الآداب
للدراسـات اللغـويـة والأدبيـة

Poetry is the Last Resort of Translation in the Age of Artificial Intelligence

Dr. Sadek Saleh Alsalemi [*] (ID)   **Dr. Mofareh Dhaher Alhazmi**** (ID)

Sadekalsalmi5@gmail.com   mfarih.alhazimiu@nbu.edu.sa

Dr. Mustafa Ahmed Al-Humari*** (ID)

Mustafa.Alhomari@nbu.edu.sa

Abstract

This study investigates the potential and limitations of artificial intelligence (AI) in translating poetry from English to Arabic compared to human translations. Focusing on Khalil Gibran's *Sand and Foam*, the research compares a human-translated version by Antonious Basheer with an AI-generated version produced by ChatGPT-4. Using Multidimensional Quality Metrics (MQM) and qualitative assessments, the study analyzes accuracy and fluency, reflecting the poetic fidelity of both translations. Findings reveal that while ChatGPT-4 excels in grammatical consistency and structural clarity, it often lacks emotional depth, stylistic features, and cultural resonance in human translation. Conversely, although more expressive, the human version contains notable inconsistencies and errors. The results highlight both AI's promise and current limitations in literary reproduction, advocating for a hybrid approach that combines AI efficiency with human creativity. Ultimately, this research contributes to a deeper understanding of poetic translation and offers pathways to enhance AI's role in preserving literary artistry.

Keywords: Poetry translation, Artificial Intelligence, Error severity, Human translation.

[*] Assistant Professor of Literature, Department of English Studies, College of Education, Ibb University, Republic of Yemen.

**Assistant Professor of Linguistics, Department of Languages and Translation, College of Humanities and Social Sciences, Northern Border University, Kingdom of Saudi Arabia.

*** Assistant Professor of Linguistics, Department of Languages and Translation, College of Humanities and Social Sciences, Northern Border University, Kingdom of Saudi Arabia.

660

(EISSN): 2708-5783 ISSN: 2707-5508 | Arts for Linguistic & Literary Studies, Faculty of Arts, Thamar University, Volume 7, Issue 3, September 2025

الآداب

للدراسات اللغوية والأدبية

الشعر هو الملاذ الأخير للترجمة في عصر الذكاء الاصطناعي

د. مفرح ظاهر الحازمي** ⓘ

mfarih.alhazimiu@nbu.edu.sa

د. صادق صالح السالمي* ⓘ

sadekalsalmi5@gmail.com

د. مصطفى أحمد الحمري*** ⓘ

Mustafa.Alhomari@nbu.edu.sa

ملخص

تهدف هذه الدراسة إلى استكشاف إمكانات وحدود استخدام الذكاء الاصطناعي في ترجمة الشعر من اللغة الإنجليزية إلى اللغة العربية، مقارنة بالترجمات البشرية. وقد ركز البحث على نص "رمل وزبد" لجبران خليل جبران، من خلال مقارنة الترجمة البشرية التي أنجزها أنطونيوس بشير بالترجمة الآليَّة باستخدام نموذج ChatGPT-4. واستنادًا إلى مقاييس الجودة متعددة الأبعاد (MQM) والتقييمات النوعية، تم تحليل مدى الدقة والسلاسة في كل من الترجمتين، بما يعكس مدى الحفاظ على الطابع الشعري للنص الأصلي. كشفت النتائج أن ترجمة ChatGPT-4 تتميز بالوضوح البنيوي والاتساق النحوي، إلا أنها تفتقر في كثير من الأحيان إلى العمق العاطفي واللمسات الأسلوبية والصدى الثقافي الذي تتسم به الترجمة البشرية. في المقابل، اتسمت الترجمة البشرية بدرجة أعلى من التعبير الفني، رغم ما شابها من تباينات وأخطاء. وتؤكد النتائج على ضرورة اعتماد نهج هجين يجمع بين كفاءة الذكاء الاصطناعي والإبداع الإنساني في ترجمة الأعمال الأدبية، كما تسهم هذه الدراسة في تعميق الفهم لعملية الترجمة الشعرية، وتطرح رؤى لتطوير دور الذكاء الاصطناعي في حفظ الجماليات الأدبية.

الكلمات المفتاحية: ترجمة الشعر، الذكاء الاصطناعي، شدة الخطأ، الترجمة البشرية.

661

الآداب للدراسات اللغوية والأدبية، كلية الآداب، جامعة ذمار، المجلد7، العدد3، سبتمبر 2025 | **ISSN: 2707-5508 (EISSN): 2708-5783**

Dr. Sadek Saleh Alsalemi, & Dr. Mofareh Dhaher Alhazmi, & Dr. Mustafa Ahmed Al-Humari

الآداب

للدراسات اللغوية والأدبية

1. Introduction

As the world becomes increasingly automated, artificial intelligence (AI) has become an integral part of daily life, influencing everything from children's toys to scientific missions. Its impact on the humanities is just as significant as on the sciences, affecting fields such as literature, linguistics, and translation studies. In this evolving landscape, engagement with AI is no longer optional—it is a necessity for scholars, professionals, and educators alike.

The integration of machines in literature and translation has a deep historical trajectory, from the Renaissance printing press to the digital revolution. AI technologies like ChatGPT represent the latest frontier, offering advanced capabilities that surpass traditional rule-based or statistical systems. With their ability to grasp linguistic context and produce coherent, fluent output, generative models such as ChatGPT-4 raise important questions about the future of literary and poetic translation (Balasubramanian, 2023; Sutskever et al., 2014).

Nonetheless, poetry translation has been a topic of interest due to its structural form, semantic richness, emotional nuance, and limitations. This study examines the capabilities and subtleties of AI-generated poetry translations compared to those produced by humans, aiming to understand both the advantages and limitations of AI in capturing poetic essence. However, the study tests artificial intelligence generative language models against human translators to understand how they manage the complexities of poetic translation. By examining these translations, the researchers aim to gain a deeper understanding of the pros and cons of these state-of-the-art translation tools — ChatGPT-4, refine research methods, and identify the challenges that AI poetry translation encounters. Accordingly, this study investigates how well AI captures the poetic integrity of *Sand and Foam* —a seminal English-language collection by the Arabic poet Khalil Gibran—when compared to a human translation. By evaluating the translations using the Multidimensional Quality Metrics (MQM) and Scalar Quality Metrics frameworks, the study offers an in-depth analysis of performance in terms of accuracy, fluency, and aesthetic fidelity.

The importance of this study lies in its contribution to a growing and urgent academic discourse at the intersection of artificial intelligence, literary studies, and translation theory. While prior research has examined AI's performance in technical or prose translation, few studies have systematically compared AI and human translations of poetry, especially between English and Arabic—a language pair marked by deep structural and cultural divergence. This research not only tests the limits of current AI capabilities but also sheds light on the broader philosophical question: Can machines convey the soul of poetry? The findings have implications for translation pedagogy, AI development, and cross-cultural literary exchange, and they help determine whether AI should complement or compete with human creativity in literary contexts.

To that end, the following research questions guide this study:

1 - How does ChatGPT perform in translating poetic texts from English to Arabic compared to a human translator in terms of accuracy and fluency based on the MQM framework?

662

(EISSN): 2708-5783 ISSN: 2707-5508 | Arts for Linguistic & Literary Studies, Faculty of Arts, Thamar University, Volume 7, Issue 3, September 2025

2 - What are the predominant types and severities of translation errors found in AI-generated versus human-generated poetry translations?

3 - How well does AI, particularly ChatGPT, preserve the poetic, semantic, and cultural nuances of Khalil Gibran's *Sand and Foam* in Arabic translation?

4 - To what extent can AI-generated translations match or surpass human translations in stylistic coherence and creative fidelity in poetry?

These questions aim to uncover not only the technical competence of AI in poetic translation, but also its limitations in aesthetic judgment, cultural sensitivity, and emotional nuance. Ultimately, this research contributes to a deeper understanding of how artificial intelligence might shape the future of literary translation—and where the human voice remains irreplaceable.

2. Literature Review

2.1 Poetry Translation

Translating poetry is a delicate and complex art that demands a balance between creativity and fidelity to the original text. It places the translator in a dual role: a guardian of the poem's essence and a co-creator who reimagines it in another language (Buçpapaj & Koni, 2024; Lahiani, 2022; Lefevere, 1992). The original poem, woven from intricate threads of rhythm, tone, and meaning, becomes a mosaic that beckons the translator to uncover its mysteries with sensitivity and precision (Bassnett, 2014; Jones, 2011). This process requires going beyond literal interpretation to grasp the source language's subtle cultural references, emotional undertones, and stylistic nuances (Bahrami, 2012; Niknasab, 2011).

Translation, especially of poetry, transforms reading into an act of deep listening (Venuti, 1998). The translator tunes into the poem's rhythm, silence, and melody to recreate its musicality and emotive power in a new linguistic context (Jones, 2011; Tymoczko, 2017). This involves making difficult decisions—such as preserving meter, imagery, or resonance—while adapting elements that may not have direct equivalents. Every choice demonstrates the translator's tact and ability as they work to arouse the same awe and feeling that the original poem evoked (AlAfnan & Alshakhs, 2025).

Poetry translation is ultimately an act of love and a cultural bridge-building endeavour. It celebrates diversity and connection, allowing the soul of one culture to resonate with another (Bassnett, 2014; Tymoczko, 2017). Utilizing his delicate alchemy, the translator presents readers with the gift of uncovering new shades of meaning in the interaction between languages and bridging the familiar with the unfamiliar. The translation of poetry becomes a labor of love and a subtle statement of our shared humanity (Jones, 2011).

2.2 Poetry Translation's Challenges

Poetry translation is widely considered one of the most intricate and demanding forms of translation, even more challenging than composing original poetry (Algobaei et al., 2024; Alzain et al., 2024; Bahrami, 2012; Khalifa, 2015; Naghiyeva, 2015; Othman, 2023). Dryden's claim that translating a poem is more complicated than writing one is echoed by scholars such as Fasiullah (2019) and Newmark (1991). This difficulty arises from poetry's rich interplay of form and content—rhyme, rhythm, metaphor, tone, and cultural nuance—which requires the translator to be both a skilled linguist and a creative poet.

663

الآداب للدراسات اللغوية والأدبية، كلية الآداب، جامعة ذمار، المجلد7، العدد3، سبتمبر 2025 | **(EISSN): 2708-5783 ISSN: 2707-5508**

Fasiullah (2019) and Nair (2018) argue that poetic skill is essential, while Ovidiu (2008) highlights the importance of talent and training. Newmark (1991) maintains that only poets can effectively translate poetry. Musicality is another significant challenge, as noted by Soong (1973), who emphasizes the difficulty of maintaining rhythm and rhyme. Frost (1969) and Venuti (2000) explore the tension between maintaining poetic form and conveying meaning, reflecting the broader debate between formal and dynamic equivalence.

