Ta'lim al-'Arabiyyah: Jurnal Pendidikan Bahasa Arab dan Kebahasaaraban
DOI: https://doi.org/10.15575/jpba.v9i2.51510

Ta'lim al-'Arabiyyah
JURNAL PENDIDIKAN BAHASA ARAB & KEBAHASAARABAN
P-ISSN: 2721-5237 | E-ISSN: 2549-208X // Vol. 9 No. 2 | 207-223
https://journal.uinsgd.ac.id/index.php/Talim/index

INDONESIAN–ARABIC ACADEMIC TRANSLATION QUALITY: A COMPARATIVE CONTENT ANALYSIS OF CHATGPT AND GOOGLE TRANSLATE

Alfitri*[1], Gamal Abdul Nasir Zakaria[2], Misran[3], Talqis Nurdianto[4]
[1]Universitas Islam Riau, Indonesia
[2]Universiti Brunei Darussalam, Brunei Darussalam
[3]Politeknik Pariwisata NHI Bandung, Indonesia
[4]Universitas Muhammadiyah Yogyakarta, Indonesia

Article Information

Article History:

Received	:	20-October-2025
Revised	:	14-December-2025
Accepted	:	19-December-2025
Published	:	25-December-2025

Keywords:
Arabic Language;
ChatGPT;
Google Translate;
Indonesian-Arabic Translation;
Machine Translation;
Translation Quality Evaluation.

Articles Available Online:

ABSTRACT

This study compares the quality of Indonesian–Arabic academic translations produced by Google Translate and ChatGPT (GPT-4), a topic rarely examined despite widespread MT use in Arabic Language Education programs. Using a qualitative descriptive design with content analysis, the data comprised ten purposively selected undergraduate thesis titles from Indonesian university repositories, including UIN Sunan Kalijaga, UIN Suska Riau, UIN Imam Bonjol Padang, and Universitas Muhammadiyah Makassar. Titles were selected for their academic rigor and need for conceptual precision and formal Arabic register. Translations were analyzed at the phrase level using a Hybrid MQM–Nababan–Baker rubric encompassing seven dimensions: accuracy, acceptability, readability, lexical equivalence, grammatical equivalence, cohesion and coherence, and academic fluency, each rated on a three-point scale. Validity was ensured through alignment with Arabic translation theory, equivalence frameworks, and MQM standards, while iterative consistency checks supported reliability. Results reveal clear differences. ChatGPT achieved an average score of 20.0 out of 21 points (93%), which is classified as Very Good and indicates strong suitability for academic publication. Google Translate scored 13.3 out of 21 points (63%), classified as Good, but requiring post-editing. ChatGPT excelled in contextual meaning, syntactic restructuring, accurate *idhafah*, case governance, consistent terminology, and scholarly style, whereas Google Translate showed literal transfer. This study enriches AI-assisted translation discourse by grounding evaluation in Arabic translation theory and pedagogy, emphasizing generative AI's pedagogical potential as a complementary tool while reaffirming the indispensable role of human expertise in maintaining linguistic accuracy, rigor, and academic standards.

CORRESPONDING AUTHOR:
Alfitri,
Department of Arabic Language Education,
Universitas Islam Riau,
Jl. Kaharudin Nasution No. 113, Marpoyan Pekanbaru, Riau, Indonesia.
Email: alfitri2018@fis.uir.ac.id

How to Cite:
Alfitri, A., Abdul Nasir Zakaria, G., Misran, M., & Nurdianto, T. (2025). Indonesian–Arabic Academic Translation Quality: A Comparative Content Analysis of ChatGPT and Google Translate. *Ta'lim al-'Arabiyyah: Jurnal Pendidikan Bahasa Arab & Kebahasaaraban, 9*(2), 207–223. https://doi.org/10.15575/jpba.v9i2.51510

INTRODUCTION

The rapid advancement of artificial intelligence (AI) has reshaped contemporary translation practices and foreign language learning, including in Arabic language education (Syafanah et al., 2025). Over the past decade, machine translation (MT) systems such as Google Translate have become widely accessible and increasingly relied upon for multilingual communication and academic work (Jooste et al., 2021). The emergence of generative AI models, most notably ChatGPT, built upon Large Language Models (LLMs), marks a new phase in MT development, offering context-aware processing and improved semantic and pragmatic interpretation (Ataman et al., 2025). These technological shifts have prompted renewed attention to the reliability of MT outputs, particularly for languages with complex morphology and diglossic features such as Arabic (Alayba, 2025).

In Arabic Language Education programs, MT tools are now used extensively by students to complete coursework, translate academic texts, and comprehend both classical and modern Arabic sources (Hakiki et al., 2023). While such tools enhance accessibility and learning efficiency, concerns persist about the accuracy, naturalness, and terminological consistency of MT outputs, particularly when used for academic purposes (El-Farahaty, 2025). These concerns highlight a broader pedagogical challenge: the increasing dependence on AI-based translation tools is not matched by adequate empirical studies evaluating their quality in the Indonesian-Arabic language pair (Nurfaiza, 2024).

Several previous studies have discussed ChatGPT and Google Translate in the context of language. Nasaruddin (2024) notes that ChatGPT supports Arabic language teachers in the learning process by providing rich, modern, and engaging educational materials, including texts, questions, and images. Then, Chandra et al. (2025) evaluate Google Translate's performance in preserving sentiment and semantics. However, these studies focus on translation accuracy in other fields and have not highlighted the context of academic translation in higher education. In addition, several previous comparative studies, such as Al-Darabee et al. (2025), which compared Netflix and Google Translate translations, focused more on English-Arabic translations and did not involve Indonesian as the source language. Furthermore, studies on MT quality in Indonesia tend to use Nababan (2012) accuracy–acceptability–readability model, but without situating Indonesian-Arabic MT evaluation within global translation quality frameworks such as MQM or equivalence theory.

