

Evaluating LLMs' Arabic-English Translation of Metaphor in Al-Ashmawi's Poetry

Nabil Amin Mohammed Ahmed Kassim

Assistant Professor of Applied Linguistics, Faculty of Education, Ibb University, Yemen

Email: nabilaminm2026@gmail.com

How to cite:

Kassim, N. A. M. A. (2026). Evaluating LLMs' Arabic-English Translation of Metaphor in Al-Ashmawi's Poetry. *International Journal of Linguistics and Translation Studies*, 7(3).384-305. <https://doi.org/10.36892/ijlts.v7i3.781>

ARTICLE HISTORY

Received:
15/08/2026

Accepted:
15/09/2026

Keywords:

Translation;
Arabic
Poetry;
Metaphor;
LLMs;
Human
Translators;
Al-Ashmawi

Abstract

The current study aims to evaluate LLMs' effectiveness of metaphor translation in the poetry of Al-Ashmawi, one of the prominent poets in the Kingdom of Saudi Arabia, in comparison to that of human translators. It also aims to identify LLMs' potential, strengths, and limitations when translating complex literary forms and show the significance of humans' expertise in refining LLMs' translation. For data collection, a sample of 20 poetic verses was selected purposefully, after ensuring their validity and suitability for the study objectives, from the poem of "Fortitude in the Time of Fracture". The mixed methods design (quantitative and qualitative) was adopted for data analysis with the help of the comparative analytical approach where the translation quality of both systems was evaluated by four Yemeni professors specialized in literature and translation, using Likert five-point scale, then analyzed descriptively through SPSS and Multidimensional Quality Metrics (MQM) model, focusing on three parameters: linguistic accuracy, poetic fidelity, and emotional impact. The findings showed that human translators consistently outperformed LLMs, particularly in areas that require cultural sensitivity, cognitive and creative interpretation, stylistic coherence, and emotional resonance. In comparison, though LLMs outperformed human translators in preserving lexical and semantic features and produced an accurate readable translation, they simplified and literalized metaphor translation, resulting in stripping the poem of its aesthetic beauty and artistic weight. The study called for further research on training LLMs in computational linguistics and highly symbolic poetic language besides using a hybrid approach that combines LLMs' efficiency and humans' creativity to achieve contextually relevant, linguistically accurate, and artistically authentic literary translation.

1. Introduction

In the age of global communication and technological advancements, translation process has undergone a major transformation through the integration and use of LLMs in translation. Traditional manual translation has been supplemented and replaced by advanced LLMs, such as ChatGPT and DeepSeek which are widely used by millions of people every day (Amroosh, 2025). These models provide rapid and linguistically accurate outputs though their reliability, efficiency, and potential to replace human translators remain debatable as their translation is

not error-free, motivating researchers to focus on improving their translation quality, reducing their translation errors, and enhancing their cultural and stylistic sensitivity to source texts (Al-Khesali, 2022). This endeavor has been challenged by the Arabic poetic language translation which is full of rhetorical images, idiomatic expressions, and contextual and cultural references, leading to some doubts regarding LLMs' efficiency in translating poetic language (Mohsen, 2023).

Literary translation is not a mere word for word transfer among languages; rather, it is a complex process that works as a bridge between languages and cultures and seeks to convey profound meaning, tone, effect, and subtleties of literature faithfully as possible (Gracia, 2022). Poetry translation is among the most demanding forms of literary translation due to its dependence on aesthetic and emotional qualities, layered meanings, and cultural depth (Alotaibi, 2021; Omar, 1997). Unlike prose, poetry utilizes imagery, symbolism, metaphor, and rhythm which often resist literal translation and require creative interpretation. Moreover, poetry translation poses further challenges due to its rhetorical elegance, intricate syntax, referential and aesthetic purposes, and culturally driven contexts (Farghal & Haider, 2024; Tisgam, 2014; Van der Meer, 2019).

Regarding metaphor translation, it is not a simple substitution of words but a complex cognitive mapping that reflects how a culture perceives reality. Its translation requires sensitivity to layered meanings, artistic imagery, and cultural subtleties (Mahmoud, 2018; Sofi, 2018). It requires recreating the same aesthetic and emotional effect on target texts readers, maintaining cultural specialities, classical references, and figurative expressions, and preserving the soul of literary heritage across linguistic boundaries (Al-Hasnawi, 2007; Harmah, 2023). These challenges demand a deep cultural, referential, contextual, and linguistic knowledge, and artistic integrity that LLMs lack, making their translation linguistically correct but aesthetically impoverished (Farghal & Haider, 2024).

Many different approaches to machine translation have been used in translation, starting with the rule-based translation in 1960s, moving to statistical machine translation in 1980s, then to neural machine translation in 2016, and ending with generative artificial intelligence represented by LLMs in 2022. While these approaches have made a great breakthrough in handling complex language translation, LLMs have achieved more fluent and coherent translation than earlier approaches though they still encounter some translation difficulties with idioms, figurative expressions, and culturally specific references (Elmadany & Wayne, 2018; Tawfik et al., 2022). As a result, many researchers, increasingly, advocate a hybrid approach that combines LLMs' efficiency with human creativity, confirming that human translation remains superior in accuracy, readability, artistic fidelity, and preservation of symbolic meanings. Also, human translators have the ability to question LLMs' outputs and choices and refine the translated materials until it meets a professional standard of translation (Abdelaal & Alazzawie, 2020; Oudah & Shaalan, 2020). This perspective is supported by broader research on technology adoption in translation, which highlights that professional translators' decisions to integrate AI tools are often mediated by concerns about accuracy and the need for human oversight, especially when dealing with culturally nuanced texts (Alrefaee & Al-Shammari, 2026). The evolution of technology-mediated learning in pragmatics similarly underscores that while digital tools offer new opportunities, they are most effective when combined with human expertise and pedagogical intervention (Alrefaee & Al-Shammari, 2026).

In the same context, though LLMs have improved translation accuracy through extensive training on large data sets, they often fail to preserve poetic images, such as poetic imagery, linguistic ambiguity, and stylistic rhythm, resulting in the loss of metaphorical essence and artistic effect of the poem (Al-Khalafat & Haider, 2022; Ghareb & Elimam, 2018; Malmkjær, 2020). These limitations were attributed to the huge amount of translated data by LLMs in comparison to the human mind, which deals with small amounts of data, especially in Arabic poetry where metaphor functions not merely as a decorative but also as a central device that conveys emotional, cultural, and conceptual depth (Wilson, 2014). Mariam (2024) also argued that metaphor is associated with human cognition, its translation requires cultural sensitivity and interpretive flexibility, qualities that LLMs struggle with, and Schaffner (2004) confirmed that metaphors rooted in cultural or historical contexts are often stripped of their significance by LLMs. These shortcomings are most found in Arabic poetry where metaphors often carry multiple interpretations, cultural symbolism, and religious or historical allusions that LLMs cannot adequately handle (Bergsma, 2021). This gap demonstrates the necessity of humans' expertise in evaluating and refining LLMs' generated poetic translation. In the words of Luo (2018, p. 2), "the performance of machine translator can never be primary and more creative than that of human translators". Therefore, human involvement in the translation process becomes indispensable for tasks like disambiguation and interpretation of idiomatic expressions to ensure cultural sensitivity among other complex linguistic and semantic aspects (Mirawi, 2014). This is further corroborated by recent cross-national studies on AI translation tool adoption among Arab translators, which revealed that concerns about accuracy, confidentiality, and cultural misinterpretation are significant barriers to adoption, reinforcing the need for human oversight (Alrefaee & Al-Shammari, 2026). The professional impact of AI, particularly its perceived threat to the nuanced and culturally sensitive nature of translation work, makes human intervention not just beneficial but essential (Alrefaee & Al-Shammari, 2026).

