

Lost in Semantics: Assessing the Precision of Gemini's Arabic to Indonesian Translation

Nely Rahmawati Zaimah¹, Alam Budi Kusuma²

¹ Sekolah Tinggi Agama Islam Al-Anwar Rembang, Indonesia

² Sekolah Tinggi Agama Islam Masjid Syuhada Yogyakarta, Indonesia

 alambudi.kusuma.yahoo.com

Abstract

This research critically evaluated Google Gemini's performance in translating and interpreting classical Arabic literature, specifically focusing on the books *Alfiyah Ibn Malik* and *Nadham al-Imrithi*, into Indonesian. A collect-measure-repeat methodology was employed as a taxonomic investigation, comparing AI-generated translations with those from verified translation books and human experts to assess both literal accuracy and contextual comprehension. The findings highlighted that while Gemini consistently provided precise word-for-word translations, it frequently overlooked contextual nuances and detailed interpretations. Despite demonstrating a solid understanding of overarching themes, the model struggled to capture the nuanced meanings and subtle contextual elements inherent in classical Arabic texts. An observed inconsistency in language output—such as the occasional inclusion of English responses despite Indonesian input—suggested limitations in linguistic fidelity. From a semantic standpoint, Gemini excelled in lexical semantics by accurately rendering individual words and phrases. However, it fell short in pragmatic and contextual semantics, failing to convey implied meanings and cultural references effectively. This limitation hindered its ability to faithfully reproduce the deeper thematic interconnections present in the original texts. Nevertheless, the translations generally provided a foundational understanding of the texts' basic messages, indicating considerable potential for the model. Future refinement efforts should prioritize improving contextual understanding and enhancing the depth of interpretation to make Gemini a more dependable tool for translating intricate literary texts. Future research is recommended to conduct comparative evaluations across different AI translation models, explore pragmatic and cultural semantics more deeply, and develop interdisciplinary frameworks that integrate linguistic, pedagogical, and technological perspectives.

Keywords: Artificial Intelligence; Google Gemini; Gemini translation performance, Arabic interpretation, AI's contextual understanding

ARTICLE INFO

Article history:

Received

May 28, 2026

Revised

June 20, 2026

Accepted

June 29, 2026

Published by
Website

ISSN

Copyright



STAI Masjid Syuhada Yogyakarta

<https://journal.stainsy.ac.id/index.php/Ihtimam/index>

2774-4302

This is an open access article under the CC BY SA license

<https://journal.stainsy.ac.id/index.php/Ihtimam/index>

@ 2026 by the author (s)

INTRODUCTION

The utilization of artificial intelligence in text translation has undergone significant advancement in recent years (Kolhar & Alameen, 2021; Läubli et al., 2020). One prominent platform in this domain is Google Gemini, as clearly articulated by Imran & Almusharraf (2024) and Gemini Team et al. (2024). It offers advanced text translation and other capabilities using state-of-the-art generative AI technology, as perfectly illustrated by Ajao

(2023) and Bandi et al., (2023). Despite Gemini's promising translation capabilities, there has not yet been an in-depth study specifically examining its proficiency in translating complex Arabic texts (Khoshafah, 2023). Additionally, other language learning fields (Author et al., 2024 a; Author, et al., 2024 b), particularly in the context of education and classical literature (Azmi et al., 2019; Fuad & Al-Yahya, 2022; Labadze et al., 2023), remain underexplored.

The efficacy of Google Gemini in accurately translating Arabic texts warrants scrutiny of its robustness (Ram & Verma, 2023), considering that Arabic is one of the languages with unique and intricate grammar rules and structures. Moreover, in the context of education and classical literature studies, it is essential to understand and interpret Arabic texts literally and contextually (Hajir et al., 2022). However, challenges arise due to the complex grammar rules, rich cultural nuances, and specialized terminologies inherent in the Arabic language, requiring a deep understanding (Alruqi & Alzahrani, 2023; Tiawaldi & Abdul Wahab, 2017).

The surge of information technology is fundamentally changing how we learn languages and study literature, as evidenced by Almelhes (2023), Annamalai et al. (2023) and Cantú-Ortiz et al. (2020). Among these advancements, large language models like Gemini and ChatGPT stand out as particularly influential. The widespread availability of advanced translation platforms like Google's Gemini offers researchers unprecedented opportunities to explore and analyze Arabic language and classical Arabic literature (Atwell et al., 2011).

