Fidelity Analysis of Artificial Intelligence-Powered Poetry Translation: A Comparison of Google Translate, Deepl Translate, and Libre Translate

Muhammad Chairil Imran
Pendidikan Bahasa Inggris, Universitas Islam Makassar, Indonesia

Correspondence*
E-mail: muh.chairil.imran@uim-makassar.ac.id

Abstract

The research aims to analyze the fidelity of Google Translate, DeepL Translate, and Libre Translate in conveying the meaning and intent of the poetry. Employing a mixed-method approach, data were collected from ten purposively selected lecturers at Universitas Islam Makassar and Universitas Megarezky Makassar, Quantitative data were obtained through a structured questionnaire using a five-point Likert scale to measure fidelity, while qualitative insights were gathered via thematic content analysis of the AI-powered poetry translations. Findings reveal a clear fidelity score: DeepL Translate (mean = 4.5, excellent), Google Translate (mean = 3.7, good), and Libre Translate (mean = 1.8, low). The researcher concludes that AI-powered translation tools exhibit considerable variability in their capacity to maintain the message, intent, and stylistic attributes of poetry. Of the three systems assessed, DeepL Translate shows the highest level of fidelity, supported by both quantitative evaluations and qualitative theme analyses.

Keywords: *Artificial Intelligence*, Deeple, *Fidelity, Google Translate, Poetry*

INTRODUCTION

Artificial intelligence (AI) has changed the way translation is done in the last few years. Machine translation techniques have made it possible for platforms like Google Translate, DeepL Translate, and Libre Translate to do complicated language tasks with amazing fluency and understanding of the context. These systems, trained on large multilingual data sets, can produce translations that look and sound like they were made by people (Alowedi & Al-Ahdal, 2023; Naveen & Trojovský, 2024). The quick use of AI in translation has changed how students, teachers, and researchers read texts in other languages.

Even with these improvements, translating literary works, especially poetry, using AI is still quite hard. Poetry has a lot of substance packed into a small space, as well as metaphors, rhythm, and emotional power that other types of writing do not have (Almaktary, 2025; De la Rosa et al., 2023). To translate poetry, you have to recreate both the meaning and the art. It entails not only linguistic precision but also the maintenance of the poet tone, imagery, and cultural identity. The complexity of poetry gives it a great way to look at how AI systems deal with creative and emotional parts of language.

Fidelity is an important idea in translation studies. It means how well a translation stays true to the source material in both meaning and style (Leiter et al., 2024; Zakhir, 2024). In poetry translation, fidelity goes beyond just being the same word for word. It also includes semantic accuracy, stylistic consistency, emotional depth, and cultural resonance (Ghazi, 2025). To find this balance, human translators use interpretation, intuition, and knowledge of other cultures. AI technologies, on the other hand, use computational predictions that might not pick up on nuanced connotations, analogies, or artistic details (Gao et al., 2024). This brings up an essential question on recreating the fidelity that readers demand from poetry translation.

Earlier research on AI translation (Doan, 2025; Farghal & Haider, 2024; Karaban & Karaban, 2024a; Mohammed, 2025; Naveen & Trojovský, 2024; Yuliasri et al., 2025) have predominantly concentrated on quantitative assessment, highlighting measurable factors like accuracy, lexical similarity, and grammatical correctness These evaluations show how well something works technically, but they do not often show how people feel about its artistic and emotional quality. This is especially true for poetry, where rhythm, imagery, and tone are very important (Almaktary, 2025). The increasing usage of AI translation technologies in schools and in literature shows how important it is for people to make decisions. Students and teachers are using these systems more and more to read poetry in many languages, often without realizing how the technology changes the tone or meaning (Ghazi, 2025; Mauladiana & Juniardi, 2024; Resende & Hadley, 2024).

People think about AI translation that can help them learning more about how people judge, trust, or question machine-generated writing.