In summary, although LLMs have significantly advanced translation practices, they still struggle with artistic and cultural dimensions of poetry, especially metaphor translation (Kadaoui et al., 2023; Omar & Gomaa, 2020). The absence of cultural intuition and awareness of referential and social contexts in LLMs intelligence makes them doubtful tools for translating sensitive literary heritage. To date and to the best of the researcher's knowledge, no study has specifically examined how LLMs handle the metaphor translation in the literary works of Al-Ashmawi, a distinguished Saudi poet who left a lasting treasure of literary genre, where his contributions reflect deep intellectual and cultural values. Therefore, this study aims to address this gap by evaluating the metaphor translation in his poetry using LLMs and comparing their performance with that of human translators. By doing so, it contributes to enhancing computational linguistics and translation studies and highlights the potential and limitations of LLMs in handling symbolic and poetic language. Furthermore, it advocates for a collaborative approach where LLMs' generated translation should be edited by human expertise to ensure linguistic precision and artistic integrity in target texts, where human cultural awareness and creative sensitivity are still fundamental components even in the era of advanced technology (Pan et al., 2024).

1.1. Study Questions

The current study attempted to answer the following questions:

- (1) How effective are LLMs in translating metaphors in Al-Ashmawi's poetry from Arabic into English in comparison to that of human translators?
- (2) What are the most common errors in LLMs' outputs in comparison to that of human translators?

1.2. Study Objectives

The current study attempted to achieve the following objectives:

1. Evaluating the effectiveness of LLMs in translating metaphors in Al-Ashmawi's poetry from Arabic into English in comparison to that of human translators.
2. Pointing out the most common errors in LLMs' outputs in comparison to that of human translators when translating metaphors in Al-Ashmawi's poetry from Arabic into English.

1.3. Significance of the Study

This study is significant as it evaluates LLMs' metaphor translation in the poetry of one of the greatest Arab poets, whose work conveys profound emotions of Arabic and Islamic cultural and historical heritage and poses considerable translation challenges. It is also significant as it benefits translators, scholars, LLMs developers, researchers, and educational institutions and expand their understanding of literary translation through LLMs. This study is also significant as it shows the extent to which LLMs could achieve an accurate poetic translation of Arabic poetry and reveals their contributions, limitations, and implications. Finally, it is significant as it sheds light on the effectiveness and potential of human translators' cultural, contextual, and referential knowledge and expertise in editing translation which must complement LLMs' pitfalls.

2. Review of Related Literature

Poetry is "an imaginative way of expressing a poet's feelings, experiences, customs, traditions, connotations, implications, associations, thoughts, universal truths, and hidden meanings of life in words" (Fairouz, 2024, as cited in Nair, 1991, p.93). Linguistically, it has symbolic, allegorical, connotative, and emotive meanings which represent the intended hidden messages to audience, forming the main concern of critics, analysts, researchers, etc. Functionally, it involves literary devices, such as metaphors, similes, alliteration, assonance, collocations, neologism, etc. that make it aesthetic to readers. Poetically, it shows prosodic features, such as feet, rhyme, metre, etc. which make it musical to audience (Newmark, 1988).

Translating poetry has been recognized as a highly demanding task in which maintaining the original meaning of the source text faithfully is particularly difficult even with the advancements of LLMs and human expertise (Enad, 2018; Kahrizsangi & Haddadi, 2017). Ahmed et al. (2025) highlighted the ongoing cultural and semantic translation challenges, especially when dealing with figurative and poetic language. Omar and Gomaa (2020) and Enad (2018) identified recurrent machine translation issues in literary texts, particularly in pragmatic and lexical choices besides structural weaknesses and showed that statistical and neural translation models still lack the deep cultural understanding needed to, accurately, convey subtle meanings of the source texts without any errors in translation (Lu et al., 2024).

LLMs are designed to trigger human intelligence through algorithms and programming, allow them to process information, translate materials, and act in ways similar to humans

(Lehka-Paul & Jankowiak, 2022). For Zhong et al. (2023) LLMs are comparable to junior-level human translators in the area of human quality translation. ChatGPT as an LLM was introduced in 2022, a system built on large-scale linguistic data and patterns, commonly referred to as an (LLM). By analyzing syntactic, morphological, logical, and algorithmic structures, ChatGPT could produce new outputs while adapting to contextual backgrounds when performing tasks (Çetin & Duran, 2024). Its capabilities extend to data transformation, text translation, post-editing, translation assessment, summarization, and interactive responses (Ibrahim, 2024). It could also generate outputs across different modes: written, visual, and auditory, that closely resemble human production. Similarly, DeepSeek is a transformative LLM that supports a wide range of tasks, including mathematical and structural reasoning, language processing, problem-solving, financial analysis, and medical diagnosis. It has been extensively a pre-trained tool on high-quality linguistic and syntactic corpora including code-based data sets (Gao, 2024). Though DeepSeek faces certain limitations in creative practices, it demonstrates significant advantages in grammatical precision and ability to evaluate and translate contexts effectively (Lu et al., 2024).

Furthermore, many studies highlighted a continuing tension between the efficiency of LLMs and professional experts' translation skills and showed a growing interest in comparing the benefits and the drawbacks of both humans and intelligence-based translation systems. They showed that there are some challenges LLMs face in literary genre translation, especially with complex poetic language, connotative meaning, and cultural particularities (Bergsma, 2021; Gracia, 2021; Kahrizsangi & Haddadi, 2017; Khalifa, 2015; Malmkjær, 2020). Al Kassab (2024) analyzed the benefits and the drawbacks of both systems and showed that there are some strengths and weaknesses of each system in various contexts, especially the subtleties of poetic language and emotional weight of poetry. Manetti and Rossi (2022) assessed LLMs' literary translation and showed its dependence on limited reference corpora as a major challenge and called for an extensive human validation and evaluation due to the ambiguity inherent in figurative language. Jiao et al., (2023) and Kadaoui et al., (2023) also showed that though LLMs have made strides in translating complex literary genre, they tend to focus on statistical patterns rather than preserve stylistic nuances, confirming that literary works translation requires creativity, interpretive ability, comprehensive understanding of multiple meanings of texts, and understanding poets' intentions, qualities that LLMs still consider difficult to deal with. Finally, Toral and Way (2015) argued that literary texts often require more dynamic and communicative translation which can confuse algorithms alignment in LLMs.

With respect to translation strategies used by both systems for metaphor translation, human translators used the traditional methods, involving replicating the original imagery and converting metaphors into similes or literal meanings, or using strategies such as omission, adaptation, or compensation when necessary, whereas LLMs tend to focus on literal word replacement and neglect cultural context and poetic imagery (Inaam, 2003; Khoshafah, 2023). In addition, though LLMs perform well in technical translation, they often struggle to preserve figurative, stylistic, and cultural features of Arabic poetry, especially in metaphor translation (Alotaibi & Hussein, 2017; Omar, 1991), attributing such struggles to poetry's heavy reliance on figurative language, rhythm, contextual nuances, and cultural references, areas where LLMs often fall short (Toral & Way, 2018; Venuti, 2017). Tymoczko (2007) and Al-Buqmi (2024) argued that translating poetry is not just a linguistic act but also a cultural negotiation where

LLMs' translation often sounds mechanical and fails to capture the emotional and artistic essence of the original text.

Other studies (e.g., Akasheh et al., 2024; Al-Batineh & Rabadi, 2019) confirmed that literary translation requires cognitive imagination and interpretive sensitivity that exceed LLMs' current capacities, and (Abdelaal & Alazzawie, 2020; Khoshafah, 2023; Putri & Setiajid, 2021) demonstrated that LLMs frequently fail to maintain the figurative and stylistic interpretation of Arabic poetry as they rely on literal word replacement, overlook imagery, and ignore essential components for conveying poetic resonance, such as symbolism and referential or cultural subtleties of the poem.