In previous year, Jan Kocoń et al. (2023) found that ChatGPT performs well on semantic tasks like entity recognition and sentiment analysis but struggles with complex pragmatic tasks, such as understanding context and emotions in text. Meanwhile, Feng et al., (2024) proposed the TER framework (Translate, Estimate, Refine) to improve LLM-based translation quality, addressing persistent errors found in translations by models like GPT-4, Gemini, and Claude. TER enhances translation across various languages, demonstrating its effectiveness and introducing a self-correcting mechanism for machine translation. Inel et al., (2023) introduced the Responsible AI (RAI) methodology to assess AI dataset quality, emphasizing transparency, fairness, and accountability by addressing biases and inaccuracies. Their approach includes considering annotator demographics and thorough documentation of data provenance. Future research should explore unified reproducibility scores and cognitive bias mitigation. There's a need for further research to develop robust validity checklists for AI applications across various fields. These checklists should address user acceptance, potential cognitive biases in data collection, and ensure the reliability and integrity of datasets (Khoshafah, 2023; Labadze et al., 2023; Author et al., 2024a; Author et al., 2024b).

While previous studies have explored how language models like ChatGPT and Gemini handle general tasks, there's been little focus on translating classical Arabic literature. This research fills that gap by closely evaluating how well Gemini captures the deeper meanings and cultural context of Arabic texts. By focusing on its strengths and weaknesses, this study adds fresh insights to the field and highlights Gemini's potential in AI-driven language translation.

This research evaluates Gemini's performance in translating and interpreting classical Arabic literature into Indonesian. Assessing its lexical and contextual semantics provides an opportunity to understand its strengths and limitations in capturing the depth of meaning in such texts. Therefore, the research questions focus on the direct accuracy of Gemini's translations and its ability to interpret nuanced meanings effectively.

Research Questions (RQ):

RQ1: How well does Google Gemini translate Arabic texts directly and literally into the target language?

RQ2: How deeply does Google Gemini interpret Arabic texts in terms of lexical and contextual semantics?

METHOD

This study employs a qualitative approach through literature review and taxonomy investigation (Categorization), focusing on the integration of Google Gemini in translating Arabic into Indonesian and examining its effectiveness in interpreting classical Arabic texts. The evaluation centers on Gemini model 2.5 flash, tested via eight repetitive queries drawn from *Alfiyah Ibn Malik* and *Nadham al-Imrithi*. Following Inel’s method (Inel et al., 2023), three evaluators with expertise in classical Arabic assessed Gemini’s performance, with five queries (Q1–Q5) measuring literal accuracy against Haq’s (Haq, 2022) Indonesian translation of *Alfiyah Ibn Malik*, and three queries (Q6–Q8) evaluating contextual comprehension using *Nadham al-Imrithi* alongside references from Hidayatullah & Fauji (Hidayatullah & Fauji, 2023). To minimize bias, identical queries were tested across two separate Google accounts on a Windows system, ensuring objectivity and consistency, similar to prior studies by Akter et al. (Akter et al., 2023) and Inel et al. (Inel et al., 2023).

Data collection followed the “collect–measure–repeat” framework (Inel et al., 2023) comprising Gemini’s translation outputs, repeated iterations for consistency, and secondary sources such as verified translations, academic journals, and expert reviews. These were systematically synchronized and categorized under Spradley’s taxonomy model, with human experts in Arabic linguistics supervising the thematic classification (Garrido, 2017). The evaluation emphasized translation accuracy, fluency, contextual appropriateness, and adherence to Arabic language standards. Three expert reviewers enriched the analysis with diverse perspectives, enhancing reliability. The final report documents findings, analysis, and recommendations, offering insights into Gemini’s potential role in Islamic educational institutions and guiding future research on AI integration in classical Arabic translation.

RESULT AND DISCUSSION

Reviewing the translations reveals a careful and systematic approach taken by each test model, breaking down every word with detailed explanations. This method highlights a deep understanding of Arabic language structures, including precise identification of letters and punctuation. The translations demonstrate a strong commitment to accuracy in conveying the intended meaning of the text. Through this thorough analysis, it’s evident that the test models show commendable proficiency in capturing the nuances of Arabic. While they don’t yet excel in handling analogies, metaphors, or deep contextual interpretation, Gemini’s translations are still notably effective.

Table 1. Inel’s (2023) Model for Gemini’s Translation Result

Q	Collect	Measure	Repeat
Q1	Detailed and systematic, with a breakdown of each word and its explanation Accurately identifying letters and punctuation marks.	Understands Arabic language construction, accurately identifying letters and punctuation marks. Provides additional insight into the implicit message of the verse, namely the analogy between studying Arabic Syntax and remembering Allah, which provides ease.	3 times tasks with the same answers in 3 prompts