Some scholars attribute the difficulty to poetic language; others emphasize the translator's innate abilities (Newmark, 1991). Raffel (2010) and Tahir (2008) stress that preserving a poem's expressive and aesthetic value may compromise translation adequacy. Likewise, Nida & Taber (1982) and Bahrami (2012) argue that poetry's unique style and sound often resist direct translation. Poetry translation demands exceptional creativity, cultural insight, and linguistic precision.

2.3 Artificial Intelligence

Artificial intelligence (AI) has become a central topic across academic, governmental, and societal discussions, sparking both interest and concern worldwide (Zhang & Wang, 2024; Wang, 2023; Mittelstadt, 2019). However, much of the debate takes place in unprofessional circles, leading to widespread confusion about AI's potential and limitations. Historically, each major technological breakthrough, including AI, has been met with fear and skepticism (Edgerton, 2023). This reaction is natural, but over time, society adapts, recognizing the benefits such innovations offer in enhancing efficiency and productivity.

Rather than posing a threat, AI augments human capabilities, performing tasks independently yet relying on human guidance for programming and ethical direction. Floridi (2019) suggests that AI allows humanity to rediscover its essence, moving beyond the exploitation that marked earlier eras. By automating repetitive or labor-intensive tasks, AI can free individuals to focus on personal growth and meaningful social interactions (West, 2018). It also holds the promise of reducing human involvement in destructive domains, such as warfare (Zequeira, 2024; Ono, 2019). Ultimately, AI should be seen not as a replacement for human life but as a tool to enhance it (Florida, 2019). A balanced, expert-informed approach is crucial to shaping its development responsibly (Mittelstadt, 2019; Motair et al., 2025).

2.4 Automated Poetry Translation

Studzińska (2020) acknowledged improvements in automated poetry translations but raised doubts about their comparability to human translations. Ghazvininejad et al. (2018) proposed neural methods to preserve rhythm and rhyme. Chakrabarty et al. (2021) highlighted the challenges of maintaining the semantics and style in translations, noting the effectiveness of multilingual fine-tuning. Ma and Wang (2020) suggested a framework for analyzing poetry translations without distinguishing human from machine efforts.

Comparative studies by Seljan et al. (2020) showed the success of machine poetry translation, while Dai et al. (2022) pointed out difficulties in capturing the nuances of traditional Chinese culture. Humblé (2019) praised the quality of translating Dickinson poetry via Google, and Kuzman et al. (2019) found Google Neural Machine Translation effective for low-resource languages. Gao et al. (2024) showed that ChatGPT outperformed other models in translating classical Chinese poetry. In Arabic poetry, cultural subtleties challenge AI, with studies by Alowedi and Al-Ahdal (2023) suggesting machine-assisted translations with post-editing as a viable solution.

2.5 ChatGPT

664

(EISSN): 2708-5783 ISSN: 2707-5508 | Arts for Linguistic & Literary Studies, Faculty of Arts, Thamar University, Volume 7, Issue 3, September 2025

ChatGPT, developed by OpenAI, is a powerful AI language model based on the generative pre-trained transformer (GPT) architecture (Radford et al., 2019; Radmir et al., 2024). Trained on vast textual datasets, it generates human-like responses and engages in coherent conversation (Schulman et al., 2022), attracting growing academic and professional interest (Brown et al., 2020). Its strength lies in its interactive capabilities, such as handling follow-up questions, detecting errors, and rejecting inappropriate content (Cai et al., 2023), which are enhanced by reinforcement learning with human feedback (Christiano et al., 2017). Advanced techniques, such as in-context learning and chain-of-thought prompting, further improve its reasoning performance (Brown et al., 2020; Moslem et al., 2023).

ChatGPT demonstrates creative abilities, including generating poetry and narratives (Peng et al., 2023; Zhang et al., 2023). Unlike traditional neural machine translation tools, such as Google Translate and DeepL, which rely heavily on parallel corpora and encoder-decoder models (jiaVaswani et al., 2017), ChatGPT offers greater contextual flexibility (Hendy et al., 2023). Although it may struggle with punctuation and low-resource languages (Jiao et al., 2023), it often outperforms conventional systems regarding fluency, particularly in high-resource languages (Hendy et al., 2023). Its innovative features mark a significant advancement in literary translation (Nagi et al., 2024), making ChatGPT a transformative tool in the evolving field of machine translation.

2. 6 MQM

Multidimensional Quality Metrics (MQM) is developed in order to overcome the shortcomings of the earlier human and automated schemes of evaluating machine translation outputs' quality by using a single metric (Lommel et al., 2014). It is a comprehensive and extensive framework, which provides a detailed description of quality characteristics for evaluation, including style, fluency, correctness, and terminology. MQM is an all-inclusive framework designed to provide precise overall evaluations of translation quality and offers detailed insights into the characteristics of MT output (Freitag et al., 2021; Vardaro et al., 2019). MQM is based on the most effective current translation quality evaluation methodologies and, thus, is a very methodical and consolidated approach for assessing machine translation quality (Lommel, 2018). MQM can be readily adjusted for manual, semi-automatic, and automated assessment settings. Lommel et al. (2014) assert that consumers can customize their measurements, which makes it applicable in many contexts. An additional benefit of MQM is that it is a measure independent of language and may be used with any existing language. Due to its many advantages, MQM has already been used by several scholars working on various topics (Vardaro et al., 2019). It breaks down the quality of translation into many components and their subcomponents. Errors are classified based on their categories and severity levels; their predetermined weights translate them into numerical scores.

The hierarchical enumeration of quality problem categories is essential to a comprehensive Multidimensional Quality Model framework. The hierarchy comprises over 100 errors, and many categories include all the existing translation quality level criteria (Lommel, 2018). The essence of MQM is structured into eight primary branches or dimensions: precision, fluency, terminology, locality convention, style, authenticity, design, and internationalization (Lommel et al., 2018).

The researchers followed a multidimensional approach by utilizing the MQM framework. It enabled them to evaluate the translated text analytically and holistically (Läubli et al., 2020; Toral et al., 2018). The first stage involves detecting the errors at the levels of words, sentences, and paragraphs and evaluating them in relation to the criteria of

665

الآداب للدراسات اللغوية والأدبية، كلية الآداب، جامعة ذمار، المجلد7، العدد3، سبتمبر 2025 | (EISSN): 2708-5783 ISSN: 2707-5508

each of the selected paradigm's dimensions. The second is through the aggregation of the dataset of the dimensions and assessing them in relation to the whole text (Mariana et al., 2015).

3. Methodology

3.1 Research Design

This study adopts a comparative analytical framework, utilizing both qualitative and quantitative methods to evaluate the accuracy, fluency, and stylistic quality of AI-generated versus human-translated poetry. The research specifically analyzes translations of Khalil Gibran's *Sand and Foam* from English to Arabic using the Multidimensional Quality Metrics (MQM) framework and a Scalar Quality Metric. By combining error annotation with qualitative ratings, the study aims to explore the capabilities and limitations of ChatGPT in literary translation.

3.2 Population and Sampling Method

The population of the study consists of all English poems from Khalil Gibran's *Sand and Foam* The researchers purposively selected this collection due to its significance in modern mystical poetry and its cultural relevance in both Arabic and English literary traditions. The study employed "purposive sampling" to select two distinct Arabic translations:

- "Human Translation": Antonious Basheer's 1926 Arabic translation, comprising 258 poems (5,789 words).
- "AI Translation": ChatGPT-4-generated translation of all 320 English poems (4,938 words).

This purposive approach allows for focused comparison between a historically significant human translation and a contemporary AI-generated version.

3.3 Data Collection Tools

3.3.1 Translation Corpus Development

- ChatGPT-4 was prompted to produce a translation of Gibran's *Sand and Foam* retaining poetic tone and fidelity to the original meaning, during January to May 2025.
- The two translations, Basheer's and ChatGPT's, were compiled into a parallel corpus, aligning each poem with its source text.

3.3.2 Evaluation Tools

1. Multidimensional Quality Metrics (MQM):

- A hierarchical error annotation framework assessing "accuracy" (e.g., additions, omissions, mistranslations) and "fluency" (e.g., grammar, cohesion, punctuation).
- Errors were annotated manually by the research team using 13 predefined MQM dimensions.

2. Scalar Quality Metric (SQM):

- A 6-point Likert scale (0—5) was applied to each poem by two professional translators.
- Criteria included poetic fluency, stylistic appropriateness, and semantic integrity.
- Reliability: the Cohen's Kappa score of $\kappa = 0.93$ indicates excellent agreement between the two raters.

3. Severity Weighting:

666

(EISSN): 2708-5783 ISSN: 2707-5508 | Arts for Linguistic & Literary Studies, Faculty of Arts, Thamar University, Volume 7, Issue 3, September 2025

- Errors were also weighted for severity: Minor = 1, Major = 5, Critical = 25.
- Penalty scores were normalized to assess relative translation quality.

3.4 Data Analysis Procedure

1. Quantitative Analysis:

- Errors were categorized and counted using MQM, then severity-weighted.
- Comparative metrics such as error rates per word, sentence, and poem were calculated.
- Overall translation quality was computed as a normalized penalty score.

2. Qualitative Analysis:

- The Scalar Quality Metric provided descriptive insights into aesthetic and stylistic fidelity.

3. Comparative Scoring:

- Basheer and ChatGPT's translations were compared across multiple dimensions.
- Statistical indicators such as mean quality scores and standard deviations were calculated.

4. MQM Results

4.1 Error Annotation

This study utilizes the Multidimensional Quality Metrics (MQM) framework to evaluate and compare the translations of Khalil Gibran's poetry collection *Sand and Foam* by Antonious Basheer and the AI-driven model ChatGPT. The paradigm identified various types and severities of translation errors across both versions. The researchers detected 794 errors in Basheer's translation, representing 13.65% of the 5,789 words of translated text. In contrast, ChatGPT's translation contained 146 errors, accounting for 2.96% of its 4,938-word output. The total error count is provided in Table 1 below:

Table 1

Total Error Count

	Basheer's error count	ChatGPT's error count
Total Errors	794	146

4.2 Error Frequency

Applying the MQM framework to Basheer's translation, the study found 373 sentences—108 fewer than the original—and 5,789 words, 1,644 words less than the source text. This represents a substantial reduction in both sentence and word counts. The error frequency is notably high, with 2.12 errors per sentence and approximately 3 errors per poem.

