

# Translating classical Arabic verse: human translation vs. AI large language models (Gemini and ChatGPT)

Mohammed Farghal & Ahmad S. Haider

To cite this article: Mohammed Farghal & Ahmad S. Haider (2024) Translating classical Arabic verse: human translation vs. AI large language models (Gemini and ChatGPT), Cogent Social Sciences, 10:1, 2410998, DOI: [10.1080/23311886.2024.2410998](https://doi.org/10.1080/23311886.2024.2410998)

To link to this article: <https://doi.org/10.1080/23311886.2024.2410998>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 10 Oct 2024.



Submit your article to this journal [↗](#)



Article views: 2175



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)

# Translating classical Arabic verse: human translation vs. AI large language models (Gemini and ChatGPT)

Mohammed Farghal<sup>a</sup> and Ahmad S. Haider<sup>a,b</sup>

<sup>a</sup>Department of English Language and Translation, Applied Science Private University, Amman, Jordan; <sup>b</sup>MEU Research Unit, Middle East University, Amman, Jordan

## ABSTRACT

Classical Arabic poetry is a sophisticated and complex art form, combining intricate themes with refined stylistic forms. This paper examines the translation of 15 individual Classical Arabic verses by comparing the English renditions provided by a human translator and two prominent large language models (LLMs), namely Google's Gemini (GEM) and OpenAI's ChatGPT (GPT). The human and LLMs are evaluated in relation to three variables: thematic clarity, creativity, and prosody by 54 Arab professors in Literature, Linguistics, or Translation who use English as a medium of instruction at their universities and are considered qualified as assessors. The translations are also evaluated by the two LLMs. The findings show that participants' assessment of human and GPT translations are clearly positive on the three variables, while GEM significantly lags behind in terms of prosody (M 2.83) and even more so in its own assessment (M 1). As for AI technology assessment, GPT is clearly consistent with participants' evaluation, in contrast to GEM assessment, which is markedly different in terms of prosody and creativity. The study concludes that AI technology emerges as a true competitor with human translators, and their interactive potentials need to be further explored in poetry translation.

## ARTICLE HISTORY

Received 25 May 2024  
Revised 18 August 2024  
Accepted 15 September 2024

## KEYWORDS

Arabic classical poetry;  
translation; cognition;  
prosody; ChatGPT;  
Gemini; English

## SUBJECTS

Language & Linguistics;  
Literature; Cultural  
Studies

## Introduction

Translation is a substantial instrument in intercultural communication (Haider & Shohaibar, 2024; Saideen et al., 2024; Weld-Ali et al., 2023). Translating classical Arabic poetry into English is a powerful tool for the global dissemination of a core part of the rich heritage of Arab culture. This linguistic and cultural transfer is a challenging endeavor that has been introduced to expanded potential and complexity with the entering of machines into the translation field.

Classical Arabic poetry has a rich history that represents the profound culture and traditions of Arabs. To this day, Arabs use poetry to document and unpack their history (Igbaria, 2020). Its artistic components are a challenge to transfer even to skilled human translators. The competency of machine translation still lags behind that of human translators (Mirzoyeva, 2023). The quality of machine translation is ever-evolving, however (Öner Bulut & Alimen, 2023). Large language models (LLMs), such as Google's Gemini (GEM) and OpenAI's ChatGPT (GPT), show potential in translating poetry.

Translation is a complex act that is associated with skill, knowledge, and experience (Akan et al., 2019). Traditionally, machine translation has not been an automatic process, yet it is gaining in popularity. This illustrates the transition from early machine translation systems—which were rather inflexible and mainly reliant on manually coded linguistic rules—to more recent systems, which make use of sophisticated algorithms and machine learning. Modern machine translation systems, such as neural networks, learn from enormous volumes of data and modify their translations over time, in contrast to earlier systems that needed explicit rule-setting. Even though machine translation is still essentially algorithmic, it is becoming more sophisticated and can now provide translations that are more fluid

**CONTACT** Ahmad S. Haider ✉ [ah.haider86@gmail.com](mailto:ah.haider86@gmail.com) MEU Research Unit, Middle East University, Amman, Jordan

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

and accurate in context, setting it apart from human translation, which is more intuitive and context-sensitive. Determining the capacity of machine translation to retain stylistic, linguistic, and thematic features establishes its capabilities and limitations in translating classical Arabic poetry. This may provide insight into the current place of machine translation programs in the literary translation landscape.

The programs still face challenges in translating classical Arabic into English. Examples of such challenges include sophisticated language, imagery and symbolism, and rhythm and flow. Translating these elements requires technical skill and cultural and linguistic knowledge. This study, therefore, investigates the translation of prominent LLMs, namely GEM, and GPT, examining their attempts at reflecting the essence and style of classical Arabic poetry as rhymed couplets.

This study attempts to answer the following two research questions:

- How do AI technology (GEM and GPT) compare with the human translator in terms of clarity, creativity (use of metaphors and figurative language), and prosody (musicality)?
- How did 54 experts, GEM and GPT, rate AI technology (specifically GEM and GPT) compare to human translators in terms of clarity, creativity, and prosody, using a scale of 1 to 5?

## Review of related literature

### *Theoretical background*

Classical Arabic poetry is renowned for and characterized by the complexity of its style, linguistics, and content. Arabic poetry encompasses a wide range of topics like love, society, politics, and Islam (Ahmed et al., 2019). According to Orabi et al. (2020), classical Arabic poetry may also be classified based on the Era in which it was written. The genre employs poetic devices such as imagery, metaphor, and figurative language. The style is structured rhythmically and adorned with rhymed couplets. Such features have traditionally existed at the center of Arabic literary criticism (Harb, 2020). These features must be captured in the translation if the essence of the poem is to be truly represented.

Additionally, Robinson (2017) found that a quality literary translation results in a unique piece of literature that is not beneath the original works. Machine translation is quickly evolving, with mistakes becoming more challenging to pinpoint (Läubli et al. (2018)). The capability of AI translation tools such as GPT and GEM to accomplish this task and represent the literal and aesthetic dimensions of a poem has yet to be determined. This research thus takes a critical and comparative approach to gauging the competency of machine translations with human translations. The analysis is based on three criteria: thematic clarity, creativity, and prosody.

The first area of criteria is thematic clarity. While literary translation, and poetic translation especially, prioritize style more than other genres of translation, the transfer of sense is still the essence of translation. As stated by Jones (2019), literary translation may be concerned with style or content based on the priorities of the translators. Thus, it must be established whether machine translation can decipher and maintain poetic intent and meaning with consideration to cultural and historical nuances as well as complex wording and intricate wordplay.

Since the source text is literary, creative language is embedded in its core. The translation must, therefore, reflect the range of cultural and linguistic metaphors, symbolism, allusion, and figurative language found in the original poem. Khudoyorovich et al. (2020) described translation as a creative act as it requires the recreation of themes and senses. This prospect is magnified in the literary genre.

Moreover, what distinguishes poetry from other forms of literature, like prose, is its prosody. Classical Arabic poetry is distinguished by its musicality with its ‘arud’, structured rhythms and meters that produce an artistic, musical aesthetic for the verses (Al-Shaibani et al., 2020). Al-Talabani (2020) argued that recognizing the meters ‘arud’ that devises Arabic poetic prosody (APP) is a more arduous task than the recognition of the poem’s sense. This flow, along with the poem’s rhyming couplets, shapes Arab poetry into the style it is celebrated for. Hence, it is essential that the aesthetic features must be retained in order for the translation to be a true reflection of its source.