Q2	Detailed and focused on understanding word-for-word. Highlights differences in sentence structure and provides examples. Accurately identifying letters and punctuation marks.	Accurately explains the meaning of each word. Captures the implicit meaning, but with less detail compared to Q1.	3 times tasks with the same answers in 3 prompts
Q3	Detailed and focused on understanding word-for-word. Identifies key grammatical components. Accurately identifying letters and punctuation marks.	Consistently accurate in word identification. Lacks additional contextual information.	3 times tasks with the same answers in 3 prompts with different positions
Q4	Focused on understanding word-for-word. Occasional misinterpretation of word meanings. Emphasizes semantic meaning even if not asked. Accurately identifying letters and punctuation marks.	Maintains consistent accuracy in identifying words and their meanings. Less detailed in capturing contextual nuances.	3 times tasks with the same answers in 3 prompts
Q5	Some mistake in word-for-word understanding. Integrates historical aspects. Not entirely accurate in identifying punctuation marks.	Contains minor errors in one-word interpretation. Accuracy in punctuation identification needs improvement.	3 times tasks with the same answers in 3 prompts with different formats and positions

The investigation focused on detailed collection methods for five queries (Q1-Q5). For Q1, Gemini provided a systematic breakdown of each word and explanation, accurately identifying letters and punctuation marks, and offered insights into the implicit message of the verse. Q2 emphasized word-for-word translation and sentence structure differences, providing examples and additional insights. Q3 followed a similar approach, identifying key grammatical components. Q4 focused on word-for-word understanding but occasionally misinterpreted meanings while emphasizing semantic content. Q5 exhibited some errors in word-for-word understanding, including misspellings and inaccuracies in identifying punctuation marks, though it integrated historical aspects and provided some contextual insights. Overall, while Gemini demonstrated a strong grasp of translation, areas like punctuation accuracy and contextual interpretation require further refinement.

Q1 stands out for its comprehensive understanding of Arabic language construction and detailed explanations, showcasing the model's ability to capture text nuances. Q2 provides accurate word meanings with slightly less detail than Q1. Despite consistent word identification in Q3, the translations lack contextual depth, indicating a potential improvement area. Q4 maintains high accuracy in word meanings but falls short in contextual nuances, yet it conveys the overall sentence context adequately. Q5 highlights the need for improved punctuation accuracy due to minor errors in word interpretation. Consistency was key in the study, with each query repeated three times. Q1, Q2, and Q4 exhibited consistent answers across repetitions, demonstrating the model's reliability. Q5 maintained consistent answers despite presentation variations, while Q3 provided identical answers in different positions, underscoring Gemini's overall reliability.

Play in Semantics

In explaining the three verses from the Nadham Al-Imrithi text, Gemini tends to provide fairly literal translations but occasionally includes practical examples. Gemini's approach focuses on structural and semantic analysis, offering direct translations from Arabic to Indonesian along with detailed explanations of the meaning and context of each verse. These translations include a breakdown of each word and concept found within the verses, explaining their subjective meanings and how they interact within the sentence. This approach aims to ensure a comprehensive understanding of each element, from individual words to overarching themes. However, there is variation in the quality and depth of the explanations provided by Gemini for each topic or query. While some explanations are generally informative and helpful, further evaluation by human experts is necessary to assess the true depth of understanding.

Table 2. for Gemini's Contextual Interpretation

(Q)	Collect	Measure	Repeat
Q6	The translation and explanation are systematically presented, providing accurate identification of key elements. Identifying thematic elements and linguistic nuances	Adequately grasps the meaning and messages conveyed in Arabic verses. Lacking specific contextual or detailed explanations.	Tasks repeated three times, yielding consistent literal interpretations without delving into intent or terminology.
Q7	Systematic approach emphasizing word-for-word comprehension and identification of key components.	Offers sufficient understanding of the verses' meanings with varied levels of contextual detail and emphasis.	Consistently provides literal translations without exploring intent or terminology, replicated in three tasks.
Q8	Follows a systematic approach similar to previous queries. Focusing on word-for-word comprehension and key element identification.	Exhibits a lack of deep understanding, meaningful interpretation, or coherence with previous verses. Highlighting the need for interconnection with preceding verses.	Consistently offers responses aligned with literal interpretation. Lacking exploration of intent and terminology, repeated in three tasks.

Gemini AI model tends to lean towards literal translation, occasionally supplementing translations with practical examples to facilitate a better understanding of the intended message. The approach involves a meticulous examination of the structure and significance of the verses, accompanied by supplementary explanations aimed at enriching comprehension of Arabic grammar and the cultural context surrounding the verses. Gemini's goal is to foster a comprehensive understanding by elucidating the contribution of each word and phrase to the overall meaning of the verse. Through this methodical approach, Gemini strives to enhance the reader's grasp of the intricate nuances embedded within the text, thereby facilitating a deeper appreciation and interpretation of classical Arabic literature.