In contrast, ChatGPT's translation includes 450 sentences and 4,938 words, 31 sentences and 2,495 words less than the source text. Despite the lexical reduction, ChatGPT maintained closer syntactic alignment with the source, with an error rate of 0.32 per sentence and 0.45 per poem. The difference in sentence count is 22.55% in Basheer's version

667

الآداب للدراسات اللغوية والأدبية، كلية الآداب، جامعة ذمار، المجلد7، العدد3، سبتمبر 2025 | **(EISSN): 2708-5783 ISSN: 2707-5508**

versus 6.25% in ChatGPT's, suggesting that the AI translation preserves more of the source text's syntactic structure. The error frequency of ChatGPT and Basheer's translations is included in Table 2.

Table 2

Error Frequency

Category	Human Translation (Basheer)	AI Translation (ChatGPT)	Source Text
Poem Count	262	321	321
Sentence Count	373	450	481
Word Count (Translated Text)	5,789	4,938	7,433
Accuracy Error Count	394	73	-
Fluency Error Count	400	73	-
Total Error Count	794	146	-
Error Rate per Word	13.7%	2.96%	-
Error Rate per Sentence	2.12 errors	0.32 errors	-
Error Rate per Poem	3 errors	0.45 errors	-

4.3 Error Distribution

The MQM's error typology of Antonious Basheer and ChatGPT's translations of Khalil Gibran's *Sand and Foam* comprised diverse errors that vary in type, quantity, and degree as Table 3 below shows. In Basheer's translation, the researchers detected 794 errors, which are classified into 394 for accuracy and 400 for fluency. The accuracy errors form 49.74% and the fluency errors form 50.76% out of the total count of errors. However, MQM error typology of ChatGPT's translation of the same poetry collection annotated 146 errors. The accuracy dimension included 73 errors, forming 50% of ChatGPT's detected errors, and the fluency dimension annotated errors were, also, 73 errors rated 50%.

Table 3

Error Distribution

Error type	Basherer's error count	ChatGPT's error count
Accuracy	394 (49,62%)	73(50%)
Fluency	400(50,38%)	73(50%)
Total	794 (100%)	146(100%)

4.3.1 Accuracy Errors

Table 4 below presents a comparative analysis of accuracy-related translation errors found in Antonious Basheer's human translation and ChatGPT's AI-generated version of *Sand and Foam*. Errors are classified using five key MQM dimensions——Addition, Mistranslation, Overly Literal, Omission, and Untranslated——to assess how effectively each translation preserves the semantic integrity of the original text. The analysis highlights distinct translational

668

(EISSN): 2708-5783 ISSN: 2707-5508 | Arts for Linguistic & Literary Studies, Faculty of Arts, Thamar University, Volume 7, Issue 3, September 2025

tendencies and provides insight into the respective strengths and limitations of human and AI translation in conveying poetic meaning.

Table 4

Basheer and ChatGPT's Accuracy Errors

Translator	Basheer			ChatGPT		
Dimension	Error count	Error rate of Dimension's errors	Error rate of Error Total	Error count	Error rate of Dimension's Error	Error rate of Error Total
Addition	181	45,94%	22,80%	5	6,85%	3,42%
Mistranslation	146	37,05%	18,39%	8	10,96%	5,47%
Overly literal	28	7,11%	3,52%	49	67,12%	33,56%
Omission	39	9,90%	4,91%	10	13,70%	6,85%
Untranslated	0	0%	0%	1	1,37%	0,70%
Total	394	100%	49.62%	73	100%	50%

The distinct patterns in the types and frequencies of accuracy errors found in both translations are illustrated in Table 4 above.

Basheer's Translation recorded a total of 394 errors, with the highest percentage occurring in the Addition category (45.94%), suggesting a significant tendency to introduce content not present in the source text, possibly to preserve the poetic flow. Mistranslation occurred at a rate of 37.05%, indicating notable semantic deviations. Omission and Overly Literal errors were comparatively low, showing greater interpretation flexibility, while no Untranslated segments were found.

ChatGPT's Translation exhibited 73 errors in total, showing a different error profile. The most prominent category was "Overly Literal" (67.12%), which implies that AI tended to prioritize word-for-word Translation at the expense of poetic nuance. Addition and Mistranslation rates were significantly lower (6.85% and 10.96% respectively), reflecting a more conservative rendering. However, the presence of Untranslated content (1.37%)—absent in Basheer's version—suggests occasional model limitations in processing or interpreting specific poetic elements.

Overall, Basheer's Translation introduced more content and variation, while ChatGPT's output leaned heavily on literal accuracy but sometimes failed to adapt stylistically or contextually. The total accuracy error rate for Basheer (49.62%) is slightly lower than that of ChatGPT (50%), though the nature of errors differs significantly, highlighting contrasting translation strategies.

4.3.2 Fluency Errors

The fluency-related errors identified in Antonious Basheer's human translation and ChatGPT's AI-generated version of *Sand and Foam* are included in Table 5 below. Using the MQM framework, fluency is evaluated through eight dimensions: Grammar, Punctuation, Unintelligibility, Ambiguity, Spelling, Register, Style, and coherence and cohesion. These categories evaluate how well each translation aligns with the target language's linguistic norms, stylistic flow, and

669

الآداب للدراسات اللغوية والأدبية، كلية الآداب، جامعة ذمار، المجلد7، العدد3، سبتمبر 2025   (EISSN): 2708-5783 ISSN: 2707-5508

readability. The comparison provides insight into each translator's ability to produce a smooth, natural-sounding text that aligns with the conventions and expectations of Arabic literary style.

Table 5

Baheer and ChatGPT's Fluency Errors

Translator	Basheer			ChatGPT		
Dimension	**Error count**	**Error rate of Dimension's errors**	**Error rate of Error Total**	**Error count**	**Error rate of Dimension's errors**	**Error rate of Error Total**
Grammar	58	14.5%	7.30%	10	13.70%	6.84%
Punctuation	77	19.25%	9.70%	16	21.92.	11.00%
Unintelligible	23	5.75%	2.90%	17	23.28%	11.64%
Ambiguity	6	1.5%	0.75%	4	5.48%	2.73%
Spelling	6	1.5%	0.75%	2	2.74%	1.37%
Register	14	3.5%	1.77%	2	2.74%	1.37%
Style	137	34.25%	17.26%	22	30.14%	15.05%
Coherence and cohesion	79	19.75%	9.95%	0	0%	0%
Total	400	100%	50.38%	73	100%	50%

The statistical figures of Table 5 indicated that Basheer's translation contained 400 fluency errors. The most frequent issues occurred in Style (34.25%), followed by Coherence and cohesion (19.75%) and Punctuation (19.25%). Grammar errors accounted for 14.5%, while other dimensions—unintelligibility (5.75%), Register (3.5%), Ambiguity (1.5%), and Spelling (1.5%)—were less prevalent. These results suggest that while Basheer preserved meaning, stylistic and cohesion issues impacted fluency.

ChatGPT's translation showed fewer overall fluency errors (73). Style also emerged as the most frequent error type (30.14%). However, Unintelligibility followed with (23.28%), indicating challenges in producing understandable poetic language. Punctuation errors (21.92%) and Grammatical errors (13.70%) were the most common, while Coherence and cohesion showed no recorded issues. Ambiguity recorded 5.48% and Register, and Spelling, represented less than 5% of the errors.

The analysis reveals that while ChatGPT performed better in maintaining textual Coherence, it struggled with intelligibility and stylistic nuance. Basheer's translation demonstrated greater poetic awareness but encountered consistency issues in cohesion and stylistic fluency.

4.4 Error Severity Distribution

The MQM paradigm employed in this study consists of two error issue categories, accuracy and fluency, with 13 dimensions. Accuracy included five dimensions, and fluency included eight dimensions. The severity was

670

(EISSN): 2708-5783 ISSN: 2707-5508 | Arts for Linguistic & Literary Studies, Faculty of Arts, Thamar University, Volume 7, Issue 3, September 2025

determined after annotation error. A three-level severity multiplier is applied to all the given dimensions. The assigned weights are:

- Minor error weight = 1
- Major error weight = 5
- Critical error weight = 25

Tables 7 and 5 below provide a comprehensive classification of error types and the severity degree of each type. The following lines will detail and explain the various categories of our selected scheme, accompanied by examples. However, we will pursue the same order of error distribution, starting with accuracy error typology and proceeding to fluency error typology.

4.4.1 Accuracy Error Severity Distribution

Addition, omission, and mistranslation are the most significant dimensions of accuracy, characterized by their profound impact on translation quality. Addition and omission are classified as high-risk dimensions. However, mistranslation has a more severe impact on translation quality than either. It has been classified as a high-risk coefficient dimension. Tables 6 and 7 below include ChatGPT and Basheer's distribution of accuracy errors' severity:
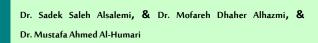
Table 6

Distribution of Accuracy errors' severity in Basheer's Translation

Dimension	Minor	Major	Critical	Total
Addition	161 (40.9%)	18 (4.6%)	2 (0.5%)	181 (45.94%)
Mistranslation	94 (23.9%)	34 (8.6%)	18 (4.6%)	146 (37.05%)
Overly Literal	28 (6.9%)	1 (0.3%)	0 (0.0%)	28 (7.11%)
Omission	33 (8.4%)	1 (0.3%)	5 (1.3%)	39 (9.9%)
Untranslated	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Total Errors	315 (79.9%)	54 (13.7%)	25 (6.3%)	394 (100%)

Table 7

Distribution of Accuracy errors' severity in ChatGPT's Translation

Dimension	Minor	Major	Critical	Total
Addition	5 (6.8%)	0 (0.0%)	0 (0.0%)	5 (6.85%)
Mistranslation	6 (8.2%)	2 (2.7%)	0 (0.0%)	8 (10.96%)
Overly Literal	44 (60.3%)	5 (6.8%)	0 (0.0%)	49 (67.12%)
Omission	9 (12.3%)	1 (1.4%)	0 (0.0%)	10 (13.7%)
Untranslated	0 (0.0%)	1 (1.4%)	0 (0.0%)	1 (1.37%)
Total Errors	64 (87.7%)	9 (12.3%)	0 (0.0%)	73 (100%)

The tables 6 and 7 above revealed that Bsheeer produced 315 minor errors representing 78,75% of accuracy and 39,67% of the error count, respectively compared to ChatGPT's 64 minor errors representing 87,6% of accuracy and

671

الآداب للدراسات اللغوية والأدبية، كلية الآداب، جامعة ذمار، المجلد7، العدد3، سبتمبر 2025 **(EISSN): 2708-5783 ISSN: 2707-5508**

43,8%, of error count, respectively. However, Basheer produced 54 major errors representing 13,5% of accuracy and 6.8% of error count compared to ChatGPT's 9 major errors representing 12% of accuracy and 6,1% of error count, respectively. Eventually Basheer produced 25 critical 6,25% of accuracy and 3.15% of error count, respectively, compared to ChatGPT's zero critical error. While Basheer's version includes more severe and meaning-altering mistakes, ChatGPT demonstrates greater accuracy and consistency despite its tendency toward literalness, which may slightly affect the natural flow without seriously distorting the meaning. The error type and severity are explained in the following examples:

4.4.1.1 Addition

Some of Basheer's addition errors are classified as low-risk due to their limited impact on the text. These additions neither enhance nor significantly harm the poetic value or clarity.