Thus, a clear *research gap* emerges: no study has systematically compared ChatGPT and Google Translate in translating Indonesian academic texts into Arabic using an integrated linguistic and translation-quality framework. This gap is academically significant for two reasons. First, Indonesian-Arabic translation presents unique linguistic challenges due to divergent syntactic structures, morphological systems, and rhetorical traditions, factors that strongly influence MT performance (Popović, 2020). Second, MT systems used by students operate on distinct algorithmic principles: Google Translate relies on neural statistical models, whereas ChatGPT utilizes contextual generative AI with advanced semantic inference. Evaluating how these differing architectures influence translation quality is crucial for understanding the pedagogical implications of MT use in Arabic language education.

Accordingly, this study aims to compare the quality of Indonesian–Arabic academic translation produced by ChatGPT and Google Translate, using a content analysis approach and a hybrid assessment model that integrates MQM, Nababan's criteria, and Baker's equivalence framework. The study addresses the following problem statement: How do

ChatGPT and Google Translate differ in producing accurate, natural, and academically appropriate translations of Indonesian academic texts into Arabic?.

The novelty of this study lies in its methodological integration of global and national translation-quality models, as well as its focus on the underexamined Indonesian–Arabic academic translation context. The findings are expected to contribute theoretically to the growing body of research on AI-assisted translation and, practically, to Arabic translation pedagogy, particularly by informing the responsible and pedagogically sound use of MT tools in higher education.

METHOD

This study adopts a qualitative descriptive design with a content analysis approach to examine the quality of Indonesian-Arabic academic translations produced by two machine translation systems: Google Translate and ChatGPT. This design is appropriate for identifying linguistic patterns, meaning shifts, and stylistic features within translated texts.

The research data consist of 10 Indonesian thesis titles purposively selected from undergraduate theses in Arabic Language Education (PBA) programs at several Indonesian universities (UIN Sunan Kalijaga, UIN Suska Riau, UIN Imam Bonjol Padang, Universitas Muhammadiyah Makassar, and others). Thesis titles were selected because they represent academically curated text units that reflect conceptual precision, terminological accuracy, and formal academic phrasing.

The unit of analysis is the Arabic translation output generated by Google Translate and ChatGPT (GPT-4 model). Each title was translated separately by the two systems to enable a point-by-point comparative evaluation. Data collection was conducted in three steps: (1) Compiling Indonesian thesis titles from institutional repositories, (2) Translating each title independently using Google Translate and ChatGPT without manual intervention, and (3) Organizing outputs into a comparative dataset for analysis.

Translation quality was assessed using a Hybrid MQM-Nababan-Baker rubric that integrates global and national evaluation dimensions. The instrument includes seven analytical dimensions: accuracy, acceptability, readability, lexical equivalence, grammatical equivalence, cohesion and coherence, fluency, and academic style. Each dimension is scored on a 3-point scale (1-3), with higher scores indicating stronger translation quality.

Content validity was ensured by aligning the rubric indicators with established theoretical frameworks in translation studies (Nababan, Baker, and MQM). Reliability was addressed through repeated evaluation cycles and cross-checking of scoring consistency. The analysis process was conducted iteratively until stable patterns emerged across the dataset. The analysis followed the qualitative stages proposed by Miles & Huberman (1992): (1) Data Reduction: identifying linguistic errors, semantic shifts, and deviations from Arabic academic norms, (2) Data Display: presenting source texts, translation outputs, scores, and analytical notes in comparative tables, and (3) Conclusion Drawing: synthesizing evaluative patterns to determine the relative strengths and weaknesses of each machine translation system.

This study is non-experimental and comparative, analyzing existing translation outputs to reveal qualitative differences attributable to the systems' underlying architectures rather than controlled manipulation. The researcher served as the primary instrument in the analysis, supported by an evaluation rubric validated against the theoretical frameworks of Nababan, Baker, and the MQM model (Lommel et al., 2014). This evaluation rubric assessed

translation quality across two key aspects: (1) assessment dimensions and (2) scoring criteria. The first aspect is illustrated in the following table:

Table 1. Assessment Dimensions

No.	Assessment Dimensions	Sub-Aspects	Ideal Performance Description	Score (1–3)
1	Accuracy	Semantic equivalence	The translation conveys the meaning of the source text completely and accurately, without loss or distortion.	1–3
		Misinterpretation	There is no shift or addition of meaning that alters the message's content.	1–3
2	Acceptability	Arabic language norms	Sentence structure, diction, and style are consistent with academic Arabic language norms.	1–3
		Natural expression	No machine-like feel; sentences flow naturally and idiomatically.	1–3
3	Readability	Level of comprehensibility	Academic Arabic readers easily understand sentences without causing ambiguity.	1–3
4	Lexical Equivalence	Equivalence of academic terms	Scientific terms are translated accurately and consistently, in line with the scientific field's context.	1–3
		Translation of idioms and collocations	Idiomatic expressions are translated with natural equivalents in Arabic.	1–3
5	Grammatical Equivalence	Sentence structure and syntactic relationships	Word structure and order follow Arabic rules without losing the original meaning.	1–3
		Morphology and word type compatibility	Word form selection (*isim–fi'il–harf*) according to function in Arabic sentences.	1–3
		Inter-sentence integration	The logical relationship between sentences and	1–3

No.	Assessment Dimensions	Sub-Aspects	Ideal Performance Description	Score (1–3)
6	Cohesion & Coherence		paragraphs is well maintained in the translation.	
		Use of conjunctions and pronouns	Discourse markers are appropriately used in Arabic.	1–3
7	Fluency & Style (MQM)	Academic style appropriateness	The academic style of Arabic is well reflected (formal, objective, clear).	1–3
		Technical machine errors (noise, literalism)	There are no traces of literalism, stiffness, or machine-like literal translations.	1–3

As for the scoring criteria, they can be observed in the following table:

Table 2. Scoring Criteria

Total Score Range	Score Interpretation
17–21	**Very Good (Highly Reliable Translation)** – Suitable for academic publication without significant revisions.
12–16	**Good (Acceptable Translation)** - Fairly good; requires minor editing.
7–11	**Moderate (Partially Acceptable)** - There are still inaccuracies in meaning and a lack of naturalness.
≤6	**Poor (Inaccurate / Machine-Literal)** - The translation is rigid and not academically acceptable.