Moreover, Khalifa (2015) and Malika (2016) highlighted the inherent challenges of translating Arabic poetry into English and showed that metaphors and imagery are deeply tied to cultural interpretation, confirming that any shifts in cultural background can alter meanings and connotations of the source text. Kadaoui et al. (2023) also showed that ChatGPT and Bard faced significant challenges when translating dialects due to limited training data, leading to frequent lexical and semantic errors. These challenges become more evident with cultural diversity and metaphorical complexity. As a result, human translators remain more adept at handling idioms, cultural allusions, and specialized terminology (Abdelaal & Alazzawie, 2020; Putri & Setiajid, 2021).

Alowedi and Al-Ahdal (2023) argued that poetic devices and figurative language remain particularly difficult for LLMs to capture, resulting in the loss of metaphorical imagery and symbolism. Similarly, Genzel et al. (2010) showed that machine translation frequently fails to reflect a poem tone and author's stylistic intent, thus weakening its outputs. Tawfik et al. (2022) also introduced multidimensional evaluation strategies for evaluating machine translation of poetry, including fluency, semantic similarity, faithfulness, and artistic merits and concluded that while these strategies improved some aspects, they still struggle with stylistic and emotional subtleties.

To overcome LLMs' translation challenges of literary forms, some researchers tried their best to find strategies to develop LLMs' quality translation. Al-Ghamdi et al. (2023), for example, employed an annotated poetry to train LLMs on translation strategies, and Oudah and Shaalan (2020) suggested a hybrid approach that integrates rule-based and statistical models with human editing. Similarly, Al-Zoghby and Elshamy (2021) merged automated analysis with human review to assess the readability of machine-translated Arabic poetry. Likewise, Elmadany and Wayne (2018) developed a specialized poetry corpus and fine-tuned translation systems. And, Amarín and AlNajjar (2021) applied neural machine translation to monolingual poetry collections and showed the artistic translation needs of parallel dataset. To support, Castilho et al. (2018) emphasized the need for approaches that move beyond simply identifying problems to proposing strategies for enhancing LLMs' ability to translate poetry. The results of such studies showed a better preservation of nuances than baseline models despite some limitations in fluency and cultural references, underscoring the indispensable role of human editing in ensuring comprehensive assessments of LLMs translation quality to achieve linguistic accuracy, cultural sensitivity, and stylistic refinement of translation and confirming that poetic or figurative texts quality translation is achieved by combining LLMs' translation with human post-editing (Muftah, 2022; Wang, 2023).

On the contrary, despite the limitations of LLMs' translation, especially large-scale ones, they have significantly improved translation systems and showed strong capabilities in generating coherent texts and promoting intercultural communication (Khoshafah, 2023). Banat and Abu Adla (2023) and Jiao et al. (2023) revealed that LLMs achieved high accuracy in European languages and performed well in Arabic-English translation, stressing the potential of human translators' post-editing to account for cultural subtleties. Thabet and Qadha (2024) evaluated LLMs' assisted translation of Al-Qusaibi's poetry and reported that LLMs' users achieved higher scores than non-users, ensuring the crucial benefits of LLMs as supportive tools. Similarly, Işım and Balcioglu (2023) tested ChatGPT across multiple benchmark data sets with 50 Turkish paragraphs and found it to be effective. Jiao et al. (2023) examined ChatGPT's translation performance of various prompts and showed that its output is stronger for high-input languages and less effective for low-input ones. Wang (2024) compared ChatGPT-4o's assessment of translation with that of humans for both literary and non-literary texts and revealed ChatGPT-4o's strong alignment in terms of fidelity, accuracy, and fluency, highlighting the necessity to improve its translation quality.

In conclusion, although substantial research has explored LLMs' translation of poetry (e.g., Alowedi & Al-Ahdal, 2023; Amini et al., 2024; Lieber et al., 2024; Obeidat & Jaradat, 2024), there remains a critical gap regarding how LLMs manage poetic devices. The failure of LLMs to process cultural subtleties, historical references, purpose of the poem, and figurative language confirms that LLMs are not a reliable translators, and human intervention is critical for maintaining faithfulness and elegance in literary works translation. Moreover, though some studies have addressed translation challenges between English and other languages (Karaban & Karaban, 2024; Obeidat & Jaradat, 2024; Thabet & Qadha, 2024), Arabic poetry translation into foreign languages is still, to some extent, unexplored. Furthermore, though many existing studies concentrated on LLMs' limitations, they did not suggest any strategies for improving their translation. Accordingly, this study seeks to fill such gaps and expand the scope by addressing the complexities of poetic language translation through LLMs and strive to develop translation strategies that more closely reflect the creativity and depth beyond the current capabilities of LLMs.

3. Study Methodology

3.1. Study Design

This study adopted a mixed methods design (qualitative and quantitative) with the help of the comparative analytical analysis to evaluate the effectiveness of LLMs in translating metaphors in Al-Ashmawi's poetry. Twenty verses from the poem of "Fortitude in the Time of Fracture" by Al-Ashmawi were selected, validated, and translated by LLMs, and their translations were compared with that of human translation. The evaluation of the translations was carried out by four professional experts in addition to MQM model, based on three parameters: linguistic accuracy, poetic fidelity, and emotional impact (Malmkjær, 2020). The MQM model is a system used for measuring the effectiveness of translation on many aspects, such as fluency, accuracy, and stylistic, facilitating a deep analysis of various translation methods, and reflecting human evaluation of translation quality (Lommel et al., 2014).

3.2. Study Procedures

This study took place in four stages. The first stage was the selection of the poetic verses from the poetry of Al-Ashmawi and the validation of them for study. The second stage was the translation of these verses where they were translated by two human translators as well as by LLMs. The third stage was the evaluation of the translation quality by four experts specialized in literature and translation, using Likert five-point scale where 5=Excellent (5-4.20), 4=very good (4.19-3.40), 3=good (3.39-2.60), 2=weak (2.59-1.80), and 1=very weak (1.79-1) based on three parameters: a) linguistic accuracy (meaning preservation), b) poetic fidelity (imagery, rhythm, tone), and c) emotional impact (preservation of figurative meaning). In addition, MQM was used to evaluate translation errors of both systems in terms of literalism, imagery ignorance, simplification, and exaggeration. The fourth stage was the comparative analysis of the translation results, where the researcher quantitatively and qualitatively compared, analyzed, and discussed the results. The results regarding translation quality were analyzed descriptively using SPSS, whereas the data regarding the translation errors were analyzed by MQM. Then many samples from the translations were also provided to illustrate where each system excelled or fell short, with a brief comment on how each system handles the translation of metaphorical Arabic poetry.

4. Data Analysis and Results Discussion

4.1. Results Related to the First Question of the Study

The first question of this study dealt with evaluating the effectiveness of LLMs' translation of metaphor in Alashmawi's poetry in comparison to human translation. To answer the first question of the study, the descriptive analysis, represented by means, standard deviations, and ranks, was employed for analyzing both systems translation quality using SPSS. The findings were analyzed quantitatively, then discussed qualitatively, at both the overall evaluation level and for each evaluation parameter level, of both systems translation separately.

Table 1: Descriptive Statistics Evaluation of LLMs' and Human Translation

System	Mean	Std.	Rank	Overall Evaluation
Human Translator 1	4.83	0.09	1st	Fidele and creative translation
Human Translator 2	4.60	0.10	2nd	Balanced and vivid translation
ChatGPT Translation	4.30	0.17	3rd	Accurate but metaphor loss translation
DeepSeek Translation	4.03	0.21	4th	Expressive, with exaggerated imagery

Starting with the results analysis at the overall evaluation level of the two systems translations, the results shown on Table 1 above demonstrated that human translators outperformed LLMs quantitatively and qualitatively in translating metaphors with noticeable differences between their translations. They obtained the top ranks with mean scores ranging (4.83-4.60) and Std. (0.9-0.10), indicating their vividness, use of balanced interpretation, excellent fidelity, and creative preservation of metaphorical layers, cultural resonance, and poetic imagery. In comparison, though LLMs demonstrated accuracy and expressiveness, they showed a relatively lower performance with lower ranks than human translators with mean scores ranging (4.30-4.03) and Std. (0.17-0.21), indicating their loss of metaphor imagery, inconsistency, and exaggeration of metaphor translation.