Table 8

Examples of Basheer's Addition Minor Errors

No	Original text	Translation	Error description
1	The Sphinx spoke only once (Gibran, 1926, p. 3).	تكلم أبو الهول مرة في حياته (p. 9).	Added word: حياته (his life). The addition of حياته does not add any poetic value. While it does not distort the meaning, it also offers no stylistic enhancement.
2	"When God threw me, a pebble" (Gibran, 1926, p. 5).	عندما رماني اللهُ حصاةً صغيرةً (p. 9).	The word "صغيرة" introduces redundancy, as the noun "pebble" already implies smallness.

The second category of Basheer's addition errors is medium-risk, which includes the most frequent and impactful type of additions, amounting to 18 instances. These additions affect the naturalness and flow of the translation, although the text remains comprehensible mainly. Table 9 contains some examples that illustrate the negative consequences of unnecessary word insertions.

Table 9

Examples of Basheer's Major Errors

No	Original text	Translation	Error description
1	A bigot is a stone-deaf orator (Gibran, 1926, p. 58).	المتعصب بالدين خطيبٌ بالغ الصمم (p. 32).	The addition "بالدين" restricts the generality of the term bigot, limiting it to religious fanaticism. This reduces the broader philosophical and universal scope of the original line.
2	And the philosopher answered, saying, 'I study man's mind, his deeds, and his desires' (Gibran, 1926, p. 65).	فأجاب الفيلسوف متبجِّحًا: إنني أدرس أخلاق الناس وطبائعهم وأبحث في أعمالهم (p. 35)	This addition "متبجحًا" alters the tone. It introduces a biased tone, portraying the philosopher as arrogant and distorting the original neutral portrayal, thus influencing reader perception.

672

(EISSN): 2708-5783 ISSN: 2707-5508 | Arts for Linguistic & Literary Studies, Faculty of Arts, Thamar University, Volume 7, Issue 3, September 2025

The critical errors are the least frequent, with only two occurrences, but their impact on the quality and intent of the translation is substantial. These additions significantly impair the original meaning and stylistic integrity. Some of the addition's critical errors are in Table 10.

Table 10

Examples of Basheer's Critical Errors

No	Original text	Translation	Error description
1	I am the flame and I am the dry brush, and one part of me consumes the other part (Gibran, 1926, p. 57).	أنا اللهيب وأنا الهشيم اليابس، وبعضي يأكل بعضي فهلَّا حوَّلت وجهك عني لكيلا يعميك دخاني! (p. 31).	The second verse is not found in the original. The added verse shifts the focus from internal struggle to external reaction, compromising the poem's thematic coherence and the author's original intent.

In contrast, ChatGPT's additions were minimal (only 5 instances), and all fell under the low-risk category. These additions served poetic and stylistic functions, clarifying implied meanings or enhancing rhythm without distorting the original tone or message. Some of ChatGPT's addition errors that are of minor severity are in Table 11 below:

Table 11

Examples of ChatGPT's Addition Minor Errors

No	Original text	Translation	Error description
1	Some of our children are our justifications and some are but our regrets (Gibran, 1926, p. 74).	بعض أبنائنا مبررات وجودنا، وبعضهم مجرد ندم (2025).	This addition "وجودنا" clarifies the phrase "justifications" by linking it explicitly to existence. It preserves the ambivalence of the original while enhancing clarity and emotional depth.
2	Your mind and my heart will never agree until your mind ceases to live in numbers and my heart in the mist (Gibran, 1926, p. 30).	لن يتفق عقلك وقلبي حتى يكف عقلك عن العيش في الأرقام، ويكف قلبي عن التوهان في الضباب(2025).	The word "التوهان" effectively captures the elusive and disoriented quality of "mist." It is a subtle and poetic enhancement that maintains fidelity to the original imagery.

4.4.1.2 Omission

Basheer's translation exhibits 39 omission errors, as categorized under accuracy-related issues. Among these, 33 are minor, 1 is medium-risk, and 5 are considered high-risk or critical. Given the significant impact of critical errors on the translation quality, we are going to confine our examples to the critical errors. However, major and minor errors are going to be included when necessary. Nonetheless, the impact of critical omissions on the translation's fidelity is crucial.

673

الآداب للدراسات اللغوية والأدبية، كلية الآداب، جامعة ذمار، المجلد7، العدد3، سبتمبر 2025 | **(EISSN): 2708-5783 ISSN: 2707-5508**

Dr. Sadek Saleh Alsalemi، & Dr. Mofareh Dhaher Alhazmi، & Dr. Mustafa Ahmed Al-Humari

Thus, the focus is on these most severe cases, with minor and medium-risk examples included only when necessary. The most serious omission by Basheer is the deletion of three entire poems without explanation or justification. These critical omissions are shown in Table 12.

Table 12

Examples of Basheer's Critical Omission Errors

No	Original text	Translation	Error description
1	Poem No. 273: (Gibran, 1926, p. 72).	No translation available	The whole poem is deleted
2	Poem No. 213: (Gibran, 1926, p. 57)	No translation available	The whole poem is deleted.
3	Poem No. 25: (Gibran, p. 8)	No translation available	The whole poem is deleted.
4	Poem No. 251: They say to me, "A bird in the hand is worth ten in the bush." But I say, "A bird and a feather in the bush is worth more than ten birds in the hand. Your seeking after that feather is life with winged feet; nay, it is life itself" (Gibran, 1926, p. 67).	يقولون لي عصفور في اليد ولا عشرة على الشجر، أما أنا فأقول لهم: إن عصفورًا واحدًا على الشجر خيرٌ من عشرة في اليد (p. 36).	This final stanza was entirely omitted by Basheer, resulting in a distorted message and denying readers a complete understanding of the poem's metaphorical richness and thematic unity.

These omissions compromise the integrity and completeness of the translated work, casting doubt on the translator's fidelity and accountability. The deliberate exclusion of entire poems leads to a substantial loss of meaning and disrupts the coherence of the original text's philosophical and poetic vision.

In contrast, ChatGPT's translation contains only 9 omission errors, all classified as minor. Some of these examples are included in Table 13.

Table 13

Examples of ChatGPT's Omission Errors

No	Original text	Translation	Error description
1	And when April returned and spring came to wed the earth, there grew in my garden beautiful flowers unlike all other flowers (Gibran, 1926, p. 82).	وحين عاد نيسان، تزوج الربيع الأرض، نبتت في حديقتي زهورٌ لا تشبه أي زهورٍ أخرى (2025).	The word "beautiful" is deleted from the translation. It had been deleted for a poetic function, as style demanded the economical use of words.
2	And they talk long; and each time Jesus of Nazareth goes away saying to Jesus of the Christian, "My friend, I fear we shall never, never agree " (Gibran, 1926, p. 77).	وفي كل مرة يقول يسوع الناصري ليسوع المسيحي: "يا صديقي أخشى أن لا نتفق أبدا"(2025).	"goes away" is the word removed from the translation, slightly affecting the overall meaning and style.

674

(EISSN): 2708-5783 ISSN: 2707-5508 | Arts for Linguistic & Literary Studies, Faculty of Arts, Thamar University, Volume 7, Issue 3, September 2025

4.4.1.3 Mistranslation

Mistranslation is the worst type of accuracy error, producing incorrect meaning and message. The main reason for mistranslation is the unskilled and unqualified translator. The translator's lack of individual and linguistic tools and poor source text comprehension hindered meaning transfer and produced bilateral contextual misunderstandings.

However, Basheer produced a large number of mistranslating errors, with 146 errors. Minor errors represented the majority, with 94 errors, followed by 34 major errors, and critical errors ranked third, with 18 errors. Nevertheless, Table 14 shows some striking mistranslation errors that harm and affect the language's meaning and fluency.

Table 14

Examples of Basheer's Mistranslation Critical Errors

No	Original text	Translation	Error description
1	And when April returned and spring came to wed the earth, Gibran, 1926, p. 82).	وعندما رجع نيسان وجاء الصيف ليتزوج الأرض (p. 43).	"spring" is mistranslated as "الصيف" while it is actually the name of another season, which is "الربيع".
2	Only the dumb envy the talkative (Gibran, 1926, p. 16).	لا يحسد الثرثار إلا الأصم (p. 14).	"الاصم" is a mistranslated word. It means the deaf, while the intended meaning of the source is dumb, which means "الأبكم".

ChatGPT also made some mistranslating errors with various degrees of severity, as shown in Table 15:

Table 15

Examples of ChatGPT's Mistranslation Errors

No	Original text	Translation	Error description
1	Frogs may bellow louder than bulls (Gibran, 1926, p. 16).	قد تنق الضفادع أعلى من ثغاء الثيران (2025).	The bull's voice is mistranslated as the voice of lambs. This error is of major severity and partially affects the meaning.
2	The first time when I saw her being meek that she might attain height (Gibran, 1926, p. 10).	الأولى، حين رأيتها تتصنع التواضع لتبلغ العلو (2025).	The word "meek" is translated as "humble," which has a positive meaning. The intended meaning is "submissive." This is an error of minor severity that slightly affected the meaning.

4.4.1.4 Overly literal

The overly literal errors are concerned with the poor performance of unskillful translators. Some examples of these errors are included in Table 16. The first example is of minor severity and produced by ChatGPT. The second is of major severity and produced by Basheer.