In Query 6, Gemini was tasked with translating and explaining specific Arabic verses. The process was systematically approached, ensuring that key elements were accurately identified, allowing for a clear understanding of the thematic components and linguistic nuances within the verses. The translation provided by the model was adequate in grasping the general meanings and messages conveyed in the Arabic text. However, the explanations lacked specific contextual details and in-depth interpretations. The responses were consistent across three repetitions, indicating reliable performance in literal translation but limited engagement with the underlying intent or specialized

terminology. In Query 7, the model maintained a systematic approach with a strong emphasis on word-for-word comprehension and the identification of key components within the text. This approach ensured that the translations provided a sufficient understanding of the verses' meanings, albeit with varying levels of contextual detail and emphasis. The model consistently delivered literal translations, focusing on the direct meanings of words and phrases without exploring the broader context or the intent behind the text. In Query 8, the AI model followed a systematic approach similar to the previous queries, focusing on word-for-word comprehension and the identification of key elements within the verses. However, the analysis revealed a notable deficiency in the model's ability to provide a deep understanding or meaningful interpretation of the text. The translations lacked coherence with preceding verses, highlighting the model's struggle to establish an interconnected narrative across multiple verses.

Discussion

The evaluation conducted by experts on Gemini's ability to interpret verses from Alfiyah Ibn Malik and Nadham Imrithy revealed consistent yet contextually inadequate translations at times. The experts focused on translation accuracy, clarity of explanations, and depth of understanding of the presented Arabic literary context. While the translations showed consistency, they occasionally lacked contextual relevance. Semantic evaluations indicated a better understanding of the discussed themes, with some translations providing more relevant explanations. However, the tendency towards overly literal translations and insufficient contextual consideration was noted across evaluations, highlighting the need for improvement in depth and detail.

Verifier 1 observed that while Gemini's translations were consistent, they sometimes lacked contextual suitability. The translations tended to be cautious and included references, albeit with similar errors related to contextual appropriateness. Semantic evaluation showed better understanding of the discussed themes, resulting in more relevant explanations. Verifier 2 similarly found that Gemini tended towards overly literal translations, lacking context in responses. Semantic evaluations indicated a better understanding of the discussed themes, albeit somewhat lacking in depth and detail. Verifier 3 noted consistency in translations but occasional contextual mismatch. Semantic evaluations revealed a better understanding of the discussed themes, with more accurate explanations. However, all test models required more detailed explanations to expand understanding, not just focusing on the given verses but also on subsequent verses crucial for understanding the context and paradigms within.

In the literal translation results using Gemini, translations were accompanied by additional interpretations regarding Arabic language comprehension and its relevance to *Nahw* studies, incorporating various available online references. Despite being more concise, these translations provided a clear overview of the source book from which the excerpts were taken and how the words in the excerpts could be projected and analysed from specific reference sources. Due to accessing unchanged reference sources, each repetition of questions or tasks in the prompt provided the same answer, with only one or two words or sentences differing in the same context, or even providing identical answers. The accuracy of word-for-word translation using Gemini was assessed by the researcher with a score of 8 out of 10, with very few translation errors and adjustments made according to context without the need for "explanation" tasks. Regarding contextual adaptation, although the answers were brief, contextual alignment was always accurate or close to correct.

Table 3. Human Expert Evaluation Metrics and Assessment Results

No.	Variable	Effort	Emphasis Level	Validation Level
-----	----------	--------	----------------	------------------