Based on Table 2, the scoring criteria are designed to classify translation quality according to the total score obtained across all assessed dimensions. A total score ranging from 17 to 21 indicates *Very Good* translation quality, reflecting high reliability and suitability for academic publication without requiring significant revision. Scores between 12 and 16 are categorized as *Good*, suggesting that the translation is generally acceptable but still requires minor editing to improve accuracy or naturalness. A score of 7 to 11 represents *Moderate* quality, indicating partial acceptability with noticeable inaccuracies in meaning and limited linguistic naturalness. Finally, translations scoring six or below are classified as *Poor*, characterized by rigid, machine-literal rendering that fails to meet academic standards.

RESULT AND DISCUSSION

After collecting the original academic texts in the form of Indonesian-language thesis titles from ten Arabic Language Education (PBA) programs at several universities in Indonesia, the texts were then translated into Arabic using both Google Translate and ChatGPT. The translation results are presented in the following table:

Table 3. Arabic Translation Results by Google Translate and ChatGPT

No.	Source of Text	Type of Text	Original Text (Indonesian)	Google Translate Result (Arabic)	Chatgpt Result (Arabic)
1	Thesis – PBA IAIN Mataram (2011)	Title	Pengaruh kelancaran membaca Al-Qur'an terhadap keterampilan membaca pada bidang studi bahasa Arab kelas VIII di MTs Yusuf Abdussatar Kediri Lombok Barat Tahun Pelajaran 2011/2012	تأثير الطلاقة في قراءة القرآن الكريم على مهارات القراءة في مجال دراسات اللغة العربية للصف الثامن في مدرسة يوسف عبد الستار كيديري غرب لومبوك في العام الدراسي ٢٠١٢/٢٠١١	تأثير طلاقة قراءة القرآن الكريم على مهارة القراءة في مادة اللغة العربية لدى طلاب الصف الثامن بمدرسة يوسف عبد الستار المتوسطة بكَدِري - لومبوك الغربية في العام الدراسي ٢٠١٢/٢٠١١م
2	Thesis – PBA IAIN Parepare (2019)	Title	Analisis kesulitan belajar bahasa Arab peserta didik kelas VII MTs DDI WANIO terhadap penguasaan kosakata bahasa Arab	تحليل صعوبات تعلم اللغة العربية لدى طلاب الصف السابع في مدرسة MTs DDI WANIO فيما يتعلق بإتقان المفردات العربية	تحليل صعوبات تعلم اللغة العربية لدى طلاب الصف السابع بمدرسة DDI وانيو المتوسطة في إتقان المفردات العربية
3	Thesis – PBA UIN SUKA Yogyakarta (2009)	Title	Problematika pengajaran kemahiran membaca bahasa Arab siswa MTsN Laboratorium UIN Sunan Kalijaga Yogyakarta	مشكلات تدريس مهارات القراءة باللغة العربية للطلاب في مختبر MTsN UIN Sunan Kalijaga Yogyakarta	إشكالية تعليم مهارة القراءة في اللغة العربية لدى طلاب المدرسة المتوسطة الحكومية النموذجية بجامعة سونان كاليجاغا الإسلامية الحكومية بيوجياكرتا

No.	Source of Text	Type of Text	Original Text (Indonesian)	Google Translate Result (Arabic)	Chatgpt Result (Arabic)
4	Thesis – PBA UIN SUSKA Pekanbaru (2021)	Title	Pengaruh aktivitas pembelajaran bahasa Arab di Pusat Bahasa terhadap kemampuan mengikuti kuliah Studi Teks Fikih di jurusan Pendidikan Agama Islam Fakultas Tarbiyah dan Keguruan UIN SUSKA Pekanbaru	تأثير أنشطة تعلم اللغة العربية في مركز اللغات على القدرة على متابعة مقرر دراسة النصوص الفقهية الإسلامية في قسم التربية الدينية الإسلامية بكلية التربية وتدريب المعلمين بجامعة سوسكا بيكانبارو	تأثير نشاطات تعلم اللغة العربية في مركز اللغات على قدرة متابعة محاضرات مادة دراسة النصوص الفقهية في قسم التربية الإسلامية بكلية التربية بجامعة السلطان شريف قاسم الإسلامية الحكومية ببيكانبارو
5	Thesis – PBA Universitas Muhammadiyah Makassar (2017)	Title	Stimulus guru dan respon siswa dalam pembelajaran bahasa Arab kelas VIII SMP UNISMUH Makassar	تحفيز المعلم واستجابة الطالب في تعلم اللغة العربية للصف الثامن في مدرسة SMP UNISMUH ماكاسار	المنبّهات التعليمية من المعلم واستجابات الطلاب في تعلم اللغة العربية لطلاب الصف الثامن بمدرسة UNISMUH المتوسطة بمكاسر
6	Thesis – PBA Universitas Jambi (2024)	Title	Implementasi metode pembelajaran Picture and Picture terhadap keterampilan membaca bahasa Arab di MTs Al-Jauharen kota Jambi	تطبيق أسلوب تعلم الصورة والصورة على مهارات القراءة باللغة العربية في مدرسة الجوهرين المتوسطة بمدينة جامبي	تطبيق طريقة التعلم Picture and Picture على مهارة القراءة في اللغة العربية بمدرسة الجوهرين المتوسطة بمدينة جمبي
7	Thesis – PBA Universitas Islam Riau Pekanbaru (2024)	Title	Strategi Sekolah Dalam Meningkatkan Minat Belajar Bahasa Arab Siswa Di SMP YLPI. P	استراتيجيات مدرسية لزيادة اهتمام الطلاب بتعلم اللغة العربية في مدرسة YLPI المتوسطة، بي ماربويان، بيكانبارو	استراتيجية المدرسة في تنمية اهتمام الطلاب بتعلم اللغة العربية في مدرسة YLPI المتوسطة بماربوين – بيكانبارو