675

الآداب للدراسات اللغوية والأدبية، كلية الآداب، جامعة ذمار، المجلد7، العدد3، سبتمبر 2025 | **(EISSN): 2708-5783 ISSN: 2707-5508**

Table 16

Examples of Overly literal Errors

No	Original text	Translation	Error description
1	There is no struggle of soul and body save in the minds of those whose souls are asleep and whose bodies are out of tune (Gibran, 1926, p. 26).	لا صراع بين الجسد والروح إلا في عقول من أرواحهم نائمة وأجسادهم نشاز (2025).	The phrase "out of tune " is translated literally as "نشاز"; the word is not appropriate to describe the human body. The intended meaning is "sick" or "tired." It is an error of minor severity, as it does not affect the overall meaning.
2	Behold here is a paradox: the deep and high are nearer to one another than the mid-level to either. (Gibran, 1926, p. 75).	إليك هذه الأُحجية: إن العميق أو العالي هما أقرب أحدهما إلى الآخر من المتوسط لأحدهما (p. 40).	The word "mid-level" is overly literal translated as "المتوسط" which results in unintelligibility and confusion. The meaning is obscured. It is an error of major severity

4.4.1.5 Untranslated

Untranslated errors concern the completeness and perfection of translation. We have only an untranslated error produced by ChatGPT. However, Basheer deleted three poems, which are classified as omissions. Table 17 shows some of the ChatGPT's untranslated errors.

Table 17.

Examples of Untranslated Major Errors

No	Original text	Translation	Error description
1	How noble is the sad heart who would sing a joyous song with joyous hearts (Gibran, 1926, p. 18).	ما noبل القلب الحزين الذي يغني مع القلوب المفرحة اغنية فرح (2025).	As the translation revealed, half of the word "noble" remained untranslated. This produced unintelligibility and prevented the message from being conveyed. It is a major error.

4.4.2 Fluency Error Severity Distribution

Fluency errors affect the message's naturalness, fluency, professionalism, and transmission. MQM stats revealed that Basheer and ChatGPT produced almost equal number of minor errors. Basheer produced 335 minor errors representing 84.5% of fluency and 42.5% of error count, respectively. However, ChatGPT made 59 minor errors representing 80% of fluency and 39.4% of error count. Basheer produced 43 major errors representing 10.8% of fluency and 5.4% of error count respectively compared to ChatGPT's 14 major errors representing 19% of fluency and 9% of error count respectively. Basheer produced 19 critical errors representing 4.75% of fluency and 2.4% of error count,

676

(EISSN): 2708-5783 ISSN: 2707-5508 | Arts for Linguistic & Literary Studies, Faculty of Arts, Thamar University, Volume 7, Issue 3, September 2025

respectively compared to ChatGPT's zero critical errors. The Tables 18 and 19 below include Basheer and ChatGPT's fluency error distribution and severity classification of these errors.

Table 18.

Distribution of fluency errors' severity in Basheer's Translation

Dimension	Minor	Major	Critical	Total
Grammar	54 (13.5%)	4 (1.0%)	0 (0.0%)	58 (14.5%)
Punctuation	77 (19.3%)	0 (0.0%)	0 (0.0%)	77 (19.25%)
Unintelligible	2 (0.5%)	13 (3.3%)	8 (2.0%)	23 (5.75%)
Ambiguity	1 (0.3%)	1 (0.3%)	4 (1.0%)	6 (1.5%)
Spelling	5 (1.3%)	1 (0.3%)	0 (0.0%)	6 (1.5%)
Register	9 (2.3%)	5 (1.3%)	0 (0.0%)	14 (3.5%)
Style	123 (30.8%)	11 (2.8%)	3 (0.8%)	137 (34.25%)
Coherence & Cohesion	67 (16.8%)	8 (2.0%)	4 (1.0%)	79 (19.75%)
Total Errors	**338 (84.5%)**	**43 (10.8%)**	**19 (4.8%)**	**400 (100%)**

Table 19

Distribution of fluency errors' severity in ChatGPT's Translation

Dimension	Minor	Major	Critical	Total
Grammar	10 (12.0%)	0 (0.0%)	0 (0.0%)	10 (13.7%)
Punctuation	16 (19.3%)	0 (0.0%)	0 (0.0%)	16 (21.92%)
Unintelligible	7 (8.4%)	10 (12.0%)	0 (0.0%)	17 (23.28%)
Ambiguity	2 (2.4%)	2 (2.4%)	0 (0.0%)	4 (5.48%)
Spelling	0 (0.0%)	2 (2.4%)	0 (0.0%)	2 (2.74%)
Register	2 (2.4%)	0 (0.0%)	0 (0.0%)	2 (2.74%)
Style	22 (26.5%)	0 (0.0%)	0 (0.0%)	22 (30.14%)
Coherence & Cohesion	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Total Errors	**59 (71.1%)**	**14 (16.9%)**	**0 (0.0%)**	**73 (100%)**

Tables 18 and 19 show Basheer's and ChatGPT translations' count and percentage of fluency errors. Basheer's version contains 400 fluency errors, with the most frequent errors being style (137, 34.25%), punctuation (77, 19.25%), and coherence and cohesion (79, 19.75%). It also includes 19 critical errors (4.8%), primarily in terms of unintelligibility

677

الآداب للدراسات اللغوية والأدبية، كلية الآداب، جامعة ذمار، المجلد7، العدد3، سبتمبر 2025 | **(EISSN): 2708-5783 ISSN: 2707-5508**

and ambiguity. In contrast, ChatGPT's translation contains only 83 fluency errors, primarily style (22, 30.14%), punctuation (16, 21.92%), and unintelligibility (17, 23.28%), with no critical errors (0%). Most ChatGPT errors are minor (71.1%), reflecting better fluency and readability. ChatGPT demonstrates significantly fewer and less severe fluency issues than Basheer's translation. Some of the fluency errors are included hereafter according to the dimensions of the selected scheme:

4.4.2.1 Grammar

Grammar errors are concerned with the naturalness and the fluency of translation. The forthcoming examples include grammatical errors of various severities. Table 20 illustrates this type of grammatical error that both Basheer, as in example (1), and ChatGPT, as in example (2) have produced.

Table 20

Examples of Grammatical Errors

No	Original text	Translation	Error description
1	When you sing, the hungry hears you with his stomach (Gibran, 1926, p. 52).	إذا غنيت للجائع سمعك بمعدته (p. 29).	The word order is changed, and the part of speech is changed. The verb "sing" in the original is intransitive and has no object. The translation alters the sentence structure, making singing directed only at the hungry, whereas the original includes both hungry and full listeners. This major error distorts the intended meaning.
2	In vain shall a poet seek the mother of the songs of his heart (Gibran, 1926, p. 22).	عبثا يبخت الشاعر عن أم لأغنيات قلبه (2025).	The prepositional direct object of the original starts with the preposition "of," while the translation employs the preposition "ل," which is the equivalent of the preposition "for." The translation changes "the mother" to "a mother," shifting the meaning from seeking a known person to an unknown one. This minor error slightly affects the intended meaning.

4.4.2.2 Punctuation

All the punctuation errors in both of the translations are of minor severity. Table 21 includes some examples of punctuation errors—Basheer's, as in example (1), and ChatGPT's, as in example (2).

Table 21

Examples of Punctuation Errors

No	Original text	Translation	Error description
1	Your saying to me, "1 do not understand you," is praise beyond my worth, and an insult you do not deserve (Gibran, 1926, p. 42).	قولك إنك لا تفهمني مديح لا أستحفه أنا، وإهانة لا تستحقها أنت (p. 25).	The quotation marks of the transposed indirect speech in the original are not used in the translation. It does not affect the meaning.
2	Life is a procession. The slow of	الحياة موكب: البطيء	The translation restructures the original three

(EISSN): 2708-5783 ISSN: 2707-5508 | Arts for Linguistic & Literary Studies, Faculty of Arts, Thamar University, Volume 7, Issue 3, September 2025

678

No	Original text	Translation	Error description
	foot finds it too swift and he steps out; And the swift of foot finds it too slow and he too steps out"(Gibran, 1926, p. 43).	يراها سريعة فيتنحى، والسريع يراها بطيئة فينصرف (2025).	sentences into compound-complex form, altering punctuation but not the meaning. These changes enhance stylistic clarity without distorting the poem's message.

4.4.2.3 Unintelligibility

Unintelligible errors are concerned with the clarity of meaning and naturalness of language. All the dimensions' errors produce them. The examples, in Table 22, represent some of these errors with severities——Basheer, as in example (1), and ChatGPT, as in example (2).

Table 22

Examples of Unintelligibility Errors

No	Original text	Translation	Error description
1	A sense of humor is a sense of proportion (Gibran, 1926, p. 14).	تعرف الفكاهة إذا عرفت اغتنام الفرص السانحة (p. 13).	The translation is meaningless, has no coherent meaning, and has nothing to do with the original text. It results from mistranslating the phrase "a sense of proportion." This critical error prevents the transmission of a message.
2	The most talkative is the least intelligent, and there is hardly a difference between an orator and an auctioneer (Gibran, 1926, p. 73).	الأكثر كلاما هو الأقل ذكاءً، ولا فرق كبير بين الخطيب والمزاد العلني (2025).	The meaning of the last stanza is partially unclear. The word auctioneer is translated المزاد العلني That is the equivalent of "auction" rather than auctioneer. It is a major error. The meaning is partially incorrect and incomplete, preventing comprehension.

4.4.2.4 Ambiguity

Ambiguity errors concern the clarity of meaning and the directness of the message. The following examples in Table 23 contain two examples of ambiguity errors of severities, example (1) for Basheer's translation, and example (2) for ChatGPT's translation.

Table 23

679

الآداب للدراسات اللغوية والأدبية، كلية الآداب، جامعة ذمار، المجلد7، العدد3، سبتمبر 2025 | **(EISSN): 2708-5783 ISSN: 2707-5508**

Examples of Ambiguity Errors

No	Original text	Translation	Error description
1	What shall I say of him who is the pursuer playing the part of the pursued (Gibran, 1926, p. 37).	ماذا أقول في المطارد الذي يلعب دور المطارد (p. 23).	The words "pursuer" and "pursued" are translated as "المطارد" without using inflectional markers to differentiate their cases. This critical severity error creates ambiguity, making it unclear who is the pursuer and who is the pursued, leading to multiple possible interpretations and reader confusion.
2	Had I filled myself with all that you know what room should I have for all that you do not know? (Gibran, 1926, p. 58).	لو ملأت نفسي بما تعرفه، فأين يكون موضع ما لا تعرفه (2025).	The letter "ه" in the word "تعرفه" can refer to "myself" as a possessive object of a genitive case and can refer at the same time to "you" as a direct object of an accusative case. It is a major error as the two meanings produced pun and ambiguity.

4.4.2.5 Spelling

Spelling errors are about the professionalism of translation. Some examples of spelling errors are included in Table 24——example (1) for Basheer's translation, and example (2) for ChatGPT's translation.