Table 2: Descriptive Statistics Evaluation of Each Parameter of Both Systems' Translation

Parameters	Human Translator 1	Human Translator 2	DeepSeek	ChatGPT
Linguistic Accuracy	4.90	4.70	4.30	4.50
Poetic fidelity	4.70	4.50	4.00	4.30
Emotional Impact	4.90	4.60	3.80	4.10
Overall Mean	4.83	4.60	4.03	4.30
Std. (Overall)	0.09	0.10	0.21	0.17

Regarding the results analysis of both systems translation at the level of each evaluation parameter, the results shown in Table 2 above revealed that Human Translator 1 got the top ranks in linguistic accuracy and poetic fidelity, with mean scores of 4.90 & 4.70 respectively, whereas DeepSeek got the last rank in the same parameters with mean scores of 4.30 & 4.00 respectively. Furthermore, with respect to emotional impact, Human Translator 1 also got the top rank with a mean score of 4.90 whereas DeepSeek got the last rank with a mean score of 3.80. Human Translator 2 and ChatGPT came in between.

For further analysis, the findings at the level of each parameter evaluation shown in Table 2 above are discussed qualitatively as follows:

1. Linguistic Accuracy

Linguistic accuracy deals with the correct use of language in the target translated texts in terms of appropriate selection of vocabulary, punctuations, and grammatical, lexical, and syntactical rules of the target language. Regarding the effectiveness of both systems in handling linguistic accuracy during translation, it has been found that both systems showed a strong and correct use of language. There is a preservation of metaphor layers and thematic interplay across verses and between contradictory words. Human translators reproduced linguistic structures smoothly, cohesively and emotionally, indicating their control of stylistic elevation, intuitive grasp of poetic language, and cultural aesthetic of the verses. As shown in Table 3 below, the choices of accurate words such as "glowed with peace; hearts entwined; cloaked in contentment" elevated the language adequately to keep it poetic without drifting into archaic or weird language.

LLMs, on the other hand, sometimes translated each poetic line independently and mishandled the translation by either producing overly literal translation or using inflated expressions that do not match the tone of the poem, such as the phrase "lived in quiet peace", "تحيا الكفاف" where the phrase "lived in peace" for example, does not mean "تحيا الكفاف". Though LLMs performed well in lexical equivalence and surface fluency, they lack interpretive intuition, especially when a metaphor is symbolic, cultural, or emotive. In addition, ChatGPT, in particular provided fluent and clear translation with a moderate figurativeness and delivered a readable poetic rhythm as in the words of "peace & cease", showing an accurate translation but more conservative, leaning toward clarity over stylistic depth, whereas DeepSeek showed a tendency toward visually rich output, indicating its capacity to recognize metaphorical expressions in spite of its over elaboration and loss of cohesion while translation. It often loses metaphoric imagery as shown in phrases like "sheltered family living in sufficiency", أسرة "مستورة تحيا الكفاف" where it adds a verbose not found in the original tone of the poem.

Table 3: An Example of Both Systems' Treatment of Linguistic Accuracy While Translation

Verse 5 : "كانت هناك أسرة مستورة.....تحيا الكفاف وبالتألف تنعم"		
System	Translation	Comment
Human 1	A humble hearth once glowed with peace; hearts entwined, living in harmony's light.	Rich textiled metaphor, deep cohesion
Human 2	There lived a humble family, cloaked in contentment, weaving harmony from threads of modest grace.	Soft retained imagery
DeepSeek	Once, there was a sheltered family, living in sufficiency, blessed with harmony.	Accurate meaning but verbose diction, loss of metaphoric equivalents
ChatGPT	A modest family lived in quiet peace, content with little, their love would never cease.	Clear and plain meaning but loss of mythic tone

This finding is in line with (Abdelaal & Alazzawie, 2020; Kadaoui et al., 2023; Khalifa, 2015; Khoshafah, 2023; Omar & Gomaa, 2020; Putri & Setiajid, 2021; Tisgam, 2014) which showed that literary works translation requires artistic creativity and interpretive ability, qualities that LLMs still consider difficult to replicate accurately in poetic translation, confirming that LLMs tend to focus on literal word replacement and neglect cultural context and poetic imagery.

2. Poetic Fidelity

Poetic fidelity deals with preserving the poetic qualities of the original poem while translation, such as tone, style, imagery, rhythm, musicality, artistic effect, etc. With respect to the effectiveness of each system in handling the poetic fidelity while translation, it has been found that human translators demonstrated an exceptional sensitivity to poetic expressions and rhetorical nature of the original poem and preserved personifications, symbolic imageries, and poetic tension better than LLMs. As shown Table 4 below, a sentence like "سحبُ تلوح ورعدها" "يتكلم" was translated by LLMs as "Clouds loom, and thunder tongues declaring truth" or "Clouds appear, their thunder speaks aloud" whereas human translators translated it as "Clouds unveil their tempests, thunder tongues proclaim their might", indicating that human translators captured both the meaning and metaphorical power which is implicit in the Arabic imagery, reproduced figurative meaning effectively in the target texts, and maintained metaphor clusters. They avoided flattering effects common in LLMs' translation and preserved metaphor layers without over clarifying or allowing the English reader to experience the same contextual, aesthetic, and cultural depth of the original text. They succeeded in combining faithfulness to meaning and poetic creativity with consistent rhythm and tone across all verses. Both of them captured the poem voice and the divine conversation between "earth, faith, and dawn" without sacrificing English fluency or aesthetic grace, taking into account that poetry metaphors operate in networks as chains of interlinked images and cultural allusions that unfold across poetic lines.

Table 4: An Example of Both Systems' Treatment of Poetic Fidelity While Translation

Verse 1: "سحب تلوح ورعدها يتكلم...والأرض تسمع ما يقال وتفهم"		
System	Translation	Comment

Human 1	Clouds unveil their tempests, thunder tongues proclaim their might, and Earth, in solemn silence, hears and comprehends the vow of skies.	Use of metaphorical expressions; personification, vivid imagery, and extended metaphors
Human 2	Clouds loom, their thunder tongues declaring truth, while the earth listens, understanding the voice of heaven.	Faithful and poetic emotions, emotional tone and cultural flavor
DeepSeek	Clouds loom, and their thunder speaks, and the earth hears what is said and understands.	Lexically and semantically good but rhythmically broken, and inflates metaphors and loses precision
ChatGPT	Clouds appear, their thunder speaks aloud, and Earth listens well, grasping what is vowed.	Smooth syntax and cohesion, simplified metaphors, literal tendencies

In contrast, though LLMs, especially ChatGPT kept rhythmic parallelism and internal rhyme found in the original poem as in words like "aloud & vowed", they make the translation prosaic and leading to stylistic drift and occasional semantic inflation. They translated the sentence "سحبُ تلوح ورعدُها يتكلم" literally as "Clouds loom/appear, and their thunder speaks aloud". Their translation is lexically faithful and grammatically good but rhythmically broken and semantically flattering. These results are consistent with (Alotaibi & Hussein, 2017; Alowedi & Al-Ahdal, 2023; Ghareb & Elimam, 2018) that showed that the poetic devices and figurative language remain particularly difficult for LLMs to capture, often resulting in the loss of metaphorical imagery and symbolism and confirming that LLMs frequently still fail to maintain the figurative and stylistic expressions of Arabic poetry.