Compared to conventional techniques, LLMs like GPT and GEM provide more precise, fluid, and contextually nuanced translations (Al-Salman & Haider, 2024). Large-scale multilingual datasets pre-train LLMs, teaching them to recognize patterns, parse syntax, and grasp contextual links between different languages. Completing this basic pre-training phase is essential because it gives these models a comprehensive comprehension of language, allowing them to produce human-like content. LLMs are refined using individual translation datasets that concentrate on certain language pairs or specialized terminology after the pre-training stage to ensure the models can manage the complexities of translating between languages (Pan et al., 2024). By refining the model, one can increase its accuracy and cultural sensitivity in translations by improving its ability to adjust to each language's distinct linguistic characteristics and subtleties.

The attention mechanism is one important feature of LLMs that improves translation skills. With the help of this method, the model can concentrate on particular portions of the input text, translating words or phrases individually rather than taking into account the entire context of a sentence or paragraph. Because of their careful consideration of context, LLMs can better provide translations appropriate for the target context and language. When translating idiomatic idioms or culturally unique references, for instance, the model can determine the broader meaning and produce a translation that respects the original meaning instead of translating something literally but sometimes misleadingly (Tannous, 2024). LLMs are especially helpful for complicated or creative translation jobs because they can also reword content and modify sentence structure to increase the translated text's readability and fluency. Moreover, learning and adaptability are built into LLMs like GPT and GEM. With time, these models improve their translation accuracy by honing their linguistic knowledge through additional data and user feedback (Işım & Balcioglu, 2023).

This flexibility is particularly useful in real-time applications where the models must quickly process and translate incoming text, such as chatbots or live translation services. LLMs are perfect for various use cases, from general communication to specialized domains like legal, medical, or technical translations. This is because they can manage enormous volumes of data and deliver reliable translations. Even so, issues must be resolved, especially concerning resolving biases present in the training set and guaranteeing that the models' decision-making processes for translation are transparent.

The use of LLMs in machine translation is a substantial departure from earlier techniques that depended on statistical or rule-based methodologies. Although these previous approaches had some success, they frequently had trouble with idioms, context, and intricate phrase patterns (Tannous, 2024). However, LLMs bring a better comprehension of language and a higher dexterity in handling linguistic complexity. These models will probably play a bigger part in machine translation as they develop further, providing even more complex and customized translation services. The ethical issues surrounding its application, including potential biases, accountability, and the effects of automating translation duties on society, must also be addressed. By tackling these issues, LLMs can be more effectively used to break down language barriers and promote international communication.

### ***Empirical studies***

Translating poetry has attracted the attention of several researchers. Khalifa (2015) examined the complexities of translating Arabic poetry into English. Arabic poetry has been affected by various trends throughout the years, such as neoclassical, romantic, and modernist trends. The neoclassical trend is characterized by following the strictness of classical poetry and its fixed forms. Romantic and modernist trends borrow from Western poetry and are more liberal with format and meter. Regardless of the trend, Arabic poetry should be seen as an amalgamation of elements that distinguish it from other forms of writing. These elements include form, style, metaphor, imagery, rhythm, and rhyme. The study found these features pose challenges in translation. For instance, imagery and metaphor are difficult to render as they are linked to a cultural understanding, and the different backgrounds can change the connotations of the words and their symbolisms. Another issue with the renditions of the words found in Arabic poetry is the richness of Arabic vocabulary, especially that of poems. Moreover, poetry conventions that seem natural due to familiarity in Arabic may be awkward or nonsensical in another language. Finally, it is difficult to balance form, style, and sense in translation, which leads to inevitable losses.

Halimah (2015) examined the effectiveness of a collaborative approach in creative translations. The results indicated that guidance by experts in both the source text and the target language improved the clarity, accuracy, communicative effects, naturalness, and stylistic features of the translations. Mahasneh (2016) found that emotive expressions are seen as problematic areas of translation due to the influence of cultural and political backgrounds on their connotations. However, some expressions can be rendered successfully if they are universal, a shared emotion, or a readable or shareable allusion. The success of the rendition also requires skill and proficiency from the translator.

Mohammed (2023) analyzed the use of compensation in translating Arabic poetry into English and found that the results were not always effective. Compensation took many forms, such as compensation in kind, place, merging, and splitting both on a semantic and syntactic level. Despite the employment of these strategies, the connotations and associations of the source text were not prioritized. Furthermore, the poetic style of the text was not reflected properly in many examples, and many instances of grammatical, pragmatic, and semantic translation losses were highlighted.

Lahiani (2022) found that poetry utilizes linguistic tools such as phonological, grammatical, or lexical devices to emphasize semantic overtones. These features of language exist as contradicting constraints which are often violated in translation. While this violation is inevitable, it exists in a hierarchy which must be followed to optimize the transfer of meaning and style. By analyzing Arabic poems and their translations it was evident that loss is inevitable due to linguistic and cultural differences and the priority of constraints in each language. However, proper application of optimization theory minimizes said loss. Alowedi & Al-Ahdal (2023) compared human and artificial intelligence machine translations of the same poems. The results showed that the machine translations failed at capturing sociocultural nuances as well as stylistic features. This emphasizes the creative skill set necessary for the rendition of poetry.

Lahiani et al. (2023) investigated the application and acceptability of hermeneutic processes in the translation of poetry. The results showed that literary translation can follow either literal, paraphrasing imagery, or creative hermeneutic re-formulations. Creative hermeneutic reformulations involve analyzing the meaning of the text and understanding where it is derived from, such as the words of the text and its stylistic features. This enriches the quality of the translated product as both a reflection of the source text and as a work of poetry in itself. Finally, and most relevantly, Farghal (2023), based on cognitive comprehension of a source poetic text and an emerging cognitive ideation in TL, proposed four types of cognitive structures in poetry translation: culture-free, culture-bound, culture-sensitive, and language-bound. By examining stanzas from Classical Arabic Poetry, he shows how the dynamic interaction and the successful fit between cognitive comprehension in SL and cognitive ideation in TL determines the degree of coherence and poetic cohesiveness of the translation. Thus, a poetic translation may range between optimally and minimally coherent or cohesive. While coherence is meant to capture content in terms of thematic clarity, poetic cohesiveness seeks to embody formal features in terms of metaphor and prosody.

Yilmaz et al. (2023) discussed the risk that artificial intelligence (AI) tools bring to human employment, especially for translators. The study looked at ChatGPT's translation capabilities as well as the neural network-based translation system. According to the findings, fewer translation tasks requiring analytical translation as opposed to emotional or cultural translation were given by these technologies.

Işim & Balcioglu (2023) evaluated ChatGPT on multiple benchmark test sets, determining its accuracy and recommended translation prompts before assessing its effectiveness in translating 50 Turkish paragraphs taken from educational papers into English. According to the study, it operates effectively. The evaluation validated the idea of potentially becoming a dependent translator and demonstrated a high degree of accuracy in translating such materials.