Table 24

Examples of Spelling Errors

No	Original text	Translation	Error description
1	I have never been envied nor hated; I am above no one (Gibran, 1926, p. 42).	ولكن لم يحسدني ولم يبغضني أحد قط، فأنا إذن لست فوق أحد (p. 25).	The word "so" is translated as "إذن," which is also a misspelling of the word "اذا."
2	But your other self grows on sorrow; so all is well (Gibran, 1926, p. 26).	لكنها تنمو بالحزن، فلا بأس إذن (2025).	The word "so" is translated as "إذن," which is also a misspelling of the word "اذا." It is identical to Basheer's above two errors, which are produced by misspelling the same word. It is an error of minor severity that does not affect the meaning.

4.4.2.6 Register

Register errors concern contextual apprehension, cultural orientation, and translator ideology. The following examples in Table 25 illustrate errors of varying severity——example (1) from Basheer's translation and example (2) from ChatGPT's translation.

Table 25

680

(EISSN): 2708-5783 ISSN: 2707-5508 | Arts for Linguistic & Literary Studies, Faculty of Arts, Thamar University, Volume 7, Issue 3, September 2025

Examples of Register Errors

No	Original text	Translation	Error description
1	Now let us play hide and seek. Should you hide in my heart it would not be difficult to find you. But should you hide behind your own shell, then it would be useless for anyone to seek you"(Gibran, 1926, p. 18).	هلم بنا نلعب لعبة "تخبأ مليح" ونفتش بعضنا عن بعض، فإن اختبأت في قلبي فليس بالصعب علي أن أجدك، ولكن إذا اختبأت وراء صدفتك فحينئذ عبثا يحاول الناس أن يهتدوا إلي (p. 15).	"تخبأ مليح" is the translation of "hide and seek", a children's game. The translation is not the Arabic name of the game which is "الغميضة", but it has given a name of national origin. It is selected out of a particular cultural frame. Thus, it reduces the overall understanding and apprehension of the text. It is a major error that has a negative impact on the text's readability and comprehension.
2	The Sphinx spoke only once, and the Sphinx said, "A grain of sand is a desert, and a desert is a grain of sand; and now let us all be silent again." I heard the Sphinx, but I did not understand (Gibran, 1926, p. 3)	تكلمت أبو الهول مرة واحدة فقط، وقالت: "حبة الرمل صحراء، والصحراء حبة رمل؛ فلنصمت جميعا من جديد." سمعت أبا الهول لكني لم أفهم (2025).	The use of the inflectional feminine pronoun "ت" in the words "قالت" and "تكلمت" is out of the cultural register. Arabic usually refers to the Sphinx with masculine pronouns and views it as a masculine symbol. The use of feminine is out of the Western cultural register. It is a minor error that does not negatively impact the text.

4.4.2.7 Style

Style errors are about the emotional effect, meaning depth, and appropriate selection of words. The following examples that Table 26 comprises highlight some of the style errors—

Table 26

Examples of Basheer's Style Errors

No	Original text	Translation	Error description
1	A pearl is a temple built by pain around a grain of sand. What longing built our bodies and around what grains? (Gibran, 1926, p. 4).	الدرة هيكل بناء الألم حول حبة رمل، فما هو الحنين الذي بنى أجسادنا؟ وما هي الحبوب التي بنيت حولها (p. 9).	The repetition of the words "بنى" affected the poetic style. The selection of the word "الحبوب" as equivalence for grains is not appropriate, as it is associated with cereals in the Arabic language. The appropriate word is "الحبات". It is a minor error that does not affect the language flow and fluency but makes the poem less poetic.

681

الآداب للدراسات اللغوية والأدبية، كلية الآداب، جامعة ذمار، المجلد7، العدد3، سبتمبر 2025 | **(EISSN): 2708-5783 ISSN: 2707-5508**

| 2 | Your mind and my heart will never agree until your mind ceases to live in numbers and my heart in the mist (Gibran, 1926, p. 30). | لن يتفق فكرك وقلبي حتى ينقطع فكرك عن أن يعيش بالأرقام، ويقف قلبي عن الحياة بالضباب (p. 21). | The use of the word "فكرك" as an equivalent of "mind" is not suitable in this context. The Arabic equivalence of the mind/heart binary is the "قلب/عقل". The use of "فكرك" lacks poetic quality and weakens emotional impact. Along with the literal translation of the rest, it causes confusion and obscures the poem's message——a critical error. |
| 3 | I had a second birth when my soul and my body loved one another and were married (Gibran, 1926, p. 5). | "قد ولدت ثانية عندما وقع جسدي بحب نفسي وتزوجا معا (p. 9). | The literal translation of the word "soul" as "نفس" is not appropriate for poetic language. The word "نفس" is associated with a religious register in Arabic. This is a major error that affects the meaning and reduces the readability of the text. |

4.4.2.8 Coherence and cohesion

Coherence and cohesion errors concern syntactic connectivity and semantic logical sequence. Some of the examples of these errors are in Table 27 below:

Table 27

Examples of Basheer's Coherence and Cohesion Errors

No	Original text	Translation	Error description
1	give me silence and I will outdare the night. And I had a second birth when my soul and my body loved one another and were married (Gibran, 1926, p. 5).	اعطني الصمت أقتحم غمرات الليل، قد ولدت ثانية عندما وقع جسدي بحب نفسي وتزوجا معا (p. 9)	The two separate poems are combined into a single unified poem. The two original poems have nothing in common and are about two different subject matters, but the translation blended them. The translation introduces incoherent, disconnected ideas and unrelated subjects, disrupting the poem's flow. This critical error severely affects readability and weakens the intended message.
2	When two women talk they say nothing; when one woman speaks she reveals all of life"(Gibran, 1926, p. 16). And " Frogs may bellow louder than bulls, but they cannot drag the plough in the field nor turn the wheel of the winepress, and of their skins you cannot make shoes (16).	إذا تكلمت امرأتان فهما لا تعلنان شيئا، وإذا تكلمت امرأة فهي نعلن الحياة كلها. قد يكون للضفادع أصواتٌ أعلى من أصوات البقر. ولكن الضفادع لا تستطيع أن تجرَّ السكة في الحقل، ولا أن تدير دولاب المعصرة ولا يمكنك أن تصنع من جلودها أحذية (pp. 14-13)	The blending of two separate poems into the form of a single unified poem is a coherent and cohesive error. The two poems are about two different subject matters. The first is about the value of love and the value of women's talk. The second is about the meaninglessness and emptiness of noise. The juxtaposition of the two poems has a negative impact on the first poem, as it compares the woman's talk to the noise of frogs. This error is critical, negatively affecting the message and distorting its meaning.

682

(EISSN): 2708-5783 ISSN: 2707-5508 | Arts for Linguistic & Literary Studies, Faculty of Arts, Thamar University, Volume 7, Issue 3, September 2025

The severity error distribution and analysis above revealed that ChatGPT and Basheer produced a large number of minor errors, with Basheer's errors being 10% more than those of ChatGPT. However, both translators produced almost equally major errors. Nonetheless, Basheer produced critical errors, whereas ChatGPT reported zero critical errors. Basheer's highest critical error rate was recorded in the mistranslation dimension, followed by omission and addition dimensions, respectively. The three dimensions are classified as high-risk error dimensions severely affecting translation quality. Basheer's highest major error rate is detected in the dimension of mistranslation, followed by the dimension of addition; both are high-risk dimensions. However, ChatGPT's highest rate of major errors was in the domain of overly literal followed by mistranslation. The overly literal dimension concerns metaphorical and stylistic use of language. Moreover, the highest rate of minor errors recorded by ChatGPT was also observed in the overly literal dimension, compared to Basheer's additional dimension. Eventually, critical and major errors are likely to appear in human translations more often than in AI translation models.

Eventually, Basheer's translation exhibits a broader distribution of errors across all severity levels, including several critical errors in fluency and accuracy. In contrast, ChatGPT's translation shows no critical errors and significantly fewer major ones. ChatGPT's output consistency and lower error rate highlight its strength in controlling and rendering text fluently. These findings validate the quantitative scoring outcomes and offer a clearer understanding of the comparative performance between human and AI-generated translations. Tables 1, 2, 3, 4, 5, 6, 7, 18and 19 will form the basis of the penalty calculations presented in the next section, providing detailed insight into the nature and severity of errors in both translations.

4.5 Translation Quality Scoring

This section evaluates the translation quality of Basheer and ChatGPT using a severity-weighted error framework based on MQM. Errors are categorized under accuracy and fluency; each is weighted using multipliers: Minor (1), Major (5), and Critical (25). Tables 28 and 29 comprised all-inclusive calculations of ChatGPT and Basheer's accuracy and fluency penalties. However, Table 30 include the measurement of quality scores of both translators.

Table 28

Accuracy Penalty (AP)

Category	Basheer (Minor×1 + Major×5 + Critical×25)	Total	ChatGPT (Minor×1 + Major×5 + Critical×25)	Total
Addition	161×1 + 18×5 + 2×25	301	5×1 + 0×5 + 0×25	5
Mistranslation	94×1 + 34×5 + 18×25	714	6×1 + 2×5 + 0×25	16
Overly Literal	27×1 + 1×5 + 0×25	32	44×1 + 5×5 + 0×25	69
Omission	33×1 + 1×5 + 5×25	163	9×1 + 1×5 + 0×25	14
Untranslated	0×1 + 0×5 + 0×25	0	0×1 + 1×5 + 0×25	5
Total AP		1210		109

683

الآداب للدراسات اللغوية والأدبية، كلية الآداب، جامعة ذمار، المجلد7، العدد3، سبتمبر 2025 | **(EISSN): 2708-5783 ISSN: 2707-5508**

Dr. Sadek Saleh Alsalemi, & Dr. Mofareh Dhaher Alhazmi, & Dr. Mustafa Ahmed Al-Humari

الآداب
للدراسات اللغوية والأدبية

Table 29

Fluency Penalty (FP)

Category	Basheer (Minor×1 + Major×5 + Critical×25)	Total	ChatGPT (Minor×1 + Major×5 + Critical×25)	Total
Grammar	54×1 + 4×5 + 0×25	74	10×1 + 0×5 + 0×25	10
Punctuation	77×1 + 0×5 + 0×25	77	16×1 + 0×5 + 0×25	16
Unintelligible	2×1 + 13×5 + 8×25	275	7×1 + 10×5 + 0×25	57
Ambiguity	1×1 + 1×5 + 4×25	106	2×1 + 2×5 + 0×25	12
Spelling	5×1 + 1×5 + 0×25	10	0×1 + 2×5 + 0×25	10
Register	9×1 + 5×5 + 0×25	34	2×1 + 0×5 + 0×25	2
Style	123×1 + 11×5 + 3×25	253	22×1 + 0×5 + 0×25	22
Coherence & Cohesion	67×1 + 8×5 + 4×25	207	0×1 + 0×5 + 0×25	0
Total FP		1036		129

Table 30

Overall Penalty and Quality Score

Translator	Basheer	ChatGPT
APT (AP+FP)	2246	238
Evaluation Word Count	5900	4900
Normalized Penalty	0.38	0.048
Raw Quality Score	0.62	0.952
Final Quality Score (%)	62	95.2

4.6 Scalar Quality Metric Result

The quantitative measures are indicators of the quality of translation in proportion to their qualitative aspects. It means that translation quality is less determined by the error count than by the severity of errors and more by error type than by error count. A closer look at the MQM's error typology of the translations under investigation provides us with outstanding stats and datasets that need to be underpinned with examples. However, Table 31 below includes the qualitative evaluation of the professional translator.