3. Emotional Impact

Emotional impact deals with the ability to convey the effect and feelings of the original poem to the target audience or readers. Without this emotional equivalents, the translation becomes as a shadow of the source text. With regard to each system effectiveness in handling the emotional impact while translation, it has been found that human translators demonstrated the closest poetic effect of the original verses. As shown in Table 5 below, it has been found that human translation was highly figurative and rhythmically faithful and sensitive to metaphorical and stylistic layers. Human translators consistently preserved the personification and metaphorical density of Arabic poetry. For example, the sentence "كان المساء حكاية ليلية" was translated by human translators as "the evening unfolded like a silken tale" and the phrase "وتنصت أنجم" was translated as "stars bent low to listen". They preserved both semantic accuracy and figurative power of the original source. This indicates their competency in interpreting layered meanings and figurative resonance within metaphorical language, maintained Arabic metaphorical associations and cohesive imagery.

Table 5: An Example of Both Systems' Treatment of Emotional Impact While Translation

Verse 6: "كان المساء حكاية ليلية... يهدي بها قمر وتنصت أنجم"		
System	Translation	Comment
Human 1	The evening was a tale, the moon would whisper, and the stars, enraptured, leaned to listen.	Personification perfectly balanced, metaphorical images

Human 2	Evening unfolded like a silken tale; the moon recited, and the stars bent low to listen.	Elegant simile, tactile texture "silken" adds beauty to translation
DeepSeek	The evening was a nightly tale, narrated by a moon, with listening stars.	Acceptable but stilted
ChatGPT	The evening was a tale in starry light, the moon would tell, while stars listened at night.	Fluent linguistically but rhythmically loose; loses elegance

On the other hand, LLMs though keeping rhythm, they weakened the poetic elegance and vary the tone across poetic lines and often render implicit metaphors literally as shown in the sentence "يهدى بها قمر وتتصت انجم" was translated as "The moon tells, and the stars listen", leading to the loss of mythic tone and artistic beauty. In other verses, they sometimes omit rhythmic parallelism and internal rhyme found in Arabic structure, and inflated it with verbose diction, and drifted imagery semantically. The sentence "وفم الربيع الطلق يحكي قصة" was translated as "The mouth of lush spring tells a story", and it was inflated with the word "lush" which added imagery not found in source text, making inconsistent style and lyrical, rigid translation. In the same token, while ChatGPT frequently simplified metaphors and reduced the emotional and symbolic depth of the original text, DeepSeek produced expressive and visually rich language with outputs suffered from inconsistency, exaggeration, and over-embellishment, often drifting away from the intended metaphorical meaning. These findings are consistent with (Ahmed et al, 2025; Al-Hasnawi, 2007; Farghal & Haider, 2024; Mahmoud, 2018; Van der Meer, 2019; Venuti, 2017) which showed that though LLMs can manage structural and lexical aspects well and make progress in fluency and general accuracy, they still lack the interpretive and aesthetic intuition required for literary translation and often struggle to reproduce emotional effect, confirming that capturing poetic language accurately requires a deep understanding of emotional subtleties, cultural nuances, historical context, and intricate structures of poetic language.

To conclude, Arabic poetry translation is a demanding task as it carries cultural associations, political and religious concepts and symbolic natural elements. Human translators have the potential to preserve these subtleties and maintain the contextual and cultural references of the metaphor whereas LLMs often neutralize or generalize such culturally loaded images and weaken metaphor fidelity, style authenticity, and cultural specificity due to the semantic layering of the figurative language.

4.2. Results Related to Study Question Two

The second question of this study dealt with pointing out the main common errors of LLMs' translation in comparison to that of human translators in terms of literalism, imagery ignorance, simplification, and exaggeration. To achieve adequate data, the findings were evaluated quantitatively by MQM with the use of percentages then discussed qualitatively at the level of the overall evaluation errors of translation of both systems and at the level of each evaluation aspect of translation errors of each system, where the percentages of errors were calculated based on the total errors observed in the translation of each system after classifying them according to MQM model criteria mentioned above.

Table 6: Descriptive Statistics of MQM Evaluation of Both Systems' Errors While Translation

Error Aspect (%)	Human Translator 1	Human Translator 2	DeepSeek	ChatGPT
Literalism	3%	5%	18%	13%
Imagery Ignorance	4%	6%	22%	15%
Simplification	4%	3%	12%	9%
Exaggeration	3%	4%	8%	6%

Starting with the total evaluation of both systems common errors and based on the results shown in Table 6 above, it has been found that Human translator 1 made the least errors in the four aspects, and DeepSeek made the most errors in the similar aspects. Human translator 2 and ChatGPT came in between. To make it clear, the results are clarified and discussed qualitatively in detail at the level of each error aspect as follows:

1. Literalism Errors

Starting with the main errors in literalism, it has been found that literalism was the most significant weakness of LLMs with a percentage of 13% by ChatGPT and 18% by DeepSeek versus human translators' errors of literalism with a percentage of 3% by Human Translator 1 and 5% by Human Translator 2. LLMs used to prioritize literal accuracy over the preservation of the aesthetic and emotional resonance that defines true literature. They struggled to fully capture metaphor layers, often opting for surface level meaning. To make it clear, the poetic line in Table 7 below "يأتي إليها الفجر طفلاً أشقراً... ولسانه بالذكريات يتمم" was translated by Human translators as "Dawn came to her as a golden haired child" and "Dawn tiptoes in as a golden child", and it was translated by LLMs as "Dawn would arrive as a fair-haired child" and "Dawn would arrive there like a fair-haired child." Such translation indicates that LLMs tend to translate metaphor literally, resulting in the loss of metaphorical tone, flattening of imagery, and reduction of aesthetic quality. Human translators, in contrast, demonstrated clear interpretative ability and preservation of metaphorical density in appropriate ways. This finding comes in line with (Al-Batineh, & Rabadi, 2019; Kadaoui et al., 2023; Khoshafah, 2023; Putri & Setiajidi, 2021) which showed that LLMs often failed to maintain the figurative and stylistic meaning of Arabic poetry and rely on literal word replacement, overlook imagery, ignore the essential elements for conveying poetic language, and lack the critical evaluative skills required to identify the covert errors that occur in poetic and cultural texts (Zhang, 2025).

Table 7: An Example of Literalism Errors by Both Systems While Translation

Verse 9: "يأتي إليها الفجر طفلاً أشقراً... ولسانه بالذكريات يتمم"		
System	Translation	Comment
Human 1	Dawn tiptoes in, a golden child, murmuring old memories in its tender tongue.	Beautiful metaphor of dawn as innocence and remembrance
Human 2	Dawn came to her as a golden-haired child, murmuring memories in a tongue only sorrow knows.	Balanced tone and fluency
DeepSeek	Dawn would arrive there like a fair-haired child, Its tongue muttering with memories.	Literal translation, simpler diction, redundant phrasing

ChatGPT	Dawn would arrive as a fair-haired child, whispering memories soft and mild.	Poetic rhythm improved; still generic
---------	--	---------------------------------------

2. Imagery Ignorance Errors

With regard to imagery ignorance errors, it has been found that LLMs committed high imagery ignorance errors with a percentage of 22% by DeepSeek and 15% by ChatGPT, versus human translators' errors with a percentage of 4% by Human Translator 1 and 6% by Human Translator 2. For example, as shown in Table 7 below, the poetic line "كانت ربوع القدس أرضاً حرة... ترعى كرامتنا بها وتعظم" was translated by LLMs as "Domains of Jerusalem was a free land, nurtured our dignity and honored it", and it was translated by human translators as "The fields of Jerusalem once breathed freedom, guarded our honor and crowned our pride." This indicates LLMs' plain and amplified translation that overshadows poetry and semantic drift where LLMs tend to mismanage and ignore poetic imagery, over-exaggerate it beyond the original source, and alter the tone and symbolic intent. This could be attributed to their over-generative tendencies and lack of contextual interpretive ability. Human translators, in contrast, maintained closer control over imagery and preserved poetic fidelity, expressive imagery, elevated tone, and faithful translation. This finding is congruent with (Al-Khalafat & Haider, 2022; Ghareb & Elimam, 2018; Hassan, 2020; Malmkjær, 2020) which showed that machine systems often rely on literal word replacement and ignore imagery.

