A Comparative Analysis of AI and Machine Translation versus Human Translation of English Literary Works

Amel BELDJENNA[1]

University of Oran 2 Mohamed Ben Ahmed –Algeria

beldjenna.amel@univ-oran2.dz

0009-0003-2652-0194

Abstract
The study at hand investigates the quality of translation produced by two machine translation systems, namely; Google Translate and ChatGPT-4, in comparison with human translation (HT) of a literary work from English into Arabic. The corpus of the study consists of the Arabic translation of The Forty Rules of Love by Elif Shafak. Five excerpts were purposefully selected from the novel to serve as the main dataset for analysis. These excerpts were chosen because they contain figurative expressions, culturally embedded idioms, metaphorical language, and emotional nuances that typically pose challenges for machine translation. To evaluate the translations, the study adopts the Multidimensional Quality Metrics (MQM) framework, which provides a systematic assessment of translation quality through dimensions such as accuracy, fluency, style, and cultural adequacy. The findings reveal clear distinctions between the three types of translations. Human translation demonstrates the highest level of accuracy and fluency, maintaining not only the semantic content of the source text (ST) but also its literary tone, stylistic richness, and cultural depth. ChatGPT-4, while producing coherent and fluent Arabic output, struggles with cultural sensitivity and with rendering figurative language in a contextually appropriate manner. It occasionally simplifies metaphors or reinterprets them in ways that weaken the stylistic effect intended by the author. Google Translate, on the other hand, performs the weakest among the three. It frequently misinterprets idiomatic expressions, distorts metaphorical meanings, and fails to convey the emotional and contextual layers of the ST. Overall, the study highlights the limitations of current machine translation tools when dealing with literary texts and emphasizes the irreplaceable role of human translators in preserving artistic and cultural integrity.

Keywords: AI translation; ChatGPT; Google Translation; Human Translation; Machine Translation.

[1]Corresponding author: Amel BELDJENNA/ beldjenna.amel@univ-oran2.dz

Introduction

Translation plays an influential role in our world today, with the world becoming more and more of small village connected via the internet. Translation is the transference of both written and spoken Source Language (SL) texts to equivalent written or spoken Target Language (TL) texts (Richards and Schmidth, 2013). Translation is required to establish communication between people from different background, countries and cultures. Translation plays a major role in inter-cultural communication in all fields of human knowledge (Farghal, 2012). With the advancement of technologies throughout history, new methods of translation has appeared. For instance, Machine Translation (MT) first emerged in 1954, when an IBM 701 computer or "electronic brain" automatically translated 60 sentences from Russian into English, proving that machine translation was possible. Notwithstanding the time-consuming research and considerable expenses involved, this paved the way for the future of language translation.

MT refers to "computerized systems responsible for the production of translations with or without human assistance" (Hutchins, 1995, p. 1). Machine Translation uses software to translate text or speech from one language to another without human intervention. Historically, MT has evolved through several stages, each marked by different approaches and technologies. Fast forward to the 20th century, AI proved a good tool for translation.

While AI translation represents a significant leap forward in the field of machine translation, it shares common ground with its predecessors. At its core, both AI translation and traditional machine translation methods aim to facilitate cross-linguistic communication by automatically converting text from one language to another. However, what distinguishes AI translation is its use of advanced neural networks and deep learning algorithms, enabling it to surpass the limitations of earlier approaches. It incorporates artificial intelligence, especially deep learning, and neural networks, to model the entire process of translation in a way that mimics human brain functions or so is claimed.

Since the Dartmouth Conference in 1956, AI has made significant advancements across various domains, evolving into a third generation characterized by new paradigms in learning approaches and technological applications. Given the intrinsic link between language and cognition, Natural Language Processing (NLP) is regarded as a core component of AI (Yusuf and Sivanadhan, 2025). Among the earliest and most prevalent applications of NLP, machine translation has remained a central topic of interest across disciplines such as linguistics, computer science, and cognitive science; driven by technological progress and increasing user demand. The interconnected relationship among AI, NLP, and machine translation is represented in Figure 1 below.

Figure 1. The interconnected relationship among AI, NLP, and machine translation
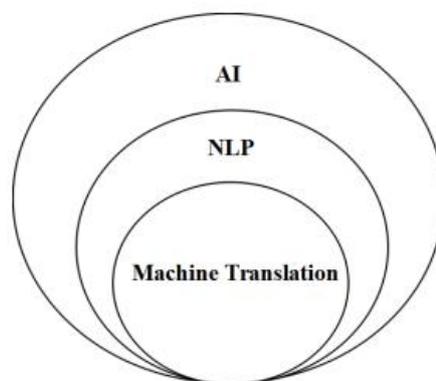
Source: Bhatt et al., 2021

In recent years, a variety of platforms have emerged to provide machine translation services, reflecting the rapid advancements in Natural Language Processing (NLP) and artificial intelligence technologies. Among the most widely used are Google Translate, DeepL, and ChatGPT, each offering different strengths in terms of accuracy, language coverage, and contextual understanding. Google

Translate is known for its broad language support and integration with other Google services, while DeepL has gained recognition for its high-quality translations, particularly for European languages. ChatGPT, on the other hand, leverages large language models to provide context-aware translations and explanations, making it a versatile tool for both casual users and language learners. The growing availability and sophistication of these platforms demonstrate the increasing role of AI-driven tools in breaking down language barriers and facilitating global communication.