The development of AI technology in translation has resulted in significant improvements in context processing, lexical accuracy, and grammatical structure. However, most existing research still focuses on quantitative metrics-based evaluations such as word equivalence, grammatical form, and automated scoring. This approach fails to capture the aesthetic and emotional complexity that characterizes poetry, including metaphor, rhythm, imagery, and inherent cultural nuances. Furthermore, comparative research specifically assessing the quality of poetry translations by various AI platforms such as Google Translate, DeepL Translate, and LibreTranslate remains very limited, particularly those that use a comprehensive fidelity assessment framework encompassing meaning, style, form, and the reader's emotional experience. Meanwhile, the use of AI-based translation technology in the context of education and literary literacy continues to grow, while understanding how translation quality affects reader perception, interpretation, and experience remains largely unexplored. This situation highlights the need for more in-depth and multidimensional research to evaluate AI's ability to translate poetry while maintaining fidelity to the form, meaning, and aesthetic value of the original work. This research aims to find out the evolving interplay between artificial intelligence and literary translation, by focusing on the fidelity aspect AI-powered poetry translation.

METHOD

The research employed a mixed-method research design to provide a comprehensive evaluation of AI-powered poetry translation. Poems were selected from Turikale dan Kisah Perjalanan (Rijal, 2023), chosen for their rich imagery and cultural nuance, making them suitable for assessing translation fidelity. Ten lecturers from Universitas Islam Makassar and Universitas Megarezky Makassar were purposively selected to ensure that all respondents possessed sufficient expertise in translation analysis and literary interpretation. Quantitative data were collected through a structured questionnaire in which participants rated the fidelity of translations produced by Google Translate, DeepL Translate, and Libre Translate using a five-point Likert scale (1 = very low to 5 = excellent), following established instruments used in translation evaluation (Chen & Lin, 2025; Gao et al., 2024). Mean scores were calculated to identify overall fidelity aspect. To complement these numerical findings, thematic content analysis was conducted to examine how each AI system rendered semantic meaning, imagery, tone, and the overall spirit of the original poems. This qualitative component enabled a deeper exploration of contextual precision, metaphorical accuracy, and stylistic fluency, offering richer insights into the strengths and limitations of each translation engine.

Quantitative analysis was conducted by calculating the mean, standard deviation, and comparing scores between platforms to identify the fidelity level of

each machine translation. If necessary, comparative statistical tests were applied to determine the significance of differences between groups of translations.

In the qualitative section, the analysis was conducted using a thematic analysis approach. Respondents were asked to provide narrative descriptions of the strengths and weaknesses of each translation based on the dimensions of meaning, imagery, tone, rhythm, and the spirit of the poem. The analysis process involved three stages: open coding, categorization, and theme abstraction to generate patterns of findings regarding the performance of each AI system in translating the stylistic and aesthetic elements of poetry. The entire research process adhered to principles of academic ethics. Participation was voluntary, with guarantees of data confidentiality, respondent anonymity, and data use solely for research purposes.

RESULT AND DISCUSSION

The structured questionnaire data indicated differing levels of perceived fidelity among the three AI translation systems. The mean ratings were derived from ten respondents utilizing the five-point Likert scale.

Table 1. Mean Fidelity Score

Translation System	Mean Fidelity Score	Interpretation
DeepL Translate	4.5	Excellent
Google Translate	3.7	Good
Libre Translate	1.8	Low

Table 2. Sample of Result from AI-Powered Translation

Asyik belalang bertengger di alang-alang Kupu-kupu bertengger di puncak menghisap sari mawar Tangkainya dan ranting tak patah Kuncupnya tetap menghadap ke timur laut (Angin Timur laut in Turikale dan Kisah Perjalanan)	*Teratai di samping sekolah SMA kita Kini tak lagi kuncup mekar ke ufuk timur Mentari pagi Kini ia layu Mungkin terlalu lama merindu (Balasan Cinta Dilan in Turikale dan Kisah Perjalanan)*
Google Translate: The grasshopper is happily perched in the reeds The butterfly is perched on the top, sucking the rose's nectar The stem and twig are unbroken The bud remains facing northeast	Google Translate: The lotus beside our high school Now it's no longer a bud blooming towards the eastern horizon of the morning sun Now it's withered Perhaps I've been missing it for too long
DeepL Translate:	DeepL Translate:

Crickets chirp in the reeds Butterflies perch on the tips, sucking the nectar from roses The stems and branches remain unbroken The buds still face northeast	The lotus beside our high school No longer blooms toward the eastern horizon of the morning sun. Now it withers. Perhaps it has longed too long.
Libre Translate: The grasshopper roars in the tarts The butterfly perched at the peak sucking the rose essence The trap and twigs don't break The lock remains facing northeast	Libre Translate: The lotus next to our high school No longer the bud blooms to the early east horizon Now it's withered. Maybe too long homesick

Respondents regarded DeepL Translate as the most dependable tool for preserving both poetic tone and meaning. Its sentence structure and lexical choices were perceived as more natural and fluid, consistent with findings from recent comparative studies that highlight DeepL Translate superior on the contextual handling and fluency in literary translation tasks (Chen & Lin, 2025; Telaumbanua et al., 2024). Google Translate generated translations that were clearly comprehensible; however, it was frequently criticized for its lack of emotional nuance and tendency toward literal rendering, a limitation also observed in cross-linguistic poetry translation analyses (Anuar, 2025; Gao et al., 2024). Libre Translate received the lowest evaluations due to frequent lexical mismatches, inconsistent rhythm, and awkward phrasing that obscured poetic imagery, which aligns with previous reports about open-source models limited vocabulary depth and contextual adaptability(Abdelhalim et al., 2025; Dost, 2023).

Descriptive statistics indicate a clear preference hierarchy: *DeepL > Google > Libre Translate*. Nevertheless, respondents emphasized that even the most advanced system did not achieve complete fidelity, especially in capturing figurative language and rhythmic patterns central to poetic expression. This reflects the inherent limitations of neural machine translation (NMT) models when processing creative texts that require affective and symbolic interpretation rather than structural precision (Gao et al., 2024). Participants frequently noted that DeepL Translate communicated meaning with higher precision and contextual relevance. Its success is attributed to its ability to capture metaphorical expressions and retain semantic depth, consistent with Chen & Lin (2025), who reported that DeepL performs better at preserving contextual meaning across creative genres. In contrast, Google Translate literal tendency often resulted in flattened symbolism, while Libre Translate restricted lexicon led to occasional misinterpretations. These outcomes reaffirm that richer training data and more sophisticated embedding models lead to more faithful renderings of metaphor and imagery (Dost, 2023; Telaumbanua et al., 2024).

In terms of stylistic continuity and rhythm, DeepL Translate maintained smoother phrasing and more cohesive diction. Similar to the results in Humanities and Social Sciences Communications, DeepL's translations were judged as more rhythmically balanced, while Google Translate often disrupted flow with abrupt lexical choices (Gao et al., 2024). Libre Translate was viewed as fragmented and lacking stylistic

unity—echoing Abdelhalim et al (2025) argument that open-access AI translators often fail to model prosodic structures found in literary texts. Respondents widely agreed that AI translations lack emotional intensity. DeepL was seen as the most emotionally resonant, yet still fell short of the poet's original tone. This finding mirrors (Ghazi, 2025), who concluded that while AI can convey propositional meaning, it struggles to reproduce affective texture. Google Translate neutral tone and Libre Translate inconsistent expressiveness further highlighted that current AI models prioritize cognitive comprehension over emotional recreation (Abdelhalim et al., 2025).

