(1).
- Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity*, 19(1).
- Wijonarko, P., Wagiman, W., Khana, R., Tundo, T., Salam, A., & James, B. (2023). Penerapan dan Kontribusi Kecerdasan Buatan ChatGPT Untuk Menafsir Teks Hukum (Studi Kasus Penafsiran Pasal 10, Pasal 13, Permenkes No. 889 Tahun 2011). *Jurnal Kajian Teknik Elektro*, 8(2).
- Yatri, D., Anugerahwati, M., & Setyowati, L. (2023). Artificial Intelligence (Ai) In Language Learning (English And Arabic Class): Students'and Teachers'experience And Perceptions. *Transformational Language, Literature, And Technology Overview In Learning*, 3(1).