Jiao et al. (2023) assessed the effectiveness of ChatGPT as a translation tool by putting its many translation prompts to the test and rating each one's output. They discovered that every prompt delivered accurate translations. ChatGPT and commercial translation services like Google Translate were contrasted in the same study. In terms of accuracy, the results demonstrated that ChatGPT performs better when translating languages with high resources, but not as effectively when translating languages with low resources. Furthermore, spoken language was where ChatGPT excelled more than biological abstracts or Reddit comments. A novel approach to addressing the issue of low-resource language translation was also investigated in this study: 'pivot prompting'. To accomplish this, ChatGPT is first asked to translate

the given text into the desired language, and then asked to convert it into a high-resource language. The study comes to the conclusion that ChatGPT-4 engine improves ChatGPT translation quality, even for languages with limited resources.

Khoshafah (2023) compared ChatGPT's translations to human translations in order to assess the system's accuracy in translating writings from various genres between English and Arabic. The results demonstrated that ChatGPT can only interpret straightforward settings; human intervention is finally required. It guarantees that caution must be exercised while employing this tool to translate materials of this nature. Additionally, human evaluation and correction are necessary to guarantee that the translated content is sufficiently accurate.

Al-Salman & Haider (2024) assessed the accuracy of well-known machine translation (MT) and artificial intelligence (AI) platforms, such as Google Translate, ChatGPT, and Gemini, in translating Arabic research titles into English for the humanities and social sciences. This study aims to assist researchers in determining which translation system works best for their research titles. They found that the three methods produced various meaning or sentence structure mistakes. It is noteworthy that although human translations contained errors as well, especially in grammar, their vocabulary was the most accurate. Of the three systems, Gemini leads the curve in this domain because it generated the fewest errors.

Mohsen (2024) assessed the performance of Google Translate (GT) when compared with Large Language Models (LLMs)—specifically, ChatGPT 3.5 and 4—in translating 20 academic abstracts between Arabic and English. Using a mixed-methods approach and utilizing some evaluation criteria that include semantic integrity, syntactic coherence, and technical adequacy, the study found that LLMs—ChatGPT 4 in particular—significantly outperform GT in both translation directions, with ChatGPT 4 demonstrating superior performance in Arabic-English translation. Similarly, Wang (2024) contrasted ChatGPT-4o's evaluations with those of human assessors for two categories of texts: literary and non-literary. According to the findings, ChatGPT-4o exhibited excellent consistency with human assessors, especially when it comes to fidelity, correctness, and fluency. The translation quality was enhanced by ChatGPT-4o's feedback and recommended changes, indicating its potential as a useful tool for translation practice and instruction. The study emphasized the necessity of human supervision to guarantee thorough evaluations of translation quality.

## Methodology

Classical Arabic poetic verses were selected to act as a base for representative data. These verses were then translated into English by human and machine translations. The translations were then evaluated and compared on the basis of thematic clarity, creativity, and prosody.

## Corpus of study

The source texts comprised 15 verses of classical Arabic poetry. The selected data represented a diverse range of historical and cultural contexts, linguistic complexity, and theme and content. The researchers carefully carried out the literal translation that is shown in Table 1. The aim of this translation was to faithfully depict every nuance of the original material while maintaining the exact wording and structure of the original work as closely as possible.

The target text comprised two sections: human translation (H) and AI technology (GEM and GPT). To ensure the desired level of quality, machine and human translators were asked to prioritize thematic clarity, creativity, and prosody in their renditions. A professor and renowned expert in the fields of linguistics and translation studies with a profound knowledge of poetic discourse translated the lines from classical poetry that are included in this paper. He is well-known for his decades-long experience and great competence in translating. He has translated several works between Arabic and English, showcasing his extensive knowledge of both languages and their subtle literary differences. In addition to countless articles, book chapters, and critical edits of translations done by other experts, his impressive portfolio consists of several translated volumes.

His skill in translating poetry is very impressive, and his works have brought great value to the subject of translation studies, demonstrating his ability to preserve the spirit and aesthetic appeal of the source



**Table 1.** The investigated verses with their Literal Translation in English.

No	Source Arabic Verse	Literal Translation
1	ولا خير في ود امرئ متلون	There is no good in the friendship of a fickle man,
2	إذا الريح مالت مال خيئ تميئ	When the wind blows, he blows in its direction
3	بالعلم والمال يبني الناس ملكهم	With knowledge and wealth, people build their kingdom,
4	لم يبن ملك على جهل وإقلال	No kingdom is built on ignorance and poverty.
5	وإذا انتك مذمتي من ناقص	If you are criticized by someone who is flawed,
6	فهي الشهادة لي باني كامل	Then it is a testament to your perfection.
7	بذا قضت الأيام ما بين أهلها	Thus the days have decreed for their people,
8	مصائب قوم عند قوم فوائد	The misfortunes of one are the gains of another.
9	كل ابن انثى وإن طالت سلامته	Every son of a woman, however long his life,
10	يوما على آلة حدياء محمول	Will one day be carried on a humped bier.
11	العلم صيد والكتابة قيده	Knowledge is a hunt, and writing is its leash,
12	قيد صيودك بالحبال الوثيقة	Secure your catch with strong ropes.
13	لكل شيء إذا ما تم نقصان	For all things, when they are complete, will lack,
14	فلا يُغر بطيب العيش إنسان	So let no man be deluded by the good life.
15	نقل فؤادك حيث شئت من الهوى	Move your heart wherever you wish from love,
16	ما الحب إلا للحبيب الأول	Love is only for the first beloved.
17	أعلل النفس بالأمال أرقها	Hope is the light that guides our way,
18	ما أضيق العيش لولا فسحة الأمل	Without it, life would be a dark and dreary day.
19	تداويت من ليلي بليلى عن الهوى	I sought to heal myself of Layla with Layla,
20	كما يتداوى شارب الخمر بالخم	Like a drunkard seeking to heal himself with wine.
21	إنما الأمم الأخلاق ما بقيت	The nations are only morals, when they remain,
22	فإن هم ذهبت أخلاقهم ذهبوا	But if they are gone, they are gone.
23	دعوت على عمرو فمات فسرتني	I cursed Omar, and he died, and that pleased me,
24	ولما أتى زيد بكيت على عمرو	But when Zayd came, I wept for Omar.
25	وتعظم في عين الصغير صغارها	Small things seem large in the eyes of the small,
26	وتصغر في عين العظيم العظائم	Great things seem small in the eyes of the great.
27	لا تحسبوا رقصي بينكم طربا	Do not think my dance among you is for joy,
28	فالطير يرقص مذبوحا من الألم	For the bird dances, though it is slaughtered with pain.
29	سئمت تكاليف الحياة ومن يعيش	I am tired of the costs of life, and whoever lives
30	ثمانين حولاً لا أبا لك يسأم	Eighty years, may you not have a father, will get tired

texts while rendering them understandable to a wider readership. His vast experience and scholarly contributions solidify his reputation as a highly esteemed person in the translation profession

The same verses were input in GEM and GPT with the following prompt: *'Based on the theme of the following Arabic verse, give an AA English rhyming couplet. Don't miss prosody. The last words in the two parts should rhyme. Suggest one and only one translation.'* The results are shown in Table 2.