No.	Source of Text	Type of Text	Original Text (Indonesian)	Google Translate Result (Arabic)	Chatgpt Result (Arabic)
			Marpoyan Pekanbaru		
8	Thesis – PBA UIN Imam Bonjol Padang (2021)	Title	Pengembangan alat tes Mufradat dengan menggunakan Wordwall untuk kelas 8 MTsN 1 Payakumbuh	تطوير أداة اختبار المفردات باستخدام Wordwall للفئة ٨ MTsN 1 Payakumbuh	تطوير أداة اختبار المفردات باستخدام منصة Wordwallلطلاب الصف الثامن بالمدرسة المتوسطة الإسلامية الحكومية الأولى بباياكمبوه
9	Thesis – PBA UIN Syarif Hidayatullah Jakarta (2015)	Title	Revitalisasi metodologi penelitian bahasa sebagai basis pengembangan pendidikan Bahasa Arab	إحياء منهجية البحث اللغوي كأساس لتطوير تعليم اللغة العربية	إعادة تفعيل منهجية البحث اللغوي بوصفها أساسًا لتطوير تعليم اللغة العربية
10	Thesis – UIN Maulana Malik Ibrahim Malang (2019)	Title	Metode pembelajaran keterampilan membaca di Madrasah Aliyah Al Ma`arif Sukorjo Pasuruan	أساليب تعلم مهارات القراءة في مدرسة المعارف سوكورجو باسوروان الإسلامية الثانوية	طريقة تدريس مهارة القراءة في المدرسة الثانوية الإسلامية العليا المعارف بسوكورجو – باسووران

The Indonesian thesis titles were deliberately selected because they constitute an essential part of a thesis. In addition to reflecting the aspects being studied, a title is intentionally crafted with precise, accurate, and scholarly diction and phrasing. This makes the thesis titles highly substantial in terms of both content and form.

The analysis of ten thesis titles reveals consistent differences in translation quality between Google Translate (GT) and ChatGPT. ChatGPT produces Arabic structures that are more natural, coherent, and aligned with academic norms, whereas GT tends to retain Indonesian syntactic patterns in a literal manner. This pattern is consistent with Popović's (2020) findings that conventional NMT systems often fail to adapt to linguistic structures in morphologically complex languages such as Arabic. Semantically, ChatGPT demonstrates greater success in capturing relational meaning and key concepts within academic texts. In nearly all cases, ChatGPT selects more appropriate academic Arabic diction, such as using *إشكالية* for "problematika" and *منصة Wordwall* for digital learning contexts. GT often translates lexically without considering disciplinary context.

These findings support Barker's equivalence theory, which posits that lexical and textual equivalence require contextual understanding, a task that LLM-based models like

ChatGPT handle more effectively. ChatGPT consistently generates *jumlah ismiyyah* and *jumlah fi'liyyah* constructions that adhere to standard Arabic grammatical norms (*nahwu–ṣarf*). In contrast, GT often preserves Indonesian word order, resulting in syntactic unnaturalness in Arabic, for example, inverted phrase sequences or incorrect *muḍāf–muḍāf ilayh* constructions.

This reinforces the critique that acceptability is determined by the translator's ability to adapt messages to target-language norms, an ability limited in GT but markedly stronger in ChatGPT. ChatGPT maintains more cohesive relationships between sentence components, including the use of appropriate conjunctions, pronouns, and logical thematic structures. GT frequently produces fragmented sentences that lack a coherent flow of information, thereby reducing readability.

These findings align closely with House's argument that translation quality in academic genres is shaped not only by semantic transfer but also by the management of informational structure and the clarity of inter-clausal relationships. In thesis titles containing longer nominal constructions and embedded modifiers, such as data 1, 4, and 5, the quality divergence between the two MT systems becomes increasingly visible. ChatGPT demonstrates stability in preserving intended meaning while restructuring Indonesian noun phrases into well-formed Arabic academic expressions that conform to disciplinary conventions.

In contrast, Google Translate tends to retain source-language surface structures, resulting in overly literal renderings, disrupted iḍāfah chains, weakened case governance, and reduced syntactic naturalness, all of which diminish scholarly acceptability in Arabic. Although Google Translate does not completely fail when processing shorter titles, its outputs still lack the terminological precision and syntactic fluidity that ChatGPT achieves through deeper contextual inference. Across the dataset, three recurring patterns emerge. First, ChatGPT shows consistent strength in linguistic and academic adaptation, particularly in phrase reordering, nominal agreement, and register alignment. Second, Google Translate operates predominantly through form-bound statistical processing, prioritizing structural resemblance over functional equivalence. Third, the errors produced by Google Translate are systemic rather than incidental, evidenced by persistent misgoverned compounds, unstable terminology, and repeated morpho-syntactic inconsistencies across all data points. The regularity of these error typologies signals that quality limitations stem from architectural constraints rather than random performance noise.

These findings substantiate Nida's concept of dynamic equivalence by showing that generative models, such as ChatGPT, are technologically better able to approximate meaning-oriented academic translation than conventional neural statistical MT systems. The results empirically confirm Chandra's 2025 conclusion that LLM-based translation yields higher academic reliability than traditional NMT engines like Google Translate. They also reinforce Al-Darabee's 2025 findings that Google Translate faces structural challenges in managing Arabic academic terminology and syntactic environments, even though those earlier studies centered on different language pairs. Moreover, this study extends Hakiki's 2023 observations originally examining Arabic-to-Indonesian MT, by demonstrating that ChatGPT maintains its superiority in the reverse direction, Indonesian-to-Arabic, when evaluated using an integrated framework combining MQM, Nababan's criteria, and Baker's equivalence model. By doing so, this study moves beyond prior field-general evaluations. It

offers corpus-specific evidence that differences in MT architecture significantly shape the semantic, grammatical, rhetorical, and stylistic quality of Indonesian–Arabic academic translation. Collectively, the study strengthens the theoretical claim that MT system design directly determines translational adequacy for Arabic academic discourse and confirms that generative MT can serve as a pedagogical scaffold, but not a replacement, for expert-driven academic post-editing and terminological standardization.