Table 8: An Example of Imagery Ignorance Errors by Both Systems While Translation

Verse 8: "كانت ربوع القدس أرضاً حرة... ترعى كرامتنا بها وتعظم"		
System	Translation	Comment
Human 1	Jerusalem's meadows once robed in freedom's green, guarded our honor and crowned our pride.	Powerful moral metaphor; elevates tone.
Human 2	The fields of Jerusalem once breathed freedom, cradling our honor, exalting our dignity like sacred fire.	Faithful, visual strong fidelity and creativity
DeepSeek	The domains of Jerusalem were once a free land, that nurtured our dignity and honored it.	Plain, political overtones overshadow poetry, loss of imager, amplified or misdirected images
ChatGPT	Jerusalem's meadows were once land of the free, guarding our honor, magnifying dignity.	Faithful but ordinary diction and imagery

3. Simplification Errors

With respect to simplification errors, it has been found that both LLMs simplify imagery due to their inability to sustain complex metaphor chains with a percentage of 12% by DeepSeek and 9% by ChatGPT, versus human translators' errors with a percentage of 4% by Human Translator 1 and 3% by Human Translator 2. For example, the poetic line "كانت هنالك أسرة مستورة... تحيا الكفاف وبالتآلف تتعم" was translated by LLMs as "There was a modest family lived in quiet peace, with little, their love would never cease and "Once, there was a sheltered family living in sufficiency, blessed with harmony", and translated by human translators as "There was a humble hearth once glowed with peace, entwined, living in harmony's light." and "There lived a humble family, cloaked in contentment". This reflects LLMs' simplified translation and

struggle with preserving metaphor networks which require integrating context, symbolism, and emotional tone. These results are congruent with (Abdelaal & Alazzawie, 2020; Fakhri et al., 2024; Wang, 2024) that showed that simplification, as a lexical error, inappropriate lexical choice, and a semantic error, is the most common error due to the ambiguity inherent in metaphors.

4. Exaggeration Errors

With respect to exaggeration errors, it has been found that LLMs tend to expand the metaphor beyond the original tone by adding explanatory phrases, reflecting its generative style and addition of descriptive elements not found in the source poem with a percentage of 6% by ChatGPT and 8% by DeepSeek versus human errors with a percentage of 3% by Human Translator 1 and 4% by Human Translator 2. As shown in Table 9 below, the poetic line كانت هنالك زهرة فواحة... وصغيرة ترعى، وطفلٌ يحلم " was translated by LLMs as "There once was a verdant grove in bloom, where nightingales sang beneath its gloom" and as "There was once a garden dressed in emerald dreams, where nightingales wove songs in the shawl of shade" by human translators. This indicates human translators' mastery of poetic language and their ability to recreate exaggerated images and stylistic expressions in English to ensure that the poem still has a provocative and engaging experience, such as the use of "nightingales" for the "girl" and the "child". On the contrary, LLMs struggled significantly with parallel structures and produced prosaic translation that shows exaggeration and over-expansion. They broke the main role of poetry which is to show rather than to tell. Such weakness negatively impacts fluency, poetic faithfulness, and emotional resonance of metaphor translation. These findings confirm that linguistic and cultural complexity, metaphorical ambiguity, and cultural symbolism make Arabic poetry difficult to translate. These results are in line with (Torralba & Way, 2018; Venuti, 2017) which revealed that the core difficulty of metaphor translation lies in its high abstractness, generality, metaphoricity, and ambiguity as well as the excessive dependence of Arabic poetry on figurative language, rhythm, contextual nuances, and cultural references, areas in which LLMs often fall short.

Table 9: An Example of Both Systems' Exaggeration Errors While Translation

Verse 4: "كانت هنالك زهرة فواحة... وصغيرة ترعى، وطفلٌ يحلم"		
System	Translation	Comment
Human 1	There was once a garden dressed in emerald dreams, where nightingales wove songs in the shawl of shade.	Strong cinematic metaphor; "tended life" symbolic of hope
Human 2	Once bloomed a garden, lush with song, where nightingales wept melodies beneath its shade.	Abstract beauty but omits "girl grazing
DeepSeek	Once, there was a verdant meadow, and nightingales sang in its shade.	Accurate but prosaic
ChatGPT	There once was a verdant grove in bloom, where nightingales sang beneath its gloom.	Syntactically smooth, emotionally thin

5. Conclusion

This study aimed at evaluating the LLMs' translation of metaphor in Al-Ashmawi's poetry from Arabic into English in comparison to human translation. Using MQM model besides professional experts for evaluating the parameters of linguistic accuracy, poetic

fidelity, and emotional impact while translation, the results showed significant differences in both systems treatments of metaphor translation, where human translators outperformed LLMs both quantitatively and qualitatively. Human translators recognized the places of metaphors and knew how to replace it by a figurative equivalent in the target language. They effectively preserved the figurative depth, emotional resonance, and cultural imagery of the Arabic poetic texts. Their translation of metaphors preserved the symbolic value, rhythm, and cohesion across the verses, indicating their proficiency and sensitivity. On the contrary, LLMs though provided a fluent and linguistically accurate translation, they showed clear weaknesses in metaphor transfer and stylistic consistency. While ChatGPT maintained acceptable rhythm, cohesion, and readability, it often tended towards literalism and metaphor simplification, leading to the flattening of emotional and cultural nuances, DeepSeek, on the other hand, displayed greater imaginative freedom, over-embellished and ignored imagery, and broke metaphorical coherence and tone. All in all, the findings confirmed that LLMs' translation can approximate surface meaning but struggle with layered metaphorical and cultural aspects, especially those that require interpretive creativity and contextual sensitivity.

6. Pedagogical Implications

Based on the findings, the current study carried many implications for human translators, LLMs users and developers, and future researchers as follows:

1. Human translators should deal with literary translation as an interpretive act not as a lexical substitution. They should select contextual, cultural, theological, emotional, and moral equivalents that meet the source text needs. They should prioritize the development of cultural and aesthetic intuition over technical proficiency
2. LLMs' users should post-edit translation and restore figurative and rhythmic integrity.
3. LLMs' developers should enhance literary forms recognition and generation modules by training them on literary and poetic corpora, integrate cross-cultural and semantic mapping to preserve figurative meaning embedded in cultural world, and combine neural translation with computational linguistic rules to handle figurative language more accurately, integrate LLMs with human expertise to get a hybrid approach for translation.
4. Future researchers have to conduct further studies on other genres translation (e.g., political, legal, medical discourses) to measure LLMs' translation proficiency of such genres as these fields are of great importance, and any mistake in LLMs' translation may lead to disastrous consequences.

7. Study Limitations

The current study has some limitations in terms of the sample, evaluation, and instruments as follows:

1. The sample of study was twenty verses from one poem in one divan. It would be better if many verses were taken from many poems in many divans, thus limiting the generalizability of the findings.
2. The study relied on four experts for comparing and evaluating the translation of both systems. Using a larger number of experts from multiple universities with high language proficiency could affect the results.

3. The study used only two LLMs and only two human translators for metaphor translation comparison. It would be better if the comparison was between many types of LLMs and many human translators.

8. Adherence to Ethical Standards

Acknowledgments: The researcher expresses his sincere gratitude to the two human translators who translated the selected verses from Arabic into English. He also thanks the four experts who evaluated the translations of both systems. He also expresses his thanks with gratitude to the Arab Observatory for Translation for funding this study.