### Evaluation and procedures

After the data is compiled, the renditions are evaluated by 54 human experts in translation, language, and literature, as well as the two AI systems. It is worth noting that the current study's ethical approval was obtained from the Deanship of Scientific Research at the Applied Science Private University in Jordan with the approval number (FAH/2023–2024/15). Also, a written informed consent was obtained from all subjects before the study was conducted.

The evaluations measure the retention of meaning and poetic features by assessing thematic clarity, creativity, and prosody, as Table 3 shows. The prompt was as follows: *'Please rate the English translations below to the Arabic verse in terms of thematic clarity, creativity, and prosody, with 5 being the highest value and 1 being the lowest.'*

We did not indicate in the questionnaire whether AI tools or humans performed the translation. This was done on purpose to prevent influencing the participants' responses.

### Analysis and findings

#### University professors' responses to the translations

Table 4 shows the Means of university professors' responses.

The overall picture in Table 4 shows that the participants' assessment of human translation on the three variables together came first (H 3.96) followed with a slight difference by GPT (3.75). GEM, however, considerably lagged behind (3.18). This finding shows that, despite the introduction of new MT systems like GPT and GEM, which are highly competitive, the human translator still maintains a small lead in

**Table 2.** Human and AI renditions to the investigated verses.

No	Human (H)	GEM	GPT
1	Honor not the company of a capricious guy; Who tilts with the wind and does always lie.	A friend who changes with the wind is not a friend, Their loyalty is as fleeting as a summer's trend.	No good in the love of one whose heart's a guise, When winds shift, love shifts, where inclination lies.
2	By knowledge and wealth people build up their dignity No status can be built upon ignorance and paucity.	Knowledge and wealth are the keys to power, Ignorance and apathy will lead to its downfall.	In knowledge and in wealth, people forge their reign, No kingdom stands on ignorance and disdain.
3	Insult aimed at me from a mean and base guy Duly testifies to my being gracious and high.	The criticism of a fool is praise, For it only confirms your worth.	When criticism comes from one with lacking sight, It testifies, my perfection stands in the light.
4	So have the days among their folks ruled; One man's plight be another man's food.	The world is a wheel that turns, For every loss, there is a gain.	Thus, days decree 'mongst those who dwell, One's woes, to others, tales to tell.
5	A woman's offspring may live long and rule; Only one day to be carried on a concave tool.	Death is the great equalizer, For all must face the final journey.	Every son of woman, though his well-being lasts, One day, like a burden, on the stretcher, he's cast.
6	Knowledge is a hunt and writing its chain Never let talk fly away after so much pain.	Knowledge is a precious thing, Write it down so it will not be lost.	Knowledge, the prey; and writing, its tether strong, Bind your hunting with secure ropes along.
7	Decline begins once perfection obtains; So, never pride yourself on luxury gains.	All things, once complete, will fade away, So do not be deceived by life's sweet sway.	For everything, when perfected, flaws appear, Let not a pleasant life delude, sincere.
8	Let heart travel as far as it can reach; First love will surely end the search.	The heart can wander far and wide, But love belongs to the first love.	Transplant your heart wherever desire leads, True love belongs to the first love it breeds.
9	I reckon with self by desiring to cope; How hard life is without room for hope.	I comfort my soul with hopes and watch over them, How narrow life would be if not for the space of hope.	Adorn your soul with hopes, watch them unfurl, Life's narrow without the expanse of hope's swirl.
10	In Layla did I seek a cure from Layla's affection; The way a drinker seeks a cure in more addiction.	I sought to cure myself of love with Layla, But it was like a drunkard trying to cure himself with wine.	I sought a cure from love through Leila's grace, As drinkers heal with the same drink they embrace.
11	Nations flourish as far as their ethics boom; If ethics are gone, soon will they suffer doom.	A nation's worth is measured by its morals, When morals fade, the nation follows.	Nations endure through ethics' lasting creed, When virtues fade, so too do nations recede.
12	I prayed against Amr and I'm glad he died; When Zayd took over, for Amr I bitterly cried.	I wished for Omar's death, and it came to pass, But when Zayd arrived, I wept for Omar's loss.	I prayed against Amr, he met his fate, Yet, when Zaid arrived, tears flowed for Amr, the great.
13	In mean men's eyes a trifle magnifies; But little is a treasure in great men's eyes.	The small things seem large to the small, The great things seem small to the great.	I prayed against Amr, he met his fate, Yet, when Zaid arrived, tears flowed for Amr, the great.
14	Think not of my dancing among you as rapture; A slaughtered bird does dance out of torture.	Do not think my dance is a sign of joy, The bird dances, but its heart is torn with grief.	Don't think my dance amid you brings delight, As birds dance, wounded, in sorrow's tight.
15	So weary have I become of life's burdens and strife Eighty years circling, no wonder, consumes one's life	The burden of life is heavy, And even the wise grow weary.	The burden of life, weary have I grown, Eighty cycles passed, no kin to call my own.

**Table 3.** Sample of the Evaluation questionnaire.

English Translation	Thematic clarity					Creativity					Prosody				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
<p>نَقَلَ فُؤَادَكَ حَيْثُ شِئْتَ مِنَ الْهَوَى</p> <p>مَا الْحُبَّ إِلَّا لِلْحَبِيبِ الْأَوَّلِ</p>															
<p>Let heart travel as far as it can reach;</p> <p>First love will surely end the search.</p> <p>The heart can wander far and wide,</p> <p>But love belongs to the first love.</p> <p>Transplant your heart wherever desire leads,</p> <p>True love belongs to the first love it breeds.</p>															



**Table 4.** Means of University professors' responses.

No	Thematic clarity			Creativity			Prosody		
	H	GEM	GPT	H	GEM	GPT	H	GEM	GPT
1	3.94	3.76	3.41	3.63	2.88	3.24	4.12	3.76	3.35
2	4.06	3.53	3.71	3.63	3.18	3.76	3.88	2.47	3.82
3	3.82	3.41	3.82	3.76	3.24	3.53	4.29	2.65	3.82
4	3.88	3.41	3.53	3.76	3.82	3.82	4	3.18	3.76
5	3.82	3.47	3.41	3.47	3.53	3.29	4.12	2.75	3.53
6	4	3.53	3.75	3.82	2.76	3.56	4.24	2.53	3.69
7	4.06	4.06	3.53	3.76	3.82	3.59	4.18	4.06	3.65
8	3.94	3.47	3.76	3.65	2.88	4.12	3.53	2.53	4.35
9	4.06	3.41	3.76	4	3	3.59	4.29	2.47	3.53
10	3.82	3.59	4	3.41	2.88	4	4.24	2.41	4.12
11	4.47	3.82	3.94	4.06	2.94	3.76	4.29	2.88	4.06
12	4	3.59	3.65	3.82	3.29	3.71	4.18	3.06	3.53
13	4.06	3.41	3.82	3.82	2.87	3.88	4.18	2.44	3.88
14	4.06	3.94	4	4	3.12	4	4.24	2.41	4.24
15	4.06	3.35	3.76	3.82	2.94	3.88	4.06	2.88	3.82
Total	4.00	3.58	3.72	3.76	3.14	3.72	4.12	2.83	3.81
Overall		H	3.96		GEM	3.18		GPT	3.75

**Table 5.** Human and AI renditions to verse 6.

Source Text	Human & AI Technology	Translation
العلم صيد والكتابة قيد قيد صيودك بالحبال الوائقة	H	Knowledge is a hunt and writing its chain; Never let talk fly away after so much pain.
	GEM	Knowledge is a precious thing; Write it down so it will not be lost.
	GPT	Knowledge, the prey; and writing, its tether strong; Bind your hunting with secure ropes along.