A recurring issue was the AI systems' inability to interpret culturally loaded metaphors and local symbolism accurately. Participants observed that expressions rooted in regional spirituality or tradition were mistranslated or rendered ambiguously. This is consistent with broader concerns about the cultural blind spots of AI translation models, which rely on probabilistic inference rather than interpretive cultural awareness (Gao et al., 2024; Ghazi, 2025). Consequently, human intervention remains essential in preserving cultural and spiritual fidelity within poetic translation.

DeepL Translate has proven to be the most reliable in both semantic and stylistic aspects. The elevated mean fidelity score corresponds with qualitative feedback highlighting naturalness and contextual precision. This finding is consistent with studies showing that DeepL Translate contextualized neural network architecture allows for more natural lexical selection and smoother syntactic generation (Chen & Lin, 2025; Gao et al., 2024). The moderate evaluations of Google Translate reflect its literal dependable conveyance of fundamental meaning, aligning with observations that Google Translate statistical corpus approach often favors accuracy over aesthetic quality (Telaumbanua et al., 2024). Libre Translate, although operational, exhibited significant deficiencies in lexical diversity, fluency, and stylistic consistency, confirming the notion that open-source models with smaller datasets produce translations with reduced semantic depth and stylistic fluidity (Abdelhalim et al., 2025; Agirrezabal et al., 2023).

Participants concurred that none of the AI technologies could fully reproduce the aesthetic and emotional core of the poems. This limitation mirrors the broader challenge identified in literary translation research AI systems can imitate linguistic structures but struggle to replicate the interpretive, emotional, and cultural dimensions of human creativity (Alowedi & Al-Ahdal, 2023; Ghazi, 2025). Both quantitative and qualitative findings in this research reaffirms that current neural translation systems exhibit strong linguistic competence but limited poetic sensibility.

Of the three translation tools assessed, DeepL Translate consistently achieved the highest fidelity ratings and received the most favourable qualitative feedback. Participants observed that DeepL's translations were more natural, contextually coherent, and stylistically refined. This outcome corresponds with previous comparative research indicating that DeepL's Transformer-based network demonstrates superior performance in lexical coherence and contextual accuracy (Chen & Lin, 2025; Gao et al., 2024). However, despite its advanced

handling of syntax and diction, DeepL remained limited in reproducing poetic aesthetics particularly rhythm, meter, and musicality, AI can approximate form, it can not replicate emotional essence. This observation shows findings by Karaban & Karaban (2024), who noted that AI translation remains largely descriptive rather than interpretive, capable of conveying information but not imagination.

Google Translate received moderate scores for both accuracy and affective impact. Participants described its outputs as grammatically correct, it is also a trait observed in earlier analyses of Google literal translation tendencies (Telaumbanua et al., 2024). The system shows reliance on massive multilingual corpora enables high lexical coverage but sacrifices stylistic adaptation. Consequently, metaphors, idioms, and personifications were often rendered into plain, denotative language, diminishing their artistic resonance. This phenomenon aligns with findings in recent studies suggesting that NMT systems such as Google Translate prioritize semantic clarity over stylistic nuance, particularly in creative texts (Gao et al., 2024; Ghazi, 2025). While effective for narrative or descriptive content, Google Translate struggled with abstract, metaphorical, or symbolic poetry confirming the broader consensus that current AI lacks interpretive awareness of literary tone and rhythm.

Libre Translate, as an open-source translation system, showed the weakest fidelity and stylistic coherence. Smaller corpus and limited neural architecture resulted in grammatically inconsistent outputs and frequent lexical repetition. This supports Abdelhalim et al (2025) argument that data scale and model complexity are major determinants of translation quality, especially in tasks demanding interpretive sensitivity. Libre Translate restricted handling of idioms and metaphors further underscores its limitations. Similar conclusions were drawn by Ghazi (2025) who found that lightweight AI models often misrepresent symbolic imagery due to inadequate exposure to figurative language during training. Thus, this research confirms that poetic fidelity requires not just linguistic capacity but aesthetic intelligence, an area still beyond current open-source systems.