Table 31

Qualitative Evaluation by Human Annotators

Quality Level	Excellent	Good	Fair	Poor	Very Poor	Unacceptable	Total
Basheer	67	184	32	19	7	12	321
ChatGPT	263	42	10	5	0	1	321

684

(EISSN): 2708-5783 ISSN: 2707-5508 | Arts for Linguistic & Literary Studies, Faculty of Arts, Thamar University, Volume 7, Issue 3, September 2025

The mean quality score and the standard deviation are included in table 15 below:

Table 32

The Mean Quality Score and the Standard Deviation

Translator	Mean Score	Standard Deviation
Basheer	3.78	1.14
ChatGPT	4.75	0.636

The mean quality score of Basheer's translation is 3.78 out of 5, indicating a moderate or average translation quality that rates between "fair" (3) and "good" (4). The high standard deviation reflects a greater variability and diversity of the quality of the translation across the translated collection. The large standard deviation indicated the spread of the quality away from the mean, with some poems higher or lower than the average. However, the mean quality score of ChatGPT's translation is 4.75 out of 5, which indicates that the translation is of high quality. It rated between the "good" (4) and the "excellent" (5), with a noticeable leaning toward excellent. The small standard deviation of ChatGPT's translation signifies that the quality scores are tightly clustered around the mean, reflecting the consistency and stability of the quality across the translated collection.

The qualitative assessment revealed that ChatGPT's excellent rank is four times higher than Basheer's. The total number of ChatGPT's excellent poems is 263, representing 81.9% of the translated poems, compared to 67 poems by Basheer, representing 20.8%. Moreover, the total number of ChatGPT's translated poems evaluated under excellent and good translation categories is 305, which forms 95% of the poetry collection compared to 251 poems produced by Basheer, representing 78%.

The annotators' scores corresponded to the MQM's quantitative measures, which indicates that the quantitative measures are reliable indicators of translation quality.

5. Discussion

This study set out to explore the potential and limitations of artificial intelligence—specifically ChatGPT—in translating poetry from English to Arabic, using Khalil Gibran's *Sand and Foam* as the focal text. The findings reveal a striking contrast in error frequency, typology, and severity between human and AI-generated translations. These differences underscore the varying strengths and shortcomings of each method.

The data show that ChatGPT's translation achieved a substantially lower error rate (2.96%) compared to Basheer's human translation (13.7%), particularly in the domains of fluency and syntactic accuracy. This aligns with recent findings by Hendy et al. (2023) and Gao et al. (2024), who observed that large language models like ChatGPT outperform traditional MT systems in high-resource language pairs, especially in terms of grammaticality, cohesion, and sentence-level coherence. Moreover, ChatGPT demonstrated a consistent quality across poems, as reflected in its lower standard deviation and higher percentage of "Excellent" ratings in qualitative assessments.

However, these quantitative advantages were accompanied by interpretive shortcomings. ChatGPT's high incidence of "Overly Literal" errors (67%) suggests a default preference for direct linguistic equivalence over poetic

flexibility. This confirms earlier concerns by Chakrabarty et al. (2021) and Studzińska (2020), who noted that AI tends to prioritize surface-level fidelity at the expense of deeper metaphorical and cultural resonances. In contrast, Basheer's human translation, while error-prone, often reflected a greater sensitivity to poetic style and cultural framing—despite introducing more additions and mistranslations.

The high number of critical errors in Basheer's work—particularly omissions of entire poems—raises questions about editorial integrity and consistency. These omissions significantly undermine the completeness and philosophical unity of Gibran's collection. In this respect, ChatGPT's consistent poem-by-poem structure demonstrates its advantage in maintaining the integrity of the source material, even if poetic nuance is occasionally sacrificed.

Furthermore, ChatGPT's zero incidence of critical errors across all MQM dimensions reveals a level of reliability and risk aversion not found in the human translation. This supports findings by Alowedi & Al-Ahdal (2023), who suggested that post-edited AI translations could be more reliable for conveying structured poetic content. Additionally, ChatGPT's superior performance in cohesion and discourse-level grammar suggests that it is increasingly capable of managing macro-level textual relationships—traditionally considered a human domain (Läubli et al., 2020; Toral et al., 2018).

Despite these strengths, AI remains limited in cultural contextualization and affective expression. For example, register mismatches and awkward metaphorical translations suggest that ChatGPT's understanding of culturally loaded or emotionally complex content is still developing. These limitations reinforce the argument advanced by Bassnett (2014) and Bahrami (2012), who argue that poetry translation is not merely a linguistic act but an interpretive, creative, and cultural endeavor—one that resists full automation.

Taken together, the findings suggest that while AI demonstrates high fluency, structural fidelity, and error-minimization, it still struggles with poetic expressiveness, cultural subtlety, and emotional depth. This confirms the view that AI should be considered a powerful—but partial—agent in literary translation. As such, a hybrid model that combines AI's technical strengths with human creativity and cultural insight may offer the most promising path forward.

6. Conclusion

The development of large language models (LLMs) has led to high-quality translations that often match human efforts. Notably, ChatGPT has demonstrated a remarkable ability to produce poetic translations that can surpass those of human translators. This study compared ChatGPT's performance with that of a human translator in translating poetry, specifically examining accuracy and fluency. The results highlighted ChatGPT's proficiency in translating free verse poetry into Arabic.

AI models, such as ChatGPT, have emerged as benchmark tools in poetry translation, exhibiting superior linguistic accuracy, grammatical integrity, and textual cohesion compared to traditional translations. The findings suggest that AI can produce efficient and stylistically coherent translations that effectively align with the structural and semantic aspects of poetic texts. However, while AI excels in formal accuracy and fluency, it struggles with the interpretive sophistication and emotional nuance found in human translation. Challenges persist in navigating metaphorical ambiguity, cultural allusion, and poetic improvisation—essential elements of the literary experience.

686

(EISSN): 2708-5783 ISSN: 2707-5508 | Arts for Linguistic & Literary Studies, Faculty of Arts, Thamar University, Volume 7, Issue 3, September 2025

ChatGPT's tendency toward literalism and occasional lack of clarity in dense passages reflect its limitations in artistic sensibility.

The implications of these findings affirm that AI translation is a powerful supplementary tool for literary translation. Moreover, they caution against relying solely on AI for final poetic renderings, especially in contexts where cultural sensitivity and creative expression are paramount. The study's statistical analysis substantiated ChatGPT's remarkable capacity to translate poetry within specific poetic contexts characterized by a low-metaphorical language and low figurative style. Eventually, a new translation of Khalil Gibran's *Sand and Foam* is a burning issue as the translation of Antonious Basheer did not meet the requirements of the poetic translation of one of the greatest and most influential Arabic poets.

Ultimately, this research advocates for a hybrid translation model that combines AI's efficiency with human translators' creativity and depth. Future research should focus on enhancing AI's literary sensibilities through collaboration, algorithmic training, and editorial oversight, enabling a more enriched future for poetic translation.

References

AlAfnan, M. A., & Alshakhs, T. (2025). Bridging linguistic and cultural nuances: A comparative study of human and AI translations of Arabic dialect poetry. *Advances in Artificial Intelligence and Machine Learning*, *5*(1), 3236–3260. https://doi.org/10.54364/AAIML.2025.51186

Algobaei, F., Alzain, E., Naji, E., & Nagi, K. A. (2024). Gender Issues between Gemini and ChatGPT: The Case of English-Arabic Translation. *World Journal of English Language*, *15*(1), 9. https://doi.org/10.5430/wjel.v15n1p9

Alowedi, N., & Al-Ahdal, A. (2023). Artificial intelligence based Arabic-to-English machine versus human translation of poetry: An analytical study of outcomes. *Journal of Namibian Studies: History Politics Culture*, (33). https://doi.org/10.59670/jns.v33i.800

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

Bahrami, N. (2012). Strategies used in the translation of allusions in Hafiz Shirazi's poetry. *Journal of Language and Culture*, *3*(1), 1–9. https://doi.org/10.5897/JLC11.058

Balasubramanian, S. (2023). Exploring the capabilities of ChatGPT in natural language processing tasks. *Journal of Artificial Intelligence and Machine Learning*, *2*(1), 717. https://doi.org/10.17605/OSF.IO/XJYMQ

Bassnett, S. (2014). The translator as mediator of poetic meaning. *Meta: Journal des Traducteurs*, *59*(1), 1–15.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, (33), 1877–1901.

687

الآداب للدراسات اللغوية والأدبية، كلية الآداب، جامعة ذمار، المجلد7، العدد3، سبتمبر 2025 | **(EISSN): 2708-5783 ISSN: 2707-5508**

Buçpapaj, E., & Koni, E. (2024). "The Art of Fidelity and Creativity in Literary Translation. *Scope Journal, 14*(04), 974—985.

Cai, Z. G., Duan, X., Haslett, D. A., Wang, S., & Pickering, M. J. (2023). *Do large language models resemble humans in language use?* arXiv preprint arXiv: 2303.08014.

Chakrabarty, T., Saakyan, A., & Muresan, S. (2021). *Don't go far off: An empirical study on neural poetry translation.* In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (pp. 577). https://doi.org/10.18653/v1/2021.emnlp-main.577

Christiano, P. F., Leike, J., Brown, T. B., Martic, M., Legg, S., & Amodei, D. (2017). Deep reinforcement learning from human preferences. In *Advances in Neural Information Processing Systems 30* (NIPS 2017) (pp. 4302—4310). Curran Associates, Inc. https://doi.org/10.5555/3294996.3295184

Dai, J., Shen, H., et al. (2022). The differences between machine translation and human translation from the perspective of literary texts. *International Journal of Arts and Social Science*, *5*(10), 112—133.

Edgerton, D. (2023). *A brief history of tech skepticism*. Strategy+Business. https://www.strategy-business.com/article/A-brief-history-of-tech-skepticism

Fasiullah, S. M. (2019). *Challenges in translating poetry: A study of two English poems translated in Urdu by Muhammad Iqbal.* Universal Review.