**Table 6.** Human and AI renditions to verse 8.

Source Text	Human & AI System	Translation
نقل فؤادك حيث شئت من الهوى ما الحب إلا للحبيب الأول	H	Let heart travel as far as it can reach; First love will surely end the search.
	GEM	The heart can wander far and wide, But love belongs to the first love.
	GPT	Transplant your heart wherever desire leads, True love belongs to the first love it breeds.

**Table 7.** Human and Machine renditions to verse 1.

Source Text	Human & AI System	Translation
ولا خير في ود امرئ متلون إذا الريخ مالت مال حيث تميل	H	Honor not the company of a capricious guy; Who tilts with the wind and does always lie.
	GEM	A friend who changes with the wind is not a friend; Their loyalty is as fleeting as a summer's trend.
	GPT	No good in the love of one whose heart's a guise; When winds shift, love shifts, where inclination lies.

translating classical Arabic poetry with regard to thematic clarity, creativity, and prosody. Table 4 also shows that this small lead is maintained for the individual variables, viz. (H 4, GEM 3.76, GPT 4.12 vs. H 3.58, GEM 3.14, GPT 2.82 vs. 3.72, 3.72, 3.81, respectively).

A closer look at Table 4 shows that, while human, GEM and GPT translators performed similarly in terms of thematic clarity (H 4 vs. GEM 3.58 vs. GPT 3.72), GEM significantly fell behind in terms of prosody (GEM 2.82) and, to a lesser degree, in creativity (GEM 3.14). Apparently, GEM is not sufficiently equipped with enough data that can handle formal poetic features, prosody in particular. By contrast, GPT showed impressive poetic potentials that came close to human translation in terms of creativity (GPT 3.72 vs. H 376). These findings indicate that the cognitive load of the verse is the easiest to render for the translators, both human and AI translators. However, formal poetic features, prosody in particular, can be challenging to both, with AI translation emerging as a real competitor.

The following examples (Tables 5–7) shed light on how human and AI translators have dealt with the three poetic features.

In example 1 in Table 5, the human and machine translators have performed similarly on thematic clarity (H 4 vs. GEM 3.82 vs. GPT 4.24, respectively), which indicates that the thematic element is reasonably transparent to the three of them. However, the assessment of creativity significantly favors both human translation (H 3.82) and GPT translation (3.56) over GEM translation (2.76). This can be readily justified by the fact that GEM translation does away with ‘the hunting metaphor’ by paraphrasing it into literal language in contrast with the other two translations in which the creative metaphor ‘likening knowledge to a hunt’ is preserved. Likewise, GEM translation seriously lags behind in terms of prosody (2.53) as opposed to the good performance of human translation (H 4.24) and GPT translation (3.69). GEM’s poor performance on prosody is consistent in the bulk of the corpus (in only 4 of the 15 verses does it score a mean of above 3, earning an overall mean of only 2.83). By contrast, human and GPT translation have performed consistently well, with no score below 3 for any of the verses, thus earning overall means of 4.12 and 3.81, respectively.

For its turn, example 2, in Table 6, shows two similar observations: the three translators have performed well on thematic clarity (H 3.94, GEM 3.47, and GPT 3.76) and GEM continues to do poorly on creativity (2.88) and prosody (2.53).

Interestingly enough, however, in this verse in Table 6, GPT outperforms the human translator on prosody (GPT 4.35 vs. H 3.53). Apparently, the high degree of metaphoricity in GPT translation (GPT 4.12) has joined forces with prosody to have it excel the human translation, which is modestly metaphorical (H 3.65). In fact, this is the only verse where GPT beats the human translator on the assessment of prosody.

Finally, example 3, in Table 7, is the only verse in which GEM beats GPT on prosody (GEM 3.76 vs. GPT 3.35) although both of them are outperformed by the human translator (H 4.12).

As Table 7 shows, while the economy and familiar vocabulary of the human translation may explain its rating better on prosody than both machine translations, GEM’s surpassing GPT here may be explained by the impact of familiar vocabulary on participants, which may have affected thematic clarity in favor of GEM (GEM 3.76 vs. GPT 3.41). In terms of creativity, the three translations have maintained the ‘wind metaphor’ around which the verse is built, despite the fact that GEM’s rating for creativity (GEM 2.88) lags behind GPT’s (GPT 3.24) and the human translator’s (H 3.63). Surprisingly, however, when GEM was asked to rate its rendition of the Arabic verse in terms of prosody, it was given only 1 while GPT’s was given 4. For its part, GPT gave GEM 4 while giving itself 3 (see below for more discussion).

### *AI-systems’ evaluation of poetic translations*

Both GEM and GPT are asked to rate the poetic translations, using the following rubric ‘**Out of 5 (1 lowest, 5 highest), rate the three poetic translations of the accompanying Arabic verse in terms of thematic clarity, creativity, and prosody.**’ Let us first examine GEM’s Means in Tables 8.

In Table 8, compared with overall participants’ assessment of the three variables (H 3.96, GEM 3.18, and GPT 3.75), GEM’s evaluation (H 2.44, GEM 2.40, and GPT 4.25) is significantly lower. In fact, it is

**Table 8.** Means of GEM.

No	Thematic clarity			Creativity			Prosody		
	Human	GEM	GPT	Human	GEM	GPT	Human	GEM	GPT
1	4	4	5	3	2	4	2	1	4
2	3	4	5	2	3	5	1	1	4
3	3	4	5	2	3	4	1	1	4
4	4	3	5	3	2	4	1	1	4
5	2	3	5	1	2	4	1	1	4
6	3	2	5	2	1	4	3	1	4
7	4	4	5	3	2	4	1	1	4
8	4	5	3	3	2	4	1	1	4
9	3	4	5	2	3	4	1	1	4
10	4	3	3	3	2	4	1	1	4
11	4	5	5	3	2	4	1	1	4
12	4	5	5	2	3	4	1	1	4
13	4	4	4	3	2	4	1	1	4
14	4	5	5	3	2	4	1	1	4
15	4	5	5	3	2	4	1	1	4
Total	3.60	4.00	4.67	2.53	2.20	4.07	1.20	1.00	4.00
Overall	H		2.44	GEM		2.40	GPT		4.25

**Table 9.** GEM's assessment of prosody.

Source Text	MT System	Translation
لُكُلْ شَيْءٌ إِذَا مَا تَمَّ نُقْصَانُ فَلَا يُغَيِّرُ بِطَيْبِ الْعَيْشِ إِنْسَانُ	Human	Decline begins once perfection obtains; So, never pride yourself on luxury gains.
	GEM	All things, once complete, will fade away, So do not be deceived by life's sweet sway.
	GPT	For everything, when perfected, flaws appear, Let not a pleasant life delude, sincere.

seriously inconsistent with the participants' assessment of two variables, viz. what is clearly positive by participants (creativity 3.76 and prosody 4.12) is downgraded by GEM (2.20 and 1 respectively) and only relatively congruent with the assessment of thematic clarity (H 4 and GEM 3.60). This finding may indicate GEM's instability and inconsistency in assessing poetic translations on two key variables (see discussion below).