The inclusion of human evaluators offered deeper insights into how fidelity is perceived beyond measurable linguistic equivalence. Participants emphasized that fidelity involves not only the accurate transfer of meaning but also the recreation of experience an emotional and cultural resonance that AI can not yet emulate. Similar to findings shows participants described feeling emotionally detached from AI-generated poems, even when the text was grammatically sound (Abdelhalim et al., 2025; Ghazi, 2025). This underscores a persistent gap between cognitive precision and aesthetic empathy. While AI can replicate syntactic patterns, it lacks creativity, cultural intuition, and emotional awareness the very traits that define human translators (Imran, 2023; Karaban & Karaban, 2024b). The respondents' reflections affirm that AI translation remains an assistive tool rather than a replacement for human literary interpretation. In poetry, meaning, structure, and emotion are inseparable; thus, machine translation can only approximate the former two but not the last.

CONCLUSION AND RECOMMENDATION

The researcher concludes that DeepL Translate offers the best and most natural translations of poetry, followed by Google Translate and Libre Translate. All three tools could get across the main idea, but they often lost the poems' beauty, rhythm, and emotion. This illustrates that AI translation still has trouble with creative or artistic language. AI translation systems should be used to help students learning, not to replace human translators. Students and teachers can use them to see how different translations mean the same thing. Developers should work on making AI better at understanding metaphorical language, emotion, and style in poetry. Future research may use additional poems and people to yield more comprehensive results.

REFERENCES

Abdelhalim, S. M., Alsahil, A. A., & Alsuhaibani, Z. A. (2025). Artificial intelligence tools and literary translation: a comparative investigation of ChatGPT and Google Translate from novice and advanced EFL student translators' perspectives. *Cogent Arts and Humanities*, *12*(1). https://doi.org/10.1080/23311983.2025.2508031

Agirrezabal, M., Gonçalo Oliveira, H., & Ormazabal, A. (2023). *Erato: Automatizing Poetry Evaluation*. https://doi.org/10.48550/arXiv.2310.20326

Almaktary, H. (2025). Reflections on Translating Poetry through ChatGPT: Insights from Arabic, English, and French. *Journal of English Studies in Arabia Felix*, *4*(1), 1–12. https://doi.org/10.56540/jesaf.v4i1.108

Alowedi, N. A., & Al-Ahdal, A. A. M. H. (2023). Artificial Intelligence-based Arabic-to-English Machine versus Human Translation of Poetry: An Analytical Study of Outcomes. *Journal of Namibian Studies*, *33*, 1523–1538.

Anuar, A. F. (2025). *Artificial Intelligence and Malaysian Literature: A Comparative Study of AI and Human Translation*. https://doi.org/10.6007/IJARPED/v14-i4/26790

Chen, S., & Lin, Y. (2025). A multidimensional comparison of ChatGPT, Google Translate, and DeepL in Chinese tourism texts translation: fidelity, fluency, cultural sensitivity, and persuasiveness. *Frontiers in Artificial Intelligence*, *8*(July), 1–15. https://doi.org/10.3389/frai.2025.1619489

De la Rosa, J., Pérez Pozo, Á., Ros, S., & González-Blanco, E. (2023). *ALBERTI: A Multilingual Domain Specific Language Model for Poetry Analysis*. https://doi.org/10.48550/arXiv.2307.01387

Doan, T. T. P. (2025). ASSESSMENT OF HUMAN TRANSLATION VS. AI TRANSLATION IN

A LITERARY WORK. *VNU Journal of Foreign Studies*, *41*(3), 126–141. https://doi.org/10.63023/2525-2445/jfs.ulis.5513

Dost, B. O. (2023). Translating poetry with artificial intelligence: Analyzing chatgpt's translations through dryden's methods. *Sosyal Bilimler Enstitüsü Dergisi*, *15*(1), 24–36. https://doi.org/10.30783/nevsosbilen.1613514