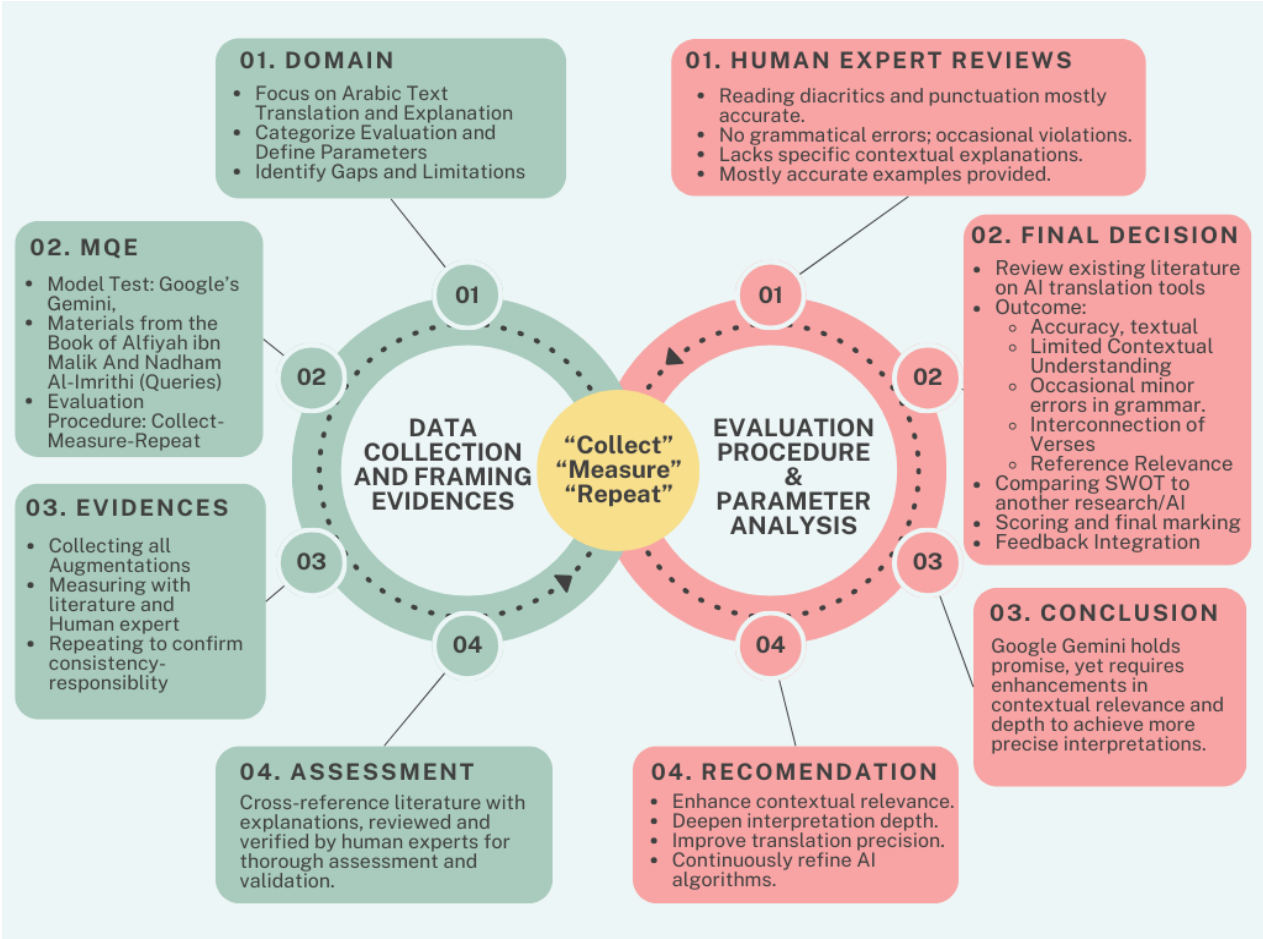
1	Accuracy of letter matching in answers	Imperfect matching Found 1 errors	Varies (rare errors)	Literal alignment, no deep meaning exploration
2	Reading of diacritics and punctuation	Mostly accurate No errors	Interpretation emphasized (some errors)	Literal interpretation, no meaning exploration
3	Spelling and Grammar	Some violations No errors present	Variable (rare occurrences) Slight attention	Seldom occurring, but present grammar errors
4	Internal Contextual Deepening	Partial understanding of context	Context understanding emphasized (incomplete)	Limited understanding of context
5	Contextual Significance	Lacks specific contextual explanations	Varies in detail and emphasis	Aligns with literal meaning; lacks deep exploration of intent
6	Provision of accurate examples	Mostly accurate examples, some inaccuracies in contextual examples	Variable accuracy (mostly accurate)	Accurate, but incomplete provision of examples
7	Understanding of interconnection with other verses	Limited understanding of interconnection	Interconnection emphasized (limited)	Limited understanding of interconnection
8	Reference Relevance	Lacks relevant contextual references	Inconsistent use of references	Relies on literal meaning without deeper contextual relevance

The testing results, as referenced in Inel et al. (2023)'s investigation into Responsible AI, demonstrate that AI models like Gemini adopt a typical and systematic approach in translation and explanation. Experts evaluated Gemini based on eight key parameters. The first parameter, accuracy of letter matching in answers, revealed rare errors with occasional discrepancies. The second, reading of diacritics and punctuation, highlighted mostly accurate interpretations with minimal errors. Third, grammar violations and errors were identified, although they were rare but significant enough to impact translation quality. The fourth parameter, internal contextual deepening, was found to be partial and incomplete. Fifth, external semantics showed that translations aligned with literal meanings but lacked deeper contextual insights. The sixth parameter, provision of accurate examples, exhibited mostly accurate examples but showed inaccuracies in contextual examples. Seventh, understanding of interconnection with other verses demonstrated limited comprehension and inadequate interconnection. Finally, the eighth parameter, reference relevance, showed a lack of consistency and deeper relevance, often relying on literal meanings. This evaluation indicates that while Gemini can produce relatively consistent translations, there are several areas requiring improvement,

particularly in terms of contextual relevance and completeness of explanations (Inel et al., 2023).