Discussion

Based on the data presented above, the translation results produced by Google Translate and ChatGPT were evaluated using the Hybrid MQM-Nababan-Baker instrument. The results of this assessment are shown in the following table:

Table 4. Scoring & Assessment

No	Text Source	Accuracy	Acceptability	Readability	Lexical Equivalence	Grammatical Equivalence	Cohesion & Coherence	Fluency & Academic Style	Total (21)	Category	Analytical Notes
1	IAIN Mataram (2011)	GT: 2/ GPT: 3	GT: 2/ GPT: 3	GT: 2/GPT: 3	GT: 2/GPT: 3	GT: 2/GPT: 3	GT: 2/GPT: 3	GT: 1/ GPT: 3	GT: 13 / GPT: 21	GT: Good / GPT: Very Good	GPT is more idiomatic and natural, GT is still literal
2	IAIN Parepare (2019)	2/3	2/3	2/3	2/3	2/3	2/3	1/3	13 / 21	Good/ Very Good	GPT correctly adapts the context of "وانيو"
3	UIN SUKA (2009)	2/3	2/3	2/3	2/3	2/3	2/3	2/3	14 / 21	Good/ Very Good	GPT uses the correct academic term "إشكالية."
4	UIN SUSKA (2021)	2/3	2/3	2/3	2/3	2/3	2/3	2/3	14 / 21	Good/ Very Good	GPT is more natural in construction " محاضرات مادة دراسة النصوص الفقهية"
5	UNISMUH Makassar (2017)	2/3	2/3	2/3	2/3	2/3	2/3	1/3	13 / 21	Good/ Very Good	GPT is more fluent and academic, GT is rigid
6	Univ. Jambi (2024)	2/3	2/3	2/3	2/3	2/3	2/3	2/3	14 / 21	Good/ Very Good	GPT captures contextual meaning "Picture and Picture."
7	Univ. Islam Riau (2024)	2/3	2/3	2/3	2/3	2/3	2/3	2/3	14 / 21	Good/ Very Good	GPT is more cohesive and natural
8	UIN Imam Bonjol (2021)	2/3	2/3	2/3	2/3	2/3	2/3	2/3	14 / 21	Good/ Very Good	GPT adds the idiomatic context "منصة Wordwall"
9	UIN Syarif Hidayatullah (2015)	02-Mar	02-Mar	02-Mar	02-Mar	02-Mar	02-Mar	02-Mar	14 / 21	Good/ Very Good	GPT chose the formal term " إعادة تفعيل"
10	UIN Maulana Malik Ibrahim (2019)	02-Mar	02-Mar	02-Mar	02-Mar	02-Mar	02-Mar	02-Mar	14 / 21	Good/ Very Good	GPT generates more formal and natural sentences

Quantitative Analysis of Translation Quality

Based on assessment results using the Hybrid MQM-Nababan-Baker instrument, the translation quality of academic texts produced by Google Translate (GT) and ChatGPT (GPT) differed significantly across all evaluation dimensions. The average total scores indicate that ChatGPT achieved 20.0 out of 21 points (93%), categorized as "Very Good", whereas Google Translate obtained an average score of 13.3 out of 21 points (63%), categorized as "Good".

This difference in scores demonstrates that ChatGPT produces translations that are more accurate, natural, and contextually appropriate for academic purposes. At the same time, Google Translate tends to generate more literal translations and often retains the syntactic structure of the source language (Indonesian). These findings are consistent with studies by Popović (2020) and Jooste et. al. (2021), which affirm that traditional neural machine translation (NMT) systems, such as Google Translate, still face limitations in handling syntax and semantics in morphologically complex languages like Arabic.

Qualitative Analysis Based on the Seven Assessment Dimensions

The accuracy dimension assesses the extent to which the translation preserves the original meaning of the source text. ChatGPT demonstrated high accuracy (average score of 3), effectively capturing the conceptual meaning and maintaining the relationships between phrases with precision.

In terms of acceptability, GPT was found to be more aligned with Arabic linguistic norms, both morphologically and syntactically (Sadiq, 2025). Most of Google Translate's outputs still retained Indonesian structural patterns (Syam et al., 2023), whereas GPT successfully adapted them into natural and standard Arabic constructions (*jumlah fi'liyah* and *jumlah ismiyyah*). As Nababan, Nuraeni, and Sumardiono state, acceptability is primarily determined by the translator's ability to adapt the source message to the conventions of the target language.

ChatGPT produced translations that were easier for native Arabic speakers to read and comprehend (Widowati et al., 2025). This high level of readability stems from GPT's ability to organize information logically and avoid the long, complex sentence structures typical of Indonesian (Ahmed et al., 2025). These findings are consistent with House's Translation Quality Assessment model, which emphasizes the relationship between sentence structure and clarity of meaning within the context of academic communication.

In terms of terminology selection, GPT demonstrated greater consistency in using academic terminology than GT. Regarding grammar, GPT also excels over Google Translate. Sentences generated by ChatGPT are more accurate. This is in accordance with the rules of Arabic syntax (*nahwu*) and morphology (*sharaf*), meeting the requirements for structural accuracy in translation (Kusumadewi et al., 2024). In contrast, GT often generated word orders inconsistent with Arabic's logical syntactic structure (Abidah et al., 2024).

GPT demonstrated a strong ability to maintain logical relationships between sentence elements and ensure robust textual cohesion (Youssef & Ismail, 2023). In contrast, GT often lost these logical connections due to fragmented sentence structures (Sadikhova & Babayev, 2025). Within the context of this study, ChatGPT demonstrated a high degree of textual equivalence, as described by Baker, in which intersentential relationships are preserved through the appropriate use of pronouns, conjunctions, and the logical sequencing of ideas.

GPT produced an academic Arabic style that was natural, formal, and aligned with the conventions of scholarly Arabic writing (Aljanah et al., 2025). Its sentences were fluent and maintained a consistent logical flow, whereas GT's translations tended to be monotonous and literal (Aghai, 2024). This highlights GPT's ability to emulate the academic register, a crucial skill for translating scientific texts. Thus, GPT excels not only in linguistic aspects but also in pragmatic and stylistic dimensions. This aligns with the principles and strategies of politeness required in academic contexts (Arifianto et al., 2022).

Across all seven dimensions, ChatGPT demonstrated comprehensive and consistent superiority in producing translations that are accurate, natural, cohesive, and academically acceptable. These findings reinforce the results of international studies by Lommel within the MQM framework, which assert that context–based AI systems possess higher evaluative and adaptive capabilities than conventional Neural Machine Translation (NMT) systems.