Table 8 also shows that GEM significantly favours GPT's translations on the three variables individually (H 3.60 vs. GEM 4 vs. GPT 4.67, H 2.53 vs. GEM 2.20. vs. GPT 4.07, and H 1.20 vs. GEM 1 vs. GPT 4). While its assessment of thematic clarity is positive for the three translations on thematic clarity, GEM has seriously downgraded the rating of creativity and prosody for both the human translator and its own translations, even more so for the latter. Apparently, GEM's AI software is confused when it comes to assessing creativity and prosody, though it is relatively equipped with recognizing thematic clarity.

Let us consider the following example (Table 9) to see how confused GEM's AI system in terms of prosody.

In Table 9, while thematic clarity has received a highly positive assessment for the three translators (H 4, GEM 4, and GPT 5) and a positive assessment on creativity for human and GPT translators (H 3 and GPT 4, but not for its own translation, GEM 2), it has accorded GPT 4 on prosody, and human translation and its own translation only a rating of 1. However, by examining the three translations in 4, we can readily see that they are similarly prosodic, following the AA rhymed scheme adopted in this study. This simply indicates that the GEM's AI system is confused and not sensitive enough to prosody in poetic discourse, which is the single key feature that distinguishes it from other genres of discourse. This is also confirmed by the participants' low assessment of GEM's performance on prosody (Table 8 above).

To check the consistency of GEM's assessment, using the same rubric, the same Arabic verse along with its three English translations is given to it a few days later for reassessing thematic clarity, creativity, and prosody. Surprisingly, the results are far from being consistent, to the point of being reversed on prosody. The new ratings for thematic clarity (H 2, GEM 4, GPT 3) are lower than the previous ones (H 4, GEM 4, GPT 3) and are even lower for creativity (H 1, GEM 2, GPT 1.5 vs. H 3, GEM 4, GPT 3). What is shocking, however, is the reassessment of prosody, which almost reversed, viz. H 3, GEM 4, GPT 2 (new) vs. H 1, GEM 1, GPT 4 (previous). In this way, what has been downgraded in the previous ratings are significantly upgraded in the new ratings. This further sheds light on GEM AI system's inconsistency in evaluating poetic features.

In contrast with GEM's results, compared with participants' overall assessment of thematic clarity, creativity and prosody, GPT's evaluation seems similar and consistent, viz. GPT 3.49, 3.42, 3.60 vs. H 3.96, 3.75, 3.18, with no rating below 3. This finding clearly indicates that GPT works with a sensitivity to key poetic features comparable to that of human informants' (Table 10).

In Table 10, taking the three variables together, the marginal difference in the overall performance between participants' assessment (H 3.63) and GPT's assessment (GPT 3.50) is negligible, which clearly proves that GPT can greatly help in translating poetic discourse and can, in fact, be a real competitor with the human translator in this regard. However, Table 10 shows that GPT's assessment of GEM's performance on prosody (GEM 3.53) is markedly higher than participants' evaluation (GEM 2.83). Apparently, GPT is less sensitive to end rhyme as a key poetic feature than human assessors and is more influenced by thematic clarity and creativity. In fact, the choice of the AA end rhyme of the two hemistichs in translating an Arabic verse is adopted in this study to reflect the importance of rhyme in the prosody of Classical Arabic poetry. Despite the clear rubric to GPT and GEM to give an AA rhymed couplet for

**Table 10.** Means of GPT.

No	Thematic clarity			Creativity			Prosody		
	Human	GEM	GPT	Human	GEM	GPT	Human	GEM	GPT
1	4	4	4	3	4	3	3	4	3
2	5	4	5	3	4	3	4	3	4
3	4	5	4	3	4	3	3	4	3
4	4	3	4	4	3	3	3	3	4
5	3	4	4	3	4	3	3	4	3
6	4	3	4	3	3	3	3	4	3
7	4	4	3	3	4	3	3	4	3
8	4	3	4	3	3	4	4	3	4
9	4	4	4	3	3	4	3	4	4
10	4	4	4	4	3	3	3	4	3
11	4	4	4	3	3	3	3	4	3
12	4	4	3	3	3	3	3	4	3
13	4	5	4	3	4	3	5	3	4
14	4	4	3	3	4	3	4	3	3
15	4	3	3	3	2	3	3	2	3
Total	4.00	3.87	3.80	3.13	3.40	3.13	3.33	3.53	3.33
Overall	Human			GEM			GPT		
			3.49			3.42			3.60

each Arabic verse, only 3 out of the 15 GEM's couplets exhibited this, in contrast with 15/15 by GPT. GPT's assessment of GEM's performance on prosody, therefore, seems to be lenient for this feature, possibly influenced by other considerations.

Finally, to do justice to GEM, the same verse in Table 8 above was given to GPT for reassessment a few days later. The results came as follows: thematic clarity (H 4, GEM 5, GPT 3), creativity (H 3, G. 4, GPT. 5), prosody (H 4, GEM 3, GPT 5). The overall ratings are clearly positive, viz. H 3.67, GEM 4, GPT 4.33, in contrast with GEM which downgraded both human translation (H 1.5) and its own translation (GEM 1) on prosody, as well as its translation on creativity (GPT 2). In terms of consistency with previous results on this verse, viz. thematic clarity (H 4, GEM 3, GPT 3), creativity (H 3, GEM 4, GPT 3), and prosody (H 3, GEM 4, GPT 3), the consistency is generally maintained, viz. overall ratings H 3.33, GEM 3.66, and GPT 3, though lower than new ones. Apparently in its reassessment, GPT found GEM's rendition more metaphorical than human and its own rendition, which concurs with the participants' assessment of this verse for creativity (H 3.76, GEM 3.82, GPT 3.59), as well as participants' evaluation of prosody (GEM 4.06 vs. GPT 3.65), which almost ties with their own judgment (H 4.18).

## Conclusions and recommendations

While the findings show that human poetic translation still maintains a small lead over AI machine systems (GEM and GPT) in poetry translation based on the evaluation by 54 university professors, the latter seem to be true competitors in this regard. In particular, GPT's assessment comes close to that of human translation in terms of thematic clarity, creativity, and prosody. This simply means that AI systems are making a huge stride not only in translating texts written for general purposes, but also for discourse produced for specific purposes, with poetry translation being viewed as the most challenging task due to both content and form complexity.

Predictably, thematic clarity and creativity, which represent the cognitive load of the verse, are the easiest to capture by both human, GEM and GPT translators and has, therefore, received a high and relatively similar assessment by participants (H 4, GEM 3.58, GPT 3.72) and (H 3.76, GEM 3.14, GPT 3.72). However, while human and GPT translators maintain participants' high assessment of prosody (H 4.12, GPT 3.81), GEM significantly lags behind (GEM 2.82). This finding indicates that GEM is not sufficiently equipped with enough data that can handle prosody as a poetic feature, which is further confirmed by the fact that GEM output has missed rhyme in 12 out of the 15 verses used in this study. By contrast, human and GPT output has captured rhyme in the entire sample.