While the translations demonstrated consistency in literal alignment, they frequently lacked contextual relevance and depth. Semantic evaluations revealed varying degrees of understanding of the discussed themes, with Gemini often providing overly literal translations and insufficient consideration of contextual nuances. This highlights the need for the AI to improve its ability to deliver more nuanced and contextually appropriate translations, particularly for complex literary texts. In terms of lexical semantics, Gemini performed well in providing precise word-for-word translations. However, the model struggled with subjective interpretations and capturing the interconnections of meanings across different verses and necessary references, as indicated in the evaluation table. These insights underscore the importance of enhancing Google Gemini's contextual understanding and interpretation depth to better handle complex literary texts like *Alfiyah Ibn Malik* and *Nadham Imrithy*. Future refinements should focus on improving the AI's ability to capture subjective nuances, intertextual references, and deeper semantic connections crucial for accurate and comprehensive translations.

Figure 1. Execution and Analysis Roadmap



Nevertheless, although AI translation models like Google Gemini face challenges in thoroughly analyzing classical Arabic literature, they represent promising avenues for language learning (Ram & Verma, 2023) with continuous enhancements (Gemini Team et al., 2024). While not as straightforward as ChatGPT, Gemini is an evolving asset for the future (Ahmed et al., 2023). Thus, models like Gemini must further refine their capabilities

to accurately interpret language nuances (Akter et al., 2023; Kishore & Shaik, 2024), and delve into the hidden meanings present in texts such as *Alfiyah Ibn Malik* and *Nadham al-Imrithi*. Despite the complexity of Arabic literature, technological advancements offer opportunities for improving translation accuracy and contextual understanding as systematically framed by Killman, (2023), focusing on machine translation and data-driven approaches to contextual accuracy. To address these challenges effectively, ongoing research and development efforts are crucial. Integration of advanced linguistic algorithms and contextual analysis tools can enhance the performance of Generative AI augmentations in capturing the richness and intricacy of Arabic literature as offered by Dalayli (2023); Khoshafah, 2023; and Lu et al., (2023) to ChatGPT. Additionally, collaboration between AI developers, linguists, and educators can facilitate the creation of more robust translation models tailored to the specific needs of language learners (Huang et al., 2023). By leveraging the capabilities of AI technology while acknowledging its limitations, stakeholders can collectively work towards advancing language learning experiences and fostering a deeper appreciation for the linguistic and cultural heritage embedded in classical Arabic literature.

CONCLUSION

Following a comprehensive evaluation of Google Gemini's performance in translating and interpreting classical Arabic texts such as *Alfiyah Ibn Malik* and *Nadham al-Imrithi*, the study found that while the platform consistently produces accurate literal translations, it often struggles with contextual relevance, thematic connections across verses, and depth of explanation. Human expert assessments noted occasional spelling and grammar errors, incomplete examples, and limited grasp of broader interpretive layers, underscoring the need for refinement. To enhance Gemini's effectiveness, future development should focus on strengthening contextual understanding, expanding reference databases of classical Arabic texts with greater syntactic and morphological detail, and integrating scholarly resources to enrich cultural and semantic interpretation. Additionally, embedding Gemini within educational platforms and designing specialized learning modules could broaden its role in supporting Arabic language education. Despite current limitations, Gemini demonstrates significant potential as a supportive tool for understanding classical Arabic literature, and continuous improvement will be essential to harness AI's capacity for advancing translation quality and educational access. Future research should investigate comparative evaluations across different AI models, explore pragmatic and cultural semantics more deeply, and develop interdisciplinary frameworks that integrate linguistic, pedagogical, and technological perspectives.

REFERENCES

- Ahmed, I., Kajol, M., Hasan, U., & Datta, P. P. (2023). *ChatGPT vs. Bard: A Comparative Study* [Preprint].
<https://doi.org/10.36227/techrxiv.23536290.v1>
- Ajao, E. (2023). *Benefits, risks of Google AI chatbot's Bard Extensions* | TechTarget. Enterprise AI.
<https://www.techtarget.com/searchenterpriseai/news/366552898/Benefits-risks-of-Google-AI-chatbots-Bard-Extensions>
- Akter, S. N., Yu, Z., Muhamed, A., Ou, T., Bäuerle, A., Cabrera, Á. A., Dholakia, K., Xiong, C., & Neubig, G. (2023). An In-depth Look at Gemini's Language Abilities. *arXiv Preprint arXiv:2312.11444*.
<https://arxiv.org/abs/2312.11444>