Moreover, from a pedagogical perspective, this study affirms that ChatGPT can serve as an effective tool in teaching academic Arabic translation, particularly within Arabic Language Education (PBA) programs. However, the integration of ChatGPT in education is most effective when accompanied by instructor supervision, such as post-editing, to ensure that the final translation meets strict academic standards and material suitability criteria (Brahmana et al., 2025; Wahyudi & Gina, 2023).

From a semantic perspective, ChatGPT consistently outperformed Google Translate in capturing relational meaning and academic terminology (Aeni et al., 2024). This aligns with Nida's notion of dynamic equivalence and confirms Baker's argument that lexical and textual equivalence require contextual interpretation rather than literal substitution. GT's literal tendencies mirror findings from studies on Arabic MT challenges (Nagoudi et al., 2022) and support broader research on semantic drift in NMT (Gupta et al., 2022). Recent evaluations of generative AI in academic translation also indicate that LLMs better interpret specialized terminology and discipline-specific concepts (Alzain et al., 2024).

Syntactically, ChatGPT demonstrated a superior ability to restructure Indonesian sentences into natural Arabic constructions (*jumlah ismiyyah* and *jumlah fi'liyyah*) (Jauhar & Setiyawan, 2025). GT, however, frequently retained Indonesian word order, resulting in unnatural syntactic patterns-echoing known limitations of NMT systems (Alenezi, 2024). ChatGPT's ability to maintain correct *iḍāfah*, verb-subject order, and nominalization structures corresponds with studies showing LLM advantages in long-range dependency modeling (Kusuma & Yulia, 2023). Other researchers similarly report that Arabic MT requires more profound syntactic transformations than those in Indo-European languages (Ameur et al., 2020).

Discourse-level differences further distinguish the two systems. ChatGPT maintained cohesive and coherent flow through accurate use of pronominal reference (Suharsono et al., 2024), connectives (Guba & Quba'a, 2025), and thematic progression, an essential component of academic readability following House's TQA framework (Skowron & Bączkowska, 2023). GT's fragmented outputs echo prior findings showing that NMT often struggles with clause linkage and discourse continuity (Kusuma & Yulia, 2023). Studies on LLM coherence modeling support these observations, as generative systems have been shown to outperform NMT at representing multi-sentence discourse relations (Wang et al., 2023).

In comparing these findings with broader MT scholarship, this study reinforces trends reported in various prior works. Chandra demonstrated that generative models outperform conventional NMT systems across low-resource languages. (Al-Darabee et al., 2025) found that GT struggles with culturally embedded expressions when translating into Arabic. Researchers examining Arabic-Indonesian MT similarly observe that ChatGPT produces more natural, contextually appropriate translations (Faris & Abdurrahman, 2023). The superiority of ChatGPT also aligns with the literature, which shows that LLMs produce more genre-appropriate academic expressions (Mohsen, 2024).

CONCLUSION

This study demonstrates that ChatGPT and Google Translate differ substantially in producing Indonesian–Arabic academic translations, revealing qualitative patterns that extend beyond previously reported numerical score distinctions. The findings affirm that ChatGPT's generative architecture enables stronger contextualized processing of semantic relations, adaptive syntactic restructuring, and alignment with the Arabic academic register, resulting in more coherent and rhetorically natural translations. In contrast, Google Translate remains constrained by literal, form-bound translation, which limits its ability to render Indonesian nominal compounds into academically acceptable Arabic constructions and often results in morpho-syntactic rigidity. These results empirically validate the initial research gap, confirming that MT architectural differences directly shape the linguistic and rhetorical adequacy of Indonesian–Arabic academic translation. This domain has been largely underexamined in Indonesian scholarship.

Theoretically, this study contributes to translation-quality research by integrating global frameworks, Multidimensional Quality Metrics (MQM), equivalence-based meaning negotiation, and House's Translation Quality Assessment (TQA) with Nababan's model of accuracy, acceptability, and readability. This hybrid framework proves relevant for evaluating MT performance in language pairs characterized by substantial typological distance and the morphological and syntactic complexity of Arabic, where translation adequacy cannot rely solely on surface-level accuracy but requires deeper functional equivalence and phrase-level adaptation. The study thus extends field-general MT evaluations by situating Indonesian-to-Arabic academic translation within a multidimensional, linguistically grounded quality assessment.

Practically, the findings confirm that generative MT models such as ChatGPT, while not a replacement for human translators or expert review, offer significant pedagogical value as scaffolding tools in Arabic academic translation training, particularly in register adaptation and conceptual phrasing. However, this study also identifies limitations related to genre scope, as the corpus was restricted to thesis titles with limited syntactic variation. Future research is encouraged to expand into more complex academic genres and compare updated generative models to deepen the understanding of MT's evolving capabilities. Collectively, this study reinforces the importance of critical and responsible MT integration in Arabic translation pedagogy and provides a validated methodological basis for future Indonesian–Arabic MT quality research.

ACKNOWLEDGMENT

The author expresses profound gratitude to Universitas Islam Riau (UIR). Sincere thanks are also extended to the lecturers, colleagues, and students of the Arabic Language Education (PBA) Program for their valuable input during data collection and analysis. In addition, the author highly appreciates the contribution of the *Ta'lim al-'Arabiyyah* academic community, which continues to foster a vibrant scholarly space for innovative research in Arabic language education and translation in the digital era. It is hoped that the findings of this study will contribute meaningfully to advancing digital literacy and academic translation pedagogy in Indonesia.

AUTHOR CONTRIBUTIONS STATEMENT

All authors made balanced and significant contributions to this research. [A] served as the lead author, responsible for designing the research framework, conducting data analysis, and preparing the final manuscript. (GANZ) contributed to developing the theoretical framework and refining the methodology in accordance with the context of Arabic language education. [TN] was involved in the linguistic validation process and theoretical review, while [M] contributed to academic language editing and citation consistency. All four authors have read and approved the final version of this article for publication. The authors reported no conflicts of interest.