Ethic declarations: The informed consent was obtained from all participants included in this study. Participation was voluntary and participants were informed for the purpose of the study and their responses were kept confidential and only used for study purposes, and participants have the right to withdraw any time without any penalty.

Data Availability: The data supporting the findings of this study are available within the manuscript. Raw dataset generated during the current study are available from the corresponding author.

Funding: This research received grant no. (697/2026) from the Arab Observatory for Translation (an affiliate of ALECSO), which is supported by the Literature, Publishing & Translation Commission in Saudi Arabia.

Consent for Publication. The author commented on previous versions of the manuscript and has read and approved the final version of the manuscript.

Conflict of Interest: The researcher declares that there are no financial, personal, or professional interests that could influence the results or interpretation of this study.

Author Contributions: The author contributed to all stages of the study, including study design, data collection, analysis, interpretation of results, and preparation of the final manuscript. The author has approved the final version submitted for publication.

References

- Abdelaal, N. M., & Alazzawie, A. (2020). Machine translation: The case of Arabic-English translation of news texts. *Theory & Practice in Language Studies*, 10(4). doi:<http://dx.doi.org/10.17507/tpls.1004.09>
- Ahmed, R., Alfaifi, A. Al-Kadi, A. & Alasmari, M. (2025). Bridging language and artificial intelligence in translating Arabic poetry. *Journal of Psycholinguistic Research*, 7 (2).
- Akasheh, W. M., Haider, A. S., Al-Saideen, B., & Sahari, Y. (2024). Artificial intelligence-generated Arabic subtitles: Insights from Veed. Io's automatic speech recognition system of Jordanian Arabic. *Texto Livre*, 17, e46952-e46952. doi:10.1590/1983-3652.2024.46952.
- Al-Batineh, M., & Rabadi, R. (2019). Will the machine understand literary translation? A glimpse into the future of literary machine translation through the lenses of artificial intelligence *Studies in Translation*, 5(1), 151–169.
- Al-Buqmi, F. M. M. (2024). Pictorial metaphor in the poetry of Iyad Al-Hakimi: The collection “I Do Not Know Their Sorrow” as a model (a cognitive approach). *Journal of Research*, 3 (74).

- Al-Ghamdi, S., Al-Khalifa, H., & Al-Salman, A. (2023). Fine-tuning Bert-based pre-trained models for Arabic dependency parsing. *Applied Sciences*, 13(7), 4225.
- Al-Hasnawi, A. R. (2007). A cognitive approach to translating metaphors. *Translation Journal*, 11(3).
- Ali, A. S., & Ben Man, S. (2012). Errors in translating the concept of metaphor in Qur'anic verses into Malay: Presentation and analysis. *Journal of Islam in Asia*, 8.
- Al-Kassab, E. (2024). *AI translation vs. human translation: The pros & Cons*. https://www.researchgate.net/publication/386345258_AI_Translation_vs_Human_Translation_the_Pros_Cons
- Al-Khalafat, L., & Haider, A. S. (2022). A corpus-assisted translation study of strategies used in rendering culture-bound expressions in the speeches of King Abdullah II. *Theory and Practice in Language Studies*, 12(1), 130-142.
- Al-Khesali, S. (2008). *Metaphors and modern Arabic poetry(1st ed.)*. Morocco: Dar Toubkal Publishing.
- Al-Khesali, S. (2022). Translation and the problems of the untranslatable. *Revue Linguistique et Référentiels Interculturels*, 3(2).
- Alotaibi, M. S. (2021). Improving Arabic machine translation using transliterated data. *Journal of Computational Science*, 54, 101-213.
- Alotaibi, S., & Hussein, A. H. (2017). Challenges in Arabic poetry machine translation: A survey. *Ain Shams Engineering Journal*, 8(4), 311-318.
- Alowedi, N., & AlAhdal, 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* · DOI: 10.59670/jns.v33i.800
- Alrefaee, Y. M. H. A., & Al-Shammari, A. H. (2026). Determinants of AI translation technology adoption among Arab professional translators across national contexts. *Discover Education*. Advance online publication. <https://doi.org/10.1007/s44217-026-01944-6>
- Alrefaee, Y. M. H. A., & Al-Shammari, A. H. (2026). Digital innovations in second language pragmatics: A systematic review (2006–2024). *Social Sciences & Humanities Open*, 13, 103004. <https://doi.org/10.1016/j.ssaho.2026.103004>
- Al-Tarawneh, H., & Al-Badawi, M. (2025). Artificial intelligence and the transformation of human translation: A study of the translation industry in Jordan. *Cogent Arts & Humanities*, 12(1).
- Al-Zoghby, A. A., & Elshamy, A. A. (2021). Readability assessment for Arabic poetry translation. *Future Computing and Informatics Journal*, 6(1).
- Amarin, R., & AlNajjar, S. (2021). Unsupervised neural machine translation of Arabic poetry. *Computers and Electrical Engineering*, 93, 107258. <https://doi.org/10.1016/j.compeleceng.2021.107258>
- Amroosh, F. (2025). Machine translation of poetry from and into Arabic in the age of artificial intelligence: A critical analysis of ChatGPT outputs in light of aesthetic challenges. *Maalim Journal*, 2 (16).

- Banat, M., & Abu Adla, Y. (2023). Exploring the effectiveness of GPT-3 in translating specialized religious text from Arabic to English: A comparative study with human translation. *Journal of Translation and Language Studies*, 4(2), 1-23.
- Bergsma, S. (2021). Machine translation for poetry: Challenges and developments. *Translation Studies*, 14(3), 213-227.
- Elmadany, A., & Wayne, A. (2018). Machine translation of Arabic poetry: Developing a poetry translation corpus and domain adaptation. *Procedia Computer Science*, 142, 160-168. <https://doi.org/10.1016/j.procs.2018.10.365>
- Enad, A. (2018). Translating rhetorical figures according to Eugene Nida. *Journal of Translation and Languages*, 2 (17).
- Fairooz, R. A. (2024). Translating Al-Mutanabbi's wisdom poetry into English: A localisation product. *International Journal of English Literature and Culture*, 12 (2), 30-55.
- Faleh, B. M. (2026). Metaphorical substitution in selected verses from Al-Gharbawi's poetry: A linguistic study. *Wasit Journal for Humanities*, 2(12).
- 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), 2410998.
- Gao, F. (2024). When minds meet machines: A pilot programme of using Chat-GPT to teach political speech translation in a Chinese-to-English translation classroom. *International Journal of Chinese and English Translation & Interpreting*. Retrieved from <https://ijceti.at-journals.com/ojs/index.php/description/article/view/94>
- Garcia, I. (2021). AI and translation: Challenges and opportunities. *Journal of Specialised Translation*, 35, 2-16.
- Genzel, D., Uszkoreit, J., & Och, F. (2010). "Poetic" statistical machine translation: rhyme and meter. In Hang Li, & Lluís Màrquez (Eds.), *Proceedings of the 2010 Conference on empirical methods in natural language processing* (Massachusetts, USA, 9–11 October 2010) (pp. 158–166). Cambridge, MA: Association for Computational Linguistics.
- Ghareb, I., & Elimam, A. (2018). Evaluation of Google Translate in translating figurative Arabic poetry into English. *Advances in Language and Literary Studies*, 9(1), 142-148.
- Harmah, A. (2023). *Challenges of poetry translation*. Journal of Al-Bayda University, 3(5).
- Hassan, S. (2020). Artificial intelligence in literary translation: A review of the field. *Journal of Translation and Technology*, 7(2), 45-61.
- Ibrahim, T. I. (2024). *Artificial intelligence and its impact on poetry translation*. Adnan Publishing and Library.
- Inaam, B. (2003). *Literary translation: Problems and solution*, V1. Beirut, Lebanon: Dar Al-Farabi.
- Işım, Ç., & Balcioglu, Y. (2023). *ChatGPT: Performance of translate [Paper presentation]*. Paper presented at the 3rd international ChaRaKa Congress on humanities and Social Sciences, acharaca.
- Jiao, W., Wang, W., Huang, J., Wang, X., & Tu, Z. (2023). Is ChatGPT a good translator? Yes with GPT-4 as the engine. *arXiv preprint arXiv:2301.08745*.