GPT outperforms GEM in handling prosody. This could be due to the models' architecture, training set, and fine-tuning procedures. Prosody necessitates a thorough comprehension of language's phonetic and semantic aspects. In translation, it requires a model that can identify and reproduce the complex patterns that give rise to rhyme, meter, and overall flow, particularly in poetic contexts.

The large and varied corpus of text data that GPT was trained on probably contains a sizable portion of poetry and other literary works emphasizing prosodic elements. Through this exposure, GPT can pick up prosody-related patterns like rhyme schemes, meter, and the rhythmic flow of words. GPT is better at maintaining prosodic components in translation because of the training data's depth and diversity, allowing it to generalize these patterns across many settings. Alternatively, GEM's training set may be less extensive or less poetic structure-focused, which would limit its capacity to precisely recreate these traits.

More complex fine-tuning procedures might contribute to GPT's exceptional prosody handling capabilities. The instance of GPT could include a targeted dataset of translated poetry or texts with a strong emphasis on prosodic elements. Fine-tuning entails further training the model using task-specific data. The deliberate refinement enables GPT to improve its proficiency in handling prosodic nuances, including preserving rhyme and rhythm, which are essential for translating poetry. GEM would find it difficult to reproduce these qualities if it were not similarly fine-tuned on prosody-rich data, resulting in the output's reported shortcomings.

Poetry frequently uses non-linear language, challenging conventional translation models with its inventive use of syntax, metaphor, and wordplay. Because of its generative characteristics, which are intended to tackle creative language tasks, GPT is more capable of navigating through these complications. Because of its adaptability, it can produce translations while preserving the style and organization of the source poetry. In comparison, GEM's methodology might be more inflexible, resulting in translations that overlook the subtle prosodic elements crucial to poetry.

As for AI systems' assessment of the translations, GPT proves to be generally consistent with the overall participants' evaluation of the three poetic variables (GEM 3.49, 3.42, 3.60 and H 3.96, 3.75, 3.18), which may indicate that its data uses a model of computing similar to that of human reasoning. GEM's assessment, however, is markedly inconsistent with participants' evaluation of prosody (GEM 1) and, to a lesser degree, creativity (GEM 2.20).

To optimize the use of both GEM and GPT in translation in general and in poetry translation in particular, the interactive nature of these AI systems should be explored. Therefore, this study may be followed by another investigation to show how the human translator can progressively prompt the machine interactively to get the best results in translation activity.

The current study mostly used GPT-3.5, which is regarded as less sophisticated than GPT-4.0 and its later editions. Furthermore, the performance of several LLMs, such as Claude 3.5, was not investigated in this work. Therefore, it is recommended that future studies use these advanced models, bearing into account that previous studies have shown that advanced LLMs have a great potential to produce high-quality translation outputs (Alkhatib & Haider, 2024; Mohsen, 2024; Wang, 2024).

Given that free verse has distinct structural and thematic qualities from classical Arabic poetry, the dynamics seen in this study may change when translating non-traditional Arabic poetry. In contrast to the inflexible metrical patterns and rhyme schemes of Classical Arabic poetry, free verse allows for more creative freedom in both form and emotion. This flexibility may bring new opportunities and challenges for thematic clarity, creative prosody, and AI model representation for both human translators and models.

Future research may take into account the following factors:

- **Thematic Flexibility:** With a looser framework, free verse frequently examines modern themes. This might free AI models from the limitations of conventional prosody, enabling models like GEM and GPT to better capture thematic nuances.
- **Prosodic Challenges:** Free verse's erratic rhythms may either test or help AI models in a different way than classical verse, which mostly depends on meter and rhyme. Future research could investigate how well AI compares to traditional forms in preserving the intended melody and flow of free verse.
- **Creativity in Translation:** Free verse may allow AI models to exhibit more creative flexibility because it has less formal constraints. Examining how AI and human translators negotiate this flexibility may provide light on the advantages and disadvantages that each has when working with less regimented literary forms.

The expansion of this study, therefore, may contribute to a more comprehensive understanding of the capabilities of AI systems in many Arabic poetry genres, enhancing the conversation on AI's place in literary translation.

This research could benefit different parties, including translators and translation scholars, technology developers, and educational institutions. Human translators and scholars of translation and literature can expand their understanding of literature and its translation through this analysis of the production of machine translation. This can guide their work and studies through the proper application of machine translation. This research provided some insight into the challenges that AI translation systems face, which can guide developers in the refinement of algorithms. This can contribute to the advancement of the quality of machine translation in this genre.

## Ethical approval

The current study's ethical approval was obtained from the Deanship of Scientific Research at the Applied Science Private University in Jordan with the approval number (FAH/2023–2024/15). Also, a written informed consent was obtained from all subjects before the study was conducted.

## Authors' contributions

Mohammed Farghal: Conceptualization; study design; data analysis and interpretation; drafting of the manuscript; final approval of the version to be published. Ahmad S. Haider: Conceptualization; study design; data analysis and interpretation; critical revision of the manuscript for intellectual content; final approval of the version to be published.

## Disclosure statement

The authors declare no competing interests.

## Funding

This research received grant no. (111/2023) from the Arab Observatory for Translation (an affiliate of ALECSO), which is supported by the Literature, Publishing & Translation Commission in Saudi Arabia. (Although the funding body, the Arab Observatory for Translation (AOT), is not affiliated with Applied Science Private University, it is an independent entity under ALECSO, supported by the Literature, Publishing, and Translation Commission. AOT provides grants to support high-quality translation research in the Arab world, addressing key topics and fostering collaboration among researchers and research groups.) see <https://aotalecso.org/en/grants>

## About the authors

**Mohammed Farghal** (PhD in General Linguistics, 1986) is currently Professor of Linguistics and Translation/Department of English and Translation/Applied Science University. He has authored/coauthored more than a 100 scholarly papers in linguistics and translation studies, which have been published in highly recognized international linguistics and/or translation studies journals. He is also the author of many books, including *Advanced Issues in Arabic-English Studies* (2012), *Contextualizing Translation Theories* (2015/with A. Almanna), and three volumes in a series of 'Arabic-English translation studies' (2015, 2017, 2019). Also, he has made many translations of important literary works, including Cormac McCarthy's *The Road* (2006).

**Ahmad S. Haider** (Corresponding) received his PhD in linguistics from the University of Canterbury, New Zealand. He is an associate professor in the Department of English Language and Translation at Applied Science Private University, Amman, Jordan. He is also a researcher at the MEU Research Unit, Middle East University, Amman, Jordan. His current research focuses on how political events are socially, discursively, and linguistically represented in media, combining corpus linguistics and (critical) discourse analysis. His main areas of interest include corpus linguistics, discourse analysis, pragmatics, and translation studies.

## ORCID

Mohammed Farghal  <http://orcid.org/0000-0002-5012-8550>

Ahmad S. Haider  <http://orcid.org/0000-0002-7763-201X>

## Data availability statement

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.