REFERENCES

Abidah, N. K. K., Hikmawati, R., & Erawanto, V. (2024). ChatGPT in Indonesia-Arabic Translation: A Quality Analysis. *Al-Arabi: Journal of teaching Arabic As A Foreign Lan*guage, *8*(1), 33–52. http://dx.doi.org/10.17977/um056v8i1p33-52

Aeni, A., Baharuddin, B., Putera, L. J., & Melani, B. Z. (2024). The Accuracy of Chatgpt in Translating Linguistics Text in. *Didaktik: Jurnal Ilmiah PGSD STKIP Subang*, *10*(1), 59–68. https://doi.org/10.36989/didaktik.v10i1.2559

Aghai, M. (2024). ChatGPT vs. Google Translate: Comparative Analysis of Translation Quality. *Iranian Journal of Translation Studies*, *22*(85), 85–100. https://dorl.net/dor/20.1001.1.17350212.1403.22.1.9.2

Ahmed, O. F., Dewi, I. K., & Anis, M. Y. (2025). Translating Arabic Literary Metaphors: A Comparative Evaluation of ChatGPT-4.0 and Google Translate 2024. *Ideas: Journal on English Language Teaching and Learning, Linguistics and Literature*, *13*(2), 3177–3199. https://doi.org/10.24256/ideas.v13i2.7439

Al-Darabee, M. A., Farghal, M., & Haider, A. S. (2025). Netflix Versus Google Translate: A Case Study of the English-Arabic Translation of Scatological Terms. *Studies in Linguistics, Culture and FLT*, *13*(2), 77–95. https://doi.org/10.46687/BCAY5747

Alayba, A. M. (2025). Arabic Natural Language Processing (NLP): A Comprehensive Review of Challenges, Techniques, and Emerging Trends. *Computers 14.11*, *14*(11), 1-32. https://doi.org/10.3390/computers14110497

Alenezi, A. M. (2024). Error Analysis of Neural Machine Translation in Technical Texts: Google Translate as A Case Study. *Journal of the North for Humanities*. *8*(2), 167–181. https://doi.org/10.12816/0061799

Aljanah, D. S., Koderi, K., & Akmansyah, M. (2025). The Mind-Machine Duality in Carving Meaning: ChatGPT and Self Regulated Learning in Arabic Research. *Al-irfan: Journal of*

Arabic Literature and Islamic Studies. *8*(1), 428–448. https://doi.org/10.58223/al-irfan.v8i1.470

Alzain, E., Nagi, K. A., & Algobaei, F. (2024). The Quality of Google Translate and ChatGPT English to Arabic Translation: The Case of Scientific Text Translation. *Forum for Linguistic Studies*. *6*(3), 837–849. https://doi.org/10.30564/fls.v6i3.6799

Ameur, M. S. H., Meziane, F., & Guessoum, A. (2020). Arabic Machine Translation: A Survey of The Latest Trends and Challenges. *Computer Science Review*, *38*(100305), 1-29. https://doi.org/10.1016/j.cosrev.2020.100305

Arifianto, M. L., Kholisin, K., Ridwan, N. A., Izzudin, I. F., Mujahidah, Z. A., & Muhtar, S. N. (2022). Designing Research Result-Based Arabic Pragmatic Learning Module On Politeness Principles And Strategies. *Ta'lim Al-'Arabiyyah: Jurnal Pendidikan Bahasa Arab & Kebahasaaraban*, *6*(2), 134–154. https://doi.org/10.15575/jpba.v6i2.20235

Ataman, D., Birch, A., Habash, N., Federico, M., & Koehn, P. (2025). Machine Translation in the Era of Large Language Models: A Survey of Historical and Emerging Problems. *Information 16.9*, *16*(9), 1–36. https://doi.org/10.3390/info16090723

Brahmana, C. R. P. S., Sofyan, R., & Mono, U. (2025). Evaluating The Strengths and Limitations of Chat Gpt-Generated Translations in Academic Post-Editing Workflows : A Case Study of English-Indonesian Academic Texts. *Language Literacy: Journal of Linguistics, Literature, and Language Teaching*, *9*(1), 135–151. https://doi.org/10.30743/ll.v9i1.10551

Chandra, R., Chaudhari, A., & Rayavarapu, Y. (2025). An Evaluation of LLMs and Google Translate for Translation of Selected Indian Languages via Sentiment and Semantic Analyses. *IEEE Access*, *13*, 122386–122407. https://doi.org/10.1109/ACCESS.2025.3585629

El-Farahaty, H. (2025). Integrating Corpora and AI Tools in the Teaching, Learning, and Assessing of Arabic/English Legal Translation. *International Journal for the Semiotics of Law*, *38*(6), 2031–2060. https://doi.org/10.1007/s11196-025-10281-0

Faris, A., & Abdurrahman, M. (2023). Analisis Penerjemahan Bahasa Arab ke Bahasa Indonesia Melalui Artificial Intelligence ChatGPT. *Al-Af'idah: Jurnal Pendidikan Bahasa Arab dan Pengajarannya*, *7*(2), 222–233. https://doi.org/10.52266/al-afidah.v7i2.1967

Guba, M. N. A., & Quba'a, A. A. (2025). AI Translation : Evaluating ChatGPT 's Reliability in Translating Arabic to English. *Jait*: *Journal of Artificial Intelligence and Technology*, 5, 551-556. https://doi.org/10.37965/jait.2025.0831

Gupta, D., Malik, S., & Rana, A. (2022). Adopting Semantic Interoperability for Improved Healthcare. *Proceedings of the International Conference on Innovative Computing & Communication (ICICC)*. 1-11. http://dx.doi.org/10.2139/ssrn.4096399

Hakiki, A. S., Sulthoni, S., & Devi, N. (2023). Analisis Hasil Terjemah Google Translate dan ChatGPT Bahasa Arab-Indonesia: Study Komparatif. *Lanedunc: Journal Of Arabic, English, and Indonesian Language Teaching, Linguistic, and Literature*, *1*(1), 23–41. https://ejournal.aksaracendikia.com/index.php/Lanedunc/article/view/3

Jauhar, A. F. A., & Setiyawan, A. (2025). Comparative Analysis of Arabic Translation Results Between ChatGPT and Deepl. *Tadris Al-'Arabiyyah: Jurnal Pendidikan Bahasa Arab dan Kebahasaaraban*, *4*(2), 287–302. https://doi.org/10.15575/ta.v4i2.46575