- Almelhes, S. A. (2023). A Review of Artificial Intelligence Adoption in Second-Language Learning. *Theory and Practice in Language Studies*, 13(5), 1259–1269. <https://doi.org/10.17507/tpls.1305.21>
- Alruqi, T. N., & Alzahrani, S. M. (2023). Evaluation of an Arabic Chatbot Based on Extractive Question-Answering Transfer Learning and Language Transformers. *AI*, 4(3), 667–691. <https://doi.org/10.3390/ai4030035>
- Annamalai, N., Rashid, R. A., Munir Hashmi, U., Mohamed, M., Harb Alqaryouti, M., & Eddin Sadeq, A. (2023). Using chatbots for English language learning in higher education. *Computers and Education: Artificial Intelligence*, 5, 100153. <https://doi.org/10.1016/j.caeai.2023.100153>
- Atwell, E. S., Brierley, C., Dukes, K., Sawalha, M., & Sharaf, A. (2011). *An artificial intelligence approach to Arabic and Islamic content on the internet* [Proceedings Paper]. Leeds. NITS 2011 3rd National Information Technology Symposium. Proceedings of NITS 3rd National Information Technology Symposium. <http://www.kokannews.org/wp-content/uploads/2014/05/An-Artificial-Intelligence.pdf>
- Azmi, A. M., Al-Qabbany, A. O., & Hussain, A. (2019). Computational and natural language processing based studies of hadith literature: A survey. *Artificial Intelligence Review*, 52, 1369–1414.
- Bandi, A., Adapa, P. V. S. R., & Kuchi, Y. E. V. P. K. (2023). The Power of Generative AI: A Review of Requirements, Models, Input–Output Formats, Evaluation Metrics, and Challenges. *Future Internet*, 15(8), Article 8. <https://doi.org/10.3390/fi15080260>
- Cantú-Ortiz, F. J., Galeano Sánchez, N., Garrido, L., Terashima-Marin, H., & Brena, R. F. (2020). An artificial intelligence educational strategy for the digital transformation. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 14, 1195–1209. <https://doi.org/https://doi.org/10.1007/s12008-020-00702-8>
- Dalayli, F. (2023). Use of NLP Techniques in Translation by ChatGPT: Case Study. In A. H. Haddad, A. R. Terryn, R. Mitkov, R. Rapp, P. Zweigenbaum, & S. Sharoff (Eds.), *Proceedings of the Workshop on Computational Terminology in NLP and Translation Studies (ConTeNTS) Incorporating the 16th Workshop on Building and Using Comparable Corpora (BUCC)* (pp. 19–25). INCOMA Ltd., Shoumen, Bulgaria. <https://aclanthology.org/2023.contents-1.3>
- Feng, Z., Zhang, Y., Li, H., Liu, W., Lang, J., Feng, Y., Wu, J., & Liu, Z. (2024). Improving LLM-based Machine Translation with Systematic Self-Correction. *arXiv Preprint arXiv:2402.16379*. <https://arxiv.org/pdf/2402.16379>
- Fuad, A., & Al-Yahya, M. (2022). Recent Developments in Arabic Conversational AI: A Literature Review. *IEEE Access*, 10, 23842–23859. <https://doi.org/10.1109/ACCESS.2022.3155521>
- Garrido, N. (2017). The method of James Spradley in qualitative research. *Enfermería: Cuidados Humanizados*, 6(SPE), 37–42.
- Gemini Team, Reid, M., Savinov, N., Teplyashin, D., Dmitry, Lepikhin, Lillicrap, T., Alayrac, J., Soricut, R., Lazaridou, A., Firat, O., Schrittwieser, J., Antonoglou, I., Anil, R., Borgeaud, S., Dai, A., Millican, K., Dyer, E., Glaese, M., ... Vinyals, O. (2024). *Gemini 1.5: Unlocking multimodal understanding across millions of*