- Kadaoui, K., Magdy, S. M., Waheed, A., Khondaker, M. T. I., El-shangiti, A. O., Nagoudi, E. M. B., & Abdul-Mageed, M. (2023). Tarjamat: Evaluation of Bard and ChatGPT on machine translation of ten Arabic varieties. *arXiv preprint arXiv:2308.03051*. <https://doi.org/10.48550/arXiv.2308.03051>
- Kahrizangi, A. A. S., & Haddadi, M. H. (2017). An inquiry into the challenges of literary translation to improve literary translation competence with reference to an anecdote by Heinrich Von Kleist. *Journal of Education and Learning*, 6(3), 350–357. <https://doi.org/10.5539/jel.v6n3p3>
- 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). doi: 10.59400/fls.v6i1.1994
- 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.
- Khoshafah, F. (2023). ChatGPT for Arabic-English translation: Evaluating the accuracy. *Research Square, Online First*, 1–20. <https://doi.org/10.21203/rs.3.rs-2814154/v2>
- Lehka-Paul, O., & Jankowiak, K. (2022). Novel metaphor translation is modulated by translation direction. *Applied Psycholinguistics*, 43(1), 177–192.
- Lieber, J., Mailly, J., Marquis, P., Prade, H., & Rollin, F. (2024). Some thoughts about artificial stupidity and artificial dumbness. *AI Communications*, 37(1), 377–391. DOI: 10.3233/AIC-220322
- Lommel, A. R., Uszkoreit, H. Burchardt, A. (2014). Multidimensional Quality Metrics (MQM): A framework for declaring and describing translation quality metrics. *Revista Tradumàtica: Traducció i Tecnologies de la Informació i la Comunicació*, 12, p.455–463.
- Lu, H., Liu, W., Zhang, B., Wang, B., Dong, K., Liu, B., Yang, H. (2024). DeepSeek-vl: Towards real-world vision language understanding. *arXiv preprint arXiv:2403.05525*.
- Luo, X. (2018). Artificial intelligence and the crisis of translation. *Asia Pacific Translation and Intercultural Studies*, 5(1), 1-2. <https://doi.org/10.1080/23306343.2018.1456440>
- Mahmoud, M. M. (2018). Challenges of translating Islamic religious texts: A case study of Qur'anic metaphors. *Arab World English Journal for Translation & Literary Studies*, 2(4), 94-106.
- Malika, B. (2016). From translating metaphors to the metaphor of translation. *Al-Mutarjim Journal*, 1 (16).
- Malmkjær, K. (2020). The challenges of translating poetry in the age of AI. *Poetry Translation Review*, 28, 34-45.
- Manetti, G., & Rossi, A. (2022). Emotional intelligence and creative nuance in AI-driven literary translation. *AI and the Arts*, 8(1), 89-104. <https://doi.org/10.xxxx/aia.2022.001>
- Mariam, B. A. (2023). *The aesthetic and cultural impact of translating literary works from Arabic into English: A case study of Ahlam Mosteghanemi's novel "Memory in the Flesh"*. Master's Thesis, Abu Bakr Belkaid University, Tlemcen, Algeria.

- Mirawi, A. (2014). The epistemological and aesthetic debate in constructing metaphor in contemporary poetic texts. *Journal of Sciences Journal* 1(17).
- Mohsen, M. A. (2024). Artificial intelligence in academic translation: A comparative study of Large Language Models and Google Translate. *Psycholinguistics*, 35 (2), 134–156.
- Muftah, M. (2022). Machine vs. human translation: A new reality or a threat to professional Arabic-English translators? *ResearchGate*.
- Newmark, P. (1988). *A Textbook of Translation*. Prentice Hall.
- Obeidat, M., & Jaradat, M. (2024). Artificial intelligence accuracy in translating resistance literature from Arabic into English: Google Translate and ChatGPT as a model. *Research Journal in Advanced Humanities*, 5(2). <https://doi.org/10.58256/9yz4dx79>
- Omar, A., & Gomaa, Y. (2020). The machine translation of literature: Implications for translation pedagogy. *International Journal of Emerging Technologies in Learning (iJET)*, 15(11), 228–235. <https://doi.org/10.3991/ijet.v15i11.13275>
- Oudah, M., & Shaalan, K. (2020). A hybrid approach for improving Arabic poetry machine translation. *ACM Transactions on Asian and Low- Resource Language Information Processing (TALLIP)*, 19(3), 1-22.
- Pan, R., Ibrahimzada, A. R., Krishna, R., Sankar, D., Wassi, L. P., Merler, M., & Jabbarvand, R. (2024). *Lost in translation: A study of bugs introduced by large language models while translating code [Paper presentation]*. Paper Presented at the Proceedings of the IEEE/ACM 46th international Conference on Software engineering, New York, United States. <https://doi.org/10.1145/3597503.3639226>
- Postigo, M. (2024). ChatGPT and MT-Systems: Advantages and limitations when translating English to Spanish and Portuguese. *Lengua Y Habla*, 28.
- Putri, A. T., & Setiajidi, H. H. (2021). Instagram Translate and human translation in the English captions of Jokowi's Account: An analysis of Koponen's error category. *Paper presented at the English Language and Literature International Conference (ELLiC) Proceedings*.
- Schaffner, Christina. (2004). Metaphor and translation: Some implications of a cognitive approach'. *Journal of Pragmatics*, 36, 1253-1269.
- Shormani, M. Q. (2024b). Artificial intelligence contribution to translation industry: Looking back and forward. *arXiv preprint arXiv:2411.19855*. 2024 Nov 29
- Soufi, S. (2018). *Translating metaphor in tourism texts from Arabic into French: An applied study*. Master's Thesis in Translation, University of Tlemcen.
- Tawfik, M., Samih, Y., Kholy, A. E., Elmadany, A., & Magdy, W. (2022). Towards a systematic framework for evaluating machine translation of poetry. *Computer Speech & Language*, 74, 101317. <https://doi.org/10.1016/j.csl.2022.101317>
- Thabet, M. A., & Qadha, A. M. (2024). The influence of AI-based translation tools on the translation of Dr. Ghazi Al-Qusaibi's poetry by Saudi EFL learners. *Journal of Educational Sciences and Humanities Studies*, 38(4), 772-799.
- Tisgam, K. H. (2014). Translating poetry: Possibility or impossibility. *Journal of College of Education for Women*, 25(2), 511-524.
- Toral, A., & Way, A. (2018). What level of quality can neural machine translation attain on literary texts? *Translation Spaces*, 7(1), 25-50.

- Tymoczko, M. (2007). *Translation and ideology: Encounters and clashes*. St. Jerome Publishing.
- Van der Meer, J. (2019). Can AI solve poetry translation?. *Nature*, 575(7781), 12-14.
- Venuti, L. (2017). *The translator's invisibility: A history of translation (2nd ed.)*. Routledge.
- Wang, J. (2024). Exploring the potential of ChatGPT-4o in translation quality assessment. *Journal of Theory and Practice in Humanities and Social Sciences*, 1(3), 18–30. <https://woodyinternational.com/index.php/jtphss/article/view/27>
- Zhong, Q., Ding, L., Liu, J., Du, B., & Tao, D. (2023). Can ChatGPT understand too? A comparative study on ChatGPT and fine-tuned BERT. *arXiv preprint arXiv:2302.10198*.