Jooste, W., Haque, R., & Way, A. (2021). Philipp Koehn: Neural Machine Translation. *Machine Translation*, *35*(2), 289–299. https://doi.org/10.1007/s10590-021-09277-x

Kusuma, E. D., & Yulia, F. (2023). Analysis of Arabic-Indonesian Translation Errors on Google Translate. *Scaffolding: Jurnal Pendidikan Islam dan Multikulturalisme*, *5*(2), 1–13. https://doi.org/10.37680/scaffolding.v5i1.2722

Kusumadewi, J. S., Anis, M. Y., & Mohammed, C. A. (2024). Translation Analysis of Negative Sentences in the Zakat Chapter of the Book Al-Fiqh Al-Muyassar. *Ta'lim Al-'Arabiyyah: Jurnal Pendidikan Bahasa Arab & Kebahasaaraban*, *8*(1), 34–55. https://doi.org/10.15575/jpba.v8i1.33848

Lommel, A., Uszkoreit, H., & Burchardt, A. (2014). Multidimensional Quality Metrics (MQM): A Framework for Declaring and Describing Translation Quality Metrics. *Tradumàtica Tecnologies de la Traducció*, *12*, 455–463. https://doi.org/10.5565/rev/tradumatica.77

Mohsen, M. (2024). Artificial Intelligence in Academic Translation: A Comparative Study of Large Language Models and Google Translate. Psycholinguistics, *35*(2), 134–156. https://doi.org/10.31470/2309-1797-2024-35-2-134-156

Nagoudi, E. B., Elmadany, A. R., & Mageed, M. A. (2022). Turjuman: A Public Toolkit for Neural Arabic Machine Translation. *Proceedings of the 5th Workshop on Open-Source Arabic Corpora and Processing Tools (Osact5)*, 1–11. https://doi.org/10.48550/arXiv.2206.03933

Nababan, M., Nuraeni, A., & Sumardiono, S. (2012). Pengembangan Model Penilaian Kualiats Terjemahan. *Kajian Linguistik dan Sastra*, *24*(1), 39–57. Retrieved from https://journals.ums.ac.id/index.php/KLS/article/view/101

Nasaruddin, N. (2024). Using ChatGPT in Teaching Arabic as a Foreign Language. *Arabiyatuna: Jurnal Bahasa Arab*, *8*(1), 1-24. https://doi.org/10.29240/jba.v8i1.9413

Nurfaiza, N. (2024). Pengaruh Penggunaan ChatGpt dalam Pembelajaran Terjemah Indonesia-Arab. *Ibtidaiyah: Jurnal Pendidikan Guru Madrasah Ibtidaiyah, 2*(2), 70-89. https://doi.org/10.52491/ibtidaiyah.v2i2.228

Popović, M. (2020). Informative Manual Evaluation of Machine Translation Output. *COLING 2020-28th International Conference on Computational Linguistics, Proceedings of the Conference*, 5059–5069. https://doi.org/10.18653/v1/2020.coling-main.444

Sadikhova, S., & Babayev, J. (2025). Challenges Encountered in Translation of Culture-bound and Subject-specific Abstract Terminology While Using Google Translate. *Euro Global Journal of Linguistics and Language Education*, *2*(3), 119–126. https://doi.org/10.69760/egjlle.2500203

Sadiq, S. (2025). Evaluating English-Arabic translation: Human translators vs Google Translate and ChatGPT Saudi Sadiq. *Journal of Languages and Translation*, *12*(1), 67–95. https://doi.org/10.21608/jltmin.2025.423147

Skowron, D., & Bączkowska, A. (2023). Assessment of Various GPT Models Versus the Human Text: A Quantitative Analysis of Lexis and Cohesion. *Forum Filologiczne Ateneum*, *1*(11), 139–156. https://doi.org/10.36575/2353-2912/1(11)2023-09

Suharsono, S., Ashadi, A., & Sudartinah, T. (2024). Comparative Study of Thematic Choice and Progression on Text Written by Humans and AI Machine. *Joll: Journal of Language and Literature*, *24*(1), 153-174. https://doi.org/10.24071/joll.v24i1.6538

Syafanah, D. N., Permatasari, I., & Rivyal, M. (2025). Google Translate dan Penerjemahan: Penggunaan Google Translate pada Penerjemahan Teks Bahasa Arab. *Blaze: Jurnal Bahasa dan Sastra dalam Pendidikan Linguistik dan Pengembangan*, *3*(3), 255–267. https://doi.org/10.59841/blaze.v3i3.2902

Syam, M. N., Isnaini, R. L., Rohmah, L., & Sa'adah, S. N. (2023). The Analysis of Google Translate Translation Error From Indonesian to Arabic and Tips for Using It. *Ijaz Arabi: Journal of Arabic Learning*, *6*(1), 247–257. https://doi.org/10.18860/ijazarabi.v6i1.16299

Wahyudi, D., & Gina, A. R. (2023). Analysis of Arabic Textbooks by H. Darsono and T. Ibrahim Based on Mackey's Theory. *Ta'lim Al-'Arabiyyah: Jurnal Pendidikan Bahasa Arab & Kebahasaaraban*, *7*(1), 110–127. https://doi.org/10.15575/jpba.v7i1.24847

Wang, L., Lyu, C., Ji, T., Zhang, Z., Yu, D., Shi, S., & Tu, Z. (2023). Document-Level Machine Translation with Large Language Models. *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 16646–16661. https://doi.org/10.18653/v1/2023.emnlp-main.1036

Widowati, D., Mubarak, M. Z., Saehudin, A., Abshar, U., & Suparno, D. (2025). Enhancing the Translation Quality of ChatGPT in Educational News: A Technique-Based Study on Adengad.Net. *Langkawi: Journal of the Association for Arabic and English*, *11*(1), 190–204. https://doi.org/10.31332/lkw.v11i1.10684

Youssef, H., & Ismail, S. (2023). Cohesion and Coherence in Essays Generated by ChatGPT : A Comparative Analysis to University Students' Writing ChatGPT. *CDELT Occasional Papers in the Development of English Education, 83*(1), 143–165. https://doi.org/10.21608/OPDE.2023.325331