- tokens of context (arXiv:2403.05530). arXiv. <https://doi.org/10.48550/arXiv.2403.05530>
- Hajir, B., Rasman, R., & McNerney, W. (2022). Digital translanguaging and Arabic-English transliteration (Arabizi): Insights from Syria and Lebanon. *Cambridge Educational Research E-Journal*, 9.
- Haq, Y. N. (2022). *Manhaj al-Imam Ibn Aqil fi Syarhi Alfiah al-Imam Ibn Malik* [bachelorThesis, Fakultas Dirasat Islamiah]. <https://repository.uinjkt.ac.id/dspace/handle/123456789/61799>
- Hidayatullah, A. S., & Fauji, I. (2023). *Bridging Theory and Practice in Arabic Language Education / Indonesian Journal of Islamic Studies*. <https://ijis.umsida.ac.id/index.php/ijis/article/view/1724>
- Huang, X., Zou, D., Cheng, G., Chen, X., & Xie, H. (2023). Trends, Research Issues and Applications of Artificial Intelligence in Language Education. *Educational Technology & Society*, 26(1), 112–131. <https://www.jstor.org/stable/48707971>
- Imran, M., & Almusharraf, N. (2024). Google Gemini as a next generation AI educational tool: A review of emerging educational technology. *Smart Learning Environments*, 11(1), 22. <https://doi.org/10.1186/s40561-024-00310-z>
- Inel, O., Draws, T., & Aroyo, L. (2023). Collect, Measure, Repeat: Reliability Factors for Responsible AI Data Collection. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, 11(1), Article 1. <https://doi.org/10.1609/hcomp.v11i1.27547>
- Khoshafah, F. (2023). ChatGPT for Arabic-English Translation: Evaluating the Accuracy. *Ministry of Education, Yemen*. <https://doi.org/10.21203/rs.3.rs-2814154/v2>
- Killman, J. (2023). Machine translation and legal terminology: Data-driven approaches to contextual accuracy. In *Handbook of Terminology* (pp. 485–510). John Benjamins. <https://www.jbe-platform.com/content/books/9789027249388-hot.3.mac2>
- Kishore, K. S., & Shaik, R. (2024). *Evaluating Telugu Proficiency in Large Language Models_ A Comparative Analysis of ChatGPT and Gemini* (arXiv:2404.19369). arXiv. <https://doi.org/10.48550/arXiv.2404.19369>
- Kocoń, J., Cichecki, I., Kaszyca, O., Kochanek, M., Szydło, D., Baran, J., Bielaniewicz, J., Gruza, M., Janz, A., Kanclerz, K., Kocoń, A., Koptyra, B., Mieleszczenko-Kowszewicz, W., Miłkowski, P., Oleksy, M., Piasecki, M., Radliński, Ł., Wojtasik, K., Woźniak, S., & Kazienko, P. (2023). ChatGPT: Jack of all trades, master of none. *Information Fusion*, 99, 101861. <https://doi.org/10.1016/j.inffus.2023.101861>
- Kolhar, M., & Alameen, A. (2021). Artificial Intelligence Based Language Translation Platform. *Intelligent Automation & Soft Computing*, 28(1). <https://doi.org/http://dx.doi.org/10.32604/iasc.2021.014995>
- Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: Systematic literature review. *International Journal of Educational Technology in Higher Education*, 20(1), 56. <https://doi.org/10.1186/s41239-023-00426-1>

- Läubli, S., Castilho, S., Neubig, G., Sennrich, R., Shen, Q., & Toral, A. (2020). A set of recommendations for assessing human-machine parity in language translation. *Journal of Artificial Intelligence Research*, 67, 653–672. <https://doi.org/https://doi.org/10.1613/jair.1.11371>
- Lu, Q., Qiu, B., Ding, L., Xie, L., & Tao, D. (2023). *Error Analysis Prompting Enables Human-Like Translation Evaluation in Large Language Models: A Case Study on ChatGPT* (No. 2023030255). Preprints. <https://doi.org/10.20944/preprints202303.0255.v1>
- Ram, B., & Verma, P. (2023). Artificial intelligence AI-based Chatbot study of ChatGPT, Google AI Bard and Baidu AI. *World Journal of Advanced Engineering Technology and Sciences*, 8(1), 258–261. <https://doi.org/10.30574/wjaets.2023.8.1.0045>
- Tiawaldi, A., & Abdul Wahab, M. (2017). PERKEMBANGAN BAHASA ARAB MODERN DALAM PERSPEKTIF SINTAKSIS DAN SEMANTIK PADA MAJALAH ALJAZEERA. *Arabiyat : Jurnal Pendidikan Bahasa Arab Dan Kebahasaaraban*, 4(1). <https://doi.org/10.15408/a.v4i1.5328>
- Zaimah, N. R., Fatchiatuzahro, & Budi Hartanto, E. (2024). ENHANCING WRITING COMPREHENSION IN L2 ARABIC LEARNERS THROUGH AI-BASED TRANSLANGUAGING CHATBOTS. *Al-Mubin: Islamic Scientific Journal*, 7(1), 21–34. <https://doi.org/10.51192/almubin.v7i1.753>
- Zaimah, N. R., Hartanto, E. B., & Zahro, F. (2024). Acceptability and Effectiveness Analysis of Large Language Model-Based Artificial Intelligence Chatbot Among Arabic Learners. *Mantiqu Tayr: Journal of Arabic Language*, 4(1), Article 1. <https://doi.org/10.25217/mantiquatayr.v4i1.3951>