Farghal, M., & Haider, A. S. (2024). Translating classical Arabic verse: human translation vs. AI large language models (Gemini and ChatGPT). *Cogent Social Sciences*, *10*(1). https://doi.org/10.1080/23311886.2024.2410998

Gao, R., Lin, Y., Zhao, N., & Cai, Z. G. (2024). Machine translation of Chinese classical poetry: a comparison among ChatGPT, Google Translate, and DeepL Translator. *Humanities and Social Sciences Communications*, *11*(1), 1–10. https://doi.org/10.1057/s41599-024-03363-0

Ghazi, A. (2025). AI and Poetry: Navigating the Challenges of Translating Mahmoud Darwish's "The Dice Player." *European Journal of Language and Culture Studies*, *4*(4), 1–10. https://doi.org/10.24018/ejlang.2025.4.4.148

Imran, M. C. (2023). BABBEL APPLICATION TO IMPROVE STUDENTS ' VOCABULARIES. *Paedagoria : Jurnal Kajian, Penelitian Dan Pengembangan Kependidikan*, *14*(2), 110–113. https://doi.org/10.31764/paedagoria.v14i2.12979

Karaban, V., & Karaban, A. (2024). AI-translated poetry: Ivan Franko's poems in GPT-3.5-driven machine and human-produced translations. *Forum for Linguistic Studies*, *6*(1), 1994. https://doi.org/10.59400/FLS.v6i1.1994

Leiter, C., Lertvittayakumjorn, P., Fomicheva, M., Zhao, W., Gao, Y., & Eger, S. (2024). *Towards Explainable Evaluation Metrics for Machine Translation*. https://doi.org/10.48550/arXiv.2306.13041

Mauladiana, I., & Juniardi, Y. (2024). Translation Techniques in Literary Poetry Texts. *Indonesian Journal of Teaching and Teacher Education (IJTTE)*, *4*(2), 76–83. https://doi.org/10.58835/ijtte.v4i2.406

Mohammed, T. A. S. (2025). Evaluating Translation Quality: A Qualitative and Quantitative Assessment of Machine and LLM-Driven Arabic–English Translations. *Information*, *16*(6), 440. https://doi.org/10.3390/info16060440

Naveen, P., & Trojovský, P. (2024). Overview and challenges of machine translation for contextually appropriate translations. *IScience*, *27*(10), 1–25. https://doi.org/10.1016/j.isci.2024.110878

Resende, N., & Hadley, J. (2024). *Using LLMs to Enhance Poetry Translation*. https://aclanthology.org/2024.amta-research.16.pdf

Rijal, A. S. (2023). *Turikale dan Kisah Perjalanan : Kumpulan Puisi*. Jejak Pustaka.

Telaumbanua, Y. A., Marpaung, A., Putri, C., Gulo, D., Kardo, D., Waruwu, W., Zalukhu, E., & Zai, N. P. (2024). Journal of Artificial Intelligence and Engineering Applications An Analysis of Two Translation Applications: Why is DeepL Translate more accurate than Google Translate? *Journal of Artificial Intelligence and Engineering Applications*, *4*(1), 2808–4519. https://ioinformatic.org/

Yuliasri, I., Purwanto, B., Hadi, M. Z. P., Nurcahyo, A. D., Rochmahwati, P., Rifqi, M., Kumara, K. A. D., & Ma'sumah, F. (2025). Student poetry translations with and without the use of AI: An evaluation of translation quality. *LITE: Jurnal Bahasa, Sastra, Dan Budaya*, *21*(2), 452–468. https://doi.org/10.33633/lite.v21i2.14336

Zakhir, M. (2024). Indirect Poetry Translation: Issues of Fidelity and Adaptation. *Translation Matters*, 6(2), 27–39. https://doi.org/10.21747/21844585/tm6_2a2