Freitag, M., Foster, G., Grangier, D., Ratnakar, V., Tan, Q., & Macherey, W. (2021). Experts, errors, and context: a large-scale study of human evaluation for machine translation. *Transactions of the Association for Computational Linguistics, 9*, 742—758. https://doi.org/10.1162/tacl_a_00437

Floridi, L. (2019). Artificial intelligence, human intelligence, and the future of work. *Philosophy & Technology, 32*(4), 545—559. https://doi.org/10.1007/s13347-019-00351-0

Frost, W. (1969). *Dryden and the art of translation.* Yale University Press.

Gao, R. Y., 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), Article 835. https://doi.org/10.1057/s41599-024-03363-0

Ghazvininejad, M., Choi, Y., & Knight, K. (2018). Neural poetry translation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)* (pp. 67—71). Association for Computational Linguistics. https://doi.org/10.18653/v1/N18-2011

Gibran, K. (1926). *Sand and foam.* Alfred A. Knopf.

Gibran, K. G. (1926—2020). *Sand and foam* (A. Basheer, Trans.). Hindawi.

Hendy, A., Abdelrehim, M., Sharaf, A., Raunak, V., Gabr, M., Matsushita, H., & Awadalla, H. H. (2023). *How good are GPT models at machine translation? A comprehensive evaluation*. arXiv preprint arXiv:2302.09210.

Humblé, P. (2019). Machine translation and poetry: The case of English and Portuguese. *Ilha do Desterro*, (72), 41—56. https://doi.org/10.5007/2175-8026.2019v72n2p41

688

(EISSN): 2708-5783 ISSN: 2707-5508 | Arts for Linguistic & Literary Studies, Faculty of Arts, Thamar University, Volume 7, Issue 3, September 2025

Jiao, W., Wang, W., Huang, J. T., & Wang, X., Tu, Z. (2023). *Is ChatGPT a good translator? A preliminary study*. arXiv preprint arXiv:2301.08745.

Jones, F. R. (2011). Translating poetry: The role of the translator as interpreter and creator. *Translation Studies*, *4*(2), 177–190. https://doi.org/10.1080/14781700.2011.584099

Khalifa, A. A. (2015). Translation studies: Some problematic aspects of Arabic poetry translation. *International Journal of Sciences: Basic and Applied Research (IJSBAR)*, *19*(1), 314–324.

Kuzman, T., Vintar, Š., & Arcan, M. (2019). *Neural machine translation of literary texts from English to Slovene.* In Proceedings of the qualities of literary machine translation (pp. 1–9). https://aclanthology.org/W19-7301.pdf

Lahiani, R. (2022). Aesthetic poetry and creative translations: A translational hermeneutic reading. *Humanities and Social Sciences Communications, 9*(1), 1–9. https://doi.org/10.1057/s41599-022-01481-1

Lommel, A. (2018). Metrics for evaluating translation quality: A case for standardising error classifications. In *Evaluation of Translation Quality* (pp. 109–127). Springer. https://doi.org/10.1007/978-3-319-91241-7_6

Lommel, A., Uszkoreit, H., & Burchardt, A. (2014). Multidimensional quality metrics (MQM): A framework for specification describing metrics for translation quality. *Tradumàtica, 12*, 455–463. https://doi.org/10.5565/rev/tradumatica.Issue23

Lefevere, A. (1992). *Translation, rewriting, and the manipulation of literary fame*. Routledge.

Läubli, S., Castilho, S., Neubig, G., et al. (2020). A set of recommendations for assessing human–machine parity in language translation. Journal of Artificial Intelligence Research, 67. https://doi.org/10.1613/jair.1.11371

Ma, Y., & Wang, B. (2020). Description and quality assessment of poetry translation: Application of a linguistic model. *Contrastive Pragmatics, 3*(1), 89–111. https://doi.org/10.1163/26660393-bja10015

Mariana, V. T., Cox, T., & Melby, A. (2015). The multidimensional quality metrics (MQM) framework: A new framework for translation quality assessment. *The Journal of Specialised Translation, 23*, 137–161. https://doi.org/10.26034/cm.jostrans.2015.343

Mittelstadt, B. D. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence, 1*(11), 501–507. https://doi.org/10.1038/s42256-019-0114-4

Moslem, Y., Haque, R., & Way, A. (2023). *Adaptive machine translation with large language models.* arXiv preprint arXiv:2301.13294.

Motair, A. A. A., Algobaei, F., & Alhazmi, M. D. (2025). Ethics in Translation: A Pathway to Integrity in Future Professionals. *Arts for Linguistic & Literary Studies, 7*(1), 711–732. https://doi.org/10.53286/arts.v7i1.2417

Naghiyeva, S. B. (2015). *Does poetry lose or gain in translation? English Language and Literature Studies, 5*(3). Canadian Center of Science and Education.

Nagi, K.A., Alzain, E., Naji, E., 2024. Informed prompts and improving ChatGPT English to Arabic translation. *Al-Andalus Journal for Humanities & Social Sciences. 98*(11). https://doi.org/10.35781/1637-000-098-007

Nair, S. K. (2018). Is poetry lost in translation? *Samyukta: A Journal of Gender and Culture, 3*(1), 35–42. https://doi.org/10.53007/SJGC.2018.V3.I1.118

689

الآداب للدراسات اللغوية والأدبية، كلية الآداب، جامعة ذمار، المجلد7، العدد3، سبتمبر 2025 | (EISSN): 2708-5783 ISSN: 2707-5508

Newmark, P. (1991). *About translation.* Multilingual Matters.

Nida, E., & Taber, C. (1982). *The theory and practice of translation*. Brill.

Niknasab, L. (2011). *Translation and culture: Allusions as culture bumps. SKASE Journal of Translation and Interpretation, 5*(1), 45–54. Retrieved from https://www.skase.sk/Volumes/JTI05/pdf_doc/03.pdf

Ono, K. (2019). Replacement of the military's intellectual labor using artificial intelligence —— Discussion about AI and human co-existence. *National Institute for Defense Studies Bulletin*, (1), 3–20. https://www.nids.mod.go.jp/english/publication/kiyo/pdf/2019/bulletin_e2019_1.pdf

Othman, A. A. M. (2023). Cohesion and coherence for poetry interpretation and translation. *AWEJ for Translation & Literary Studies, 7*(2), 176–196. https://doi.org/10.24093/awejtls/vol7no2.13

Ovidiu, M. (2008). *Translating poetry: Contemporary theories and hypotheses. Professional Communication and Translation Studies, 1.* Lucian Blaga University of Sibiu. https://doi.org/10.59168/YDRF2520

Peng, N., Li, W., Yang, B., & Li, Z. (2023). Exploring ChatGPT for automatic poetry generation. *Journal of Artificial Intelligence Research, 72*, 215–237. https://doi.org/10.1613/jair.1.13208

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI blog, 1*(8), 9.

Radmir, K., Anton, K., Aleksandr, A., Dmitry, A., & Oleg, V. (2024). Comparison of ChatGPT and Bard for using in hybrid intelligent information systems. E3S Web of Conferences. https://doi.org/10.1051/e3sconf/202454908009

Raffel, B. (2010). *The Art of Translating Poetry.* Penn State Press.

Schulman, J., Zoph, B., Kim, C., Hilton, J., Menick, J., Weng, J., & Ryder, N. (2022). *ChatGPT: Optimizing language models for dialogue*. OpenAI blog. https://openai.com/blog/chatgpt

Seljan, S., Dunder, I., & Pavlovski, M. (2020). *Human quality evaluation of machine-translated poetry*. In *2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO)* (pp. 1040–1045). IEEE. https://doi.org/10.23919/MIPRO48935.2020.9245436

Soong, S. C. (1973). Notes on translating poetry. In W. Arrowsmith & R. Shattuck (Eds.), The craft and context of translation (pp. xx–xx). Doubleday Anchor Books.

Studzińska, J. (2020). Turing test for (automatic) translation of poetry (Polish). *Porównania*, (26), 299–313. https://doi.org/10.14746/por.2020.1.17

Tahir, E. M. (2008). *Strategies for translating poetry aesthetically*. https://doi.org/10.13140/RG.2.2.22475.28961

Toral, A., Castilho, S., Hu, K., et al. (2018). *Attaining the unattainable? Reassessing claims of human parity in neural machine translation*. In Proceedings of the Third Conference on Machine Translation: Research Papers. https://doi.org/10.18653/v1/w18-6312

Tymoczko, M. (2017). Expanding translation, empowering translators. *The Translator, 23*(1), 1–15. https://doi.org/10.1080/13556509.2017.1288257

Vardaro, J., Schaeffer, M., & Hansen-Schirra, S. (2019). Translation quality and effort prediction in professional machine translation post-editing. In M. Carl, M. Schaeffer, & S. Hansen-Schirra (Eds.), *Proceedings of the Second*

690

(EISSN): 2708-5783 ISSN: 2707-5508 | Arts for Linguistic & Literary Studies, Faculty of Arts, Thamar University, Volume 7, Issue 3, September 2025

MEMENTO Workshop on Modelling Parameters of Cognitive Effort in Translation Production (pp. 7—8). European Association for Machine Translation. https://doi.org/10.18653/v1/W19-7004

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, Ł., & Polosukhin, I. (2017). *Attention is all you need.* In Advances in Neural Information Processing Systems, 30, 5998—6008.

Venuti, L. (1998). Poetry and translation. *The Translator, 4*(2), 145—166. https://doi.org/10.1080/13556509.1998.10799088

Venuti, L. (2000). *The translation studies reader.* Routledge.

Wang, L. (2023). The impacts and challenges of artificial intelligence translation tool on translation professionals. SHS Web of Conferences. https://www.shsconferences.org/articles/shsconf/pdf/2023/12/shsconf_icssed2023_02021.pdf

West, D. M. (2018). *The future of work: Robots, AI, and automation.* Brookings Institution Press.

Zequeira, M. (2024). *Artificial intelligence as a combat multiplier: Using AI to unburden army staffs.* Military Review Online Exclusive. U.S. Army University Press.

Zhang, Y., Sun, A., Rao, J., & Huang, J. (2023). Evaluating large language models in creative writing: The case of story generation. *Transactions of the Association for Computational Linguistics, 11,* 123—138. https://doi.org/10.1162/tacl_a_00434

Zhang, Y., & Wang, L. (2024). Machine translation of Chinese classical poetry: A comparison of ChatGPT and traditional systems. *Humanities and Social Sciences Communications,* (11), Article 363.

❈ ❈ ❈

691

الآداب للدراسات اللغوية والأدبية، كلية الآداب، جامعة ذمار، المجلد7، العدد3، سبتمبر 2025 | **(EISSN): 2708-5783 ISSN: 2707-5508**