## References

- Ahmed, M. A., Hasan, R. A., Ali, A. H., & Mohammed, M. A. (2019). The classification of the modern arabic poetry using machine learning. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 17(5), 2667–2674. <https://doi.org/10.12928/telkomnika.v17i5.12646>
- Akan, M. F., Karim, M. R., & Chowdhury, A. M. K. (2019). An analysis of Arabic-English translation: Problems and prospects. *Advances in Language and Literary Studies*, 10(1), 58–65. <https://doi.org/10.7575/aiac.all.v.10n.1p.58>
- Alkhatib, R., & Haider, A. S. (2024). Subtitling legal expressions in English series into Arabic by Netflix, machine, and artificial intelligence. *Pakistan Journal of Criminology*, 14(4), 513–528. <https://doi.org/10.62271/pjc.16.4.513.528>
- Alowed, N. A., & Al-Ahdal, A. A. M. H. (2023). Artificial intelligence based Arabic-to-English machine versus human translation of poetry: An analytical study of outcomes. *Journal of Namibian Studies: History Politics Culture*, 33, 1523–1538. <https://doi.org/10.59670/jns.v33i.800>
- Al-Salman, S., & Haider, A. S. (2024). Assessing the accuracy of MT and AI tools in translating humanities or social sciences Arabic research titles into English: Evidence from Google Translate, Gemini, and ChatGPT. *International Journal of Data and Network Science*, 8(4), 2483–2498. <https://doi.org/10.5267/j.ijdns.2024.5.009>
- Al-Shaibani, M. S., Alyafeai, Z., & Ahmad, I. (2020). Meter classification of Arabic poems using deep bidirectional recurrent neural networks. *Pattern Recognition Letters*, 136, 1–7. <https://doi.org/10.1016/j.patrec.2020.05.028>
- Al-Talabani, A. K. (2020). Automatic recognition of arabic poetry meter from speech signal using long short-term memory and support vector machine. *Aro-the Scientific Journal of Koya University*, 8(1), 50–54. <https://doi.org/10.14500/aro.10631>
- Farghal, M. (2023). *Highlights from Classical Arabic Poetry in English Translation: A Schematic Model*. Eliva Press, Republic of Moldova.
- Haider, A. S., & Shohaibar, R. S. (2024). Netflix English subtitling of idioms in Egyptian movies: challenges and strategies. *Humanities & Social Sciences Reviews*, 11(1), 1–13. <https://doi.org/10.1057/s41599-024-03327-4>
- Halimah, A. M. (2015). Teaching translation of poetry: A collaborative approach. *English Language and Literature Studies*, 5(1), 26–36. <https://doi.org/10.5539/ells.v5n1p26>
- Harb, L. (2020). Arabic literary theory. In J. E. Muñoz (Ed.), *Oxford research encyclopedia of literature* (pp. 1–25). NY: Oxford University Press.
- Igbaria, K. (2020). Arab spring revolutions throughout modern Arabic poetry. *Asian Journal of Humanities and Social Studies*, 8(4), 142–153. <https://doi.org/10.24203/ajhss.v8i4.6214>
- İşim, Ç., & Balcioglu, Y. (2023). *ChatGPT: Performance of translate* [Paper presentation]. Paper presented at the 3rd International ACHARAKA 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*. 1–10. [https://www.researchgate.net/publication/367359399\\_Is\\_ChateGPT\\_A\\_Good\\_Translator\\_A\\_Preliminary\\_Study](https://www.researchgate.net/publication/367359399_Is_ChateGPT_A_Good_Translator_A_Preliminary_Study)
- Jones, F. R. (2019). *Literary translation*. Routledge.
- 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>
- Khudoyorovich, K. K., Rasuljanovna, I. N., Khalmuratovna, R. Z., & Eshkobilovna, K. D. (2020). The issues of word choice in fiction translation. *International Journal of Psychosocial Rehabilitation*, 24(4), 6660–6668. <https://doi.org/10.37200/IJPR/V24I4/PR2020478>
- 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>
- Lahiani, R., Alfalasi, T., Alzaabi, A., & Alhassani, N. (2023). Translations of overtones in classical Arabic poetry: Optimality theory as a quality assessment framework. *Ampersand*, 11, 100143. <https://doi.org/10.1016/j.amper.2023.100143>
- Läubli, S., Sennrich, R., & Volk, M. (2018). Has machine translation achieved human parity? a case for document-level evaluation [Paper presentation]. Paper Presented at the Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium. <https://doi.org/10.18653/v1/D18-1512>
- Mahasneh, A. (2016). Translating Arabic political poetry into English. 831–809, (2)13, *مجلة اتحاد الجامعات العربية للآداب*. <https://doi.org/10.51405/0639-013-002-019>
- Mirzoyeva, L. (2023). Machine translation shortcomings and teaching translation. *Revista Romaneasca Pentru Educatie Multidimensionala*, 15(3), 232–242. <https://doi.org/10.18662/rrem/15.3/764>
- Mohammed, K. S. (2023). The impact of compensation in translating classical Arabic poetry into English: an analytical study. *Journal of Language Studies*, 1(4), 11–27. <https://doi.org/10.25130/jls.1.4.5>
- Mohsen, M. (2024). Artificial intelligence in academic translation: A comparative study of large language models and google translate. *PSYCHOLINGUISTICS*, 35(2), 134–156. <https://doi.org/10.31470/2309-1797-2024-35-2-134-156>
- Öner Bulut, S., & Alimen, N. (2023). Translator education as a collaborative quest for insights into the re-positioning of the human translator (educator) in the age of machine translation: the results of a learning experiment. *The Interpreter and Translator Trainer*, 17(3), 375–392. <https://doi.org/10.1080/1750399X.2023.2237837>

- Orabi, M., El Rifai, H., & Elnagar, A. (2020). *Classical arabic poetry: Classification based on Era* [Paper presentation]. Paper Presented at the 2020 IEEE/ACS 17th International Conference on Computer Systems and Applications (AICCSA), Antalya, Turkey. <https://doi.org/10.1109/AICCSA50499.2020.9316520>
- 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>
- Robinson, D. (2017). What kind of literature is a literary translation? *Target International Journal of Translation Studies*, 29(3), 440–463. <https://doi.org/10.1075/target.16064.rob>
- Saideen, B., Thalji, A., Lababaneh, A., Tartory, R., Haider, A. S., & Obeidat, M. (2024). Netflix English subtitling of the Jordanian movie 'The Alleys': Challenges and strategies. *Research Journal in Advanced Humanities*, 5(1), 157–177. <https://doi.org/10.58256/sptxps94>
- Tannous, B. (2024). *Assessing the Accuracy of AI Tools (Google Translate, ChatGPT, and Gemini) in Translating Arabic Idiomatic Expressions and Proverbs into English* [MA Thesis]. Applied Science Private University.
- 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>
- Weld-Ali, E. W., Obeidat, M. M., & Haider, A. S. (2023). Religious and cultural expressions in legal discourse: evidence from Interpreting Canadian courts hearings from Arabic into English. *International Journal for the Semiotics of Law - Revue Internationale de Sémiotique Juridique*, 36(6), 2283–2301. <https://doi.org/10.1007/s11196-023-10016-z>
- Yilmaz, E. D., Naumovska, I., & Aggarwal, V. A. (2023). AI-driven labor substitution: Evidence from google translate and ChatGPT. *INSEAD*. <https://sites.insead.edu/facultyresearch/research/doc.cfm?did=70468>