Translating literary texts between Arabic and English presents significant challenges due to the substantial differences in their grammatical rules, sentence structures, and linguistic conventions. Arabic, with its rich morphological system, complex verb forms, and flexible word order, contrasts sharply with the relatively fixed syntax and simpler morphology of English. In fact, Arabic is an introflectional (also called non-concatenative) language (lugha ᶜishtiqaᶜia) (Beldjenna and Bekaddour, 2024). These structural disparities can make it difficult to preserve the original tone, style, and nuanced meanings when rendering literary works from one language to the other. Furthermore, cultural references, idiomatic expressions, and rhetorical devices often lack direct equivalents, requiring translators to balance fidelity to the source text with clarity and naturalness in the target language. As a result, literary translation between Arabic and English demands not only linguistic expertise but also a deep understanding of both cultures and literary traditions.

With the rapid development of Artificial Intelligence (AI) and Natural Language Processing (NLP), machine translation tools have become increasingly accessible and widely used across various fields. However, while platforms such as Google Translate, DeepL, and ChatGPT have demonstrated significant improvements in translation quality, challenges persist; especially when translating between linguistically and culturally distant languages such as Arabic and English. Literary texts, in particular, present unique difficulties due to their complex syntax, figurative language, and deep cultural references. Despite the growing popularity of AI-powered translation tools, limited research has focused specifically on their effectiveness in translating literary texts between Arabic and English. Therefore, the present study seeks to evaluate the accuracy, fluency, and cultural appropriateness of machine-generated translations in this context. By doing so, it aims to shed light on the current capabilities and limitations of AI in handling nuanced literary content, thus contributing to both translation studies and the ongoing development of NLP technologies. Therefore, the following research question is asked: **Can AI-Powered Machine Translation Match the Quality of Human Translation?**

1.   Literature Review

Yulianto and Supriatnaningsih, R. (2021) evaluate the performance of Google Translate (GT) and DeepL in terms of translation accuracy and readability. They focus on French-to-English translations of the play *En attendant Godot*, with both GT and DeepL outputs assessed against the original English version (EO) of the text as a reference. They employ two quantitative methods: a manual evaluation using the SAE J2450 translation quality metric, and an automated assessment using the Coh-Metrix tool.The manual assessment revealed that both GT and DeepL produced translations that met acceptable standards, scoring 84 and 99.04 respectively. According to the CdTrubric, translations scoring between 80 and 99 are considered good. Coh-Metrix results showed variation between GT and DeepL, but statistical analysis using ANOVA indicated no significant difference between either system and the original version. The mean scores were 99.69 for original version, 100.4 for GT, and 100.78 for DeepL.

Zhao, et al, (2023) study the English-to-Chinese translation made by ChatGPT. They found that ChatGPT generally performs on par with standard machine translation models, particularly within the general domain where the language is more conversational and lacks complex terminology. In such contexts, ChatGPT produces fluent, natural translations and exhibits the least signs of translationese among one general domain (WMT22 General MT Task) and two specific domains (WMT22 biomedical and Information and Communications Technology (ICT) domains). Its performance declines

significantly in specialized domains like biomedical and ICT, where terminology is highly domain-specific. In many instances, ChatGPT fails to accurately render these technical terms. This highlights the need for further investigation into whether its translation quality could be enhanced through the integration of external knowledge sources. At present, using ChatGPT for domain-specific translation tasks is not advisable.

Turchaninov and Stefurak (2024) scrutinize the effectiveness of AI-based machine translation tools in rendering French literary texts into English. Using a 'Reverse Turing Test' approach, translations of selected excerpts from the prominent French authors; Marcel Proust, Louis-Ferdinand Céline, and Albert Camus are evaluated. Their study compares outputs from ChatGPT-4o, ChatGPT-3.5, ChatGPT-4, Google Translate, and DeepL, focusing on accuracy, stylistic fidelity, and overall quality. The findings indicate that ChatGPT-4o delivered the most accurate and stylistically faithful translations, achieving an average rating of 9 out of 10, while also offering a faster translation process that required minimal post-editing. Moreover, they used a survey and they found out that 82.6% of participants were able to correctly identify machine-generated translations, whereas only 48.3% accurately recognized human translations. These results highlight the growing sophistication of AI translation tools and their potential for use in literary contexts.

Al Rousan, Jaradat and Malkawi (2025) investigate the translation proficiency of ChatGPT in comparison to Human Translation (HT) through the lens of Arabic literary work. They specifically explore potential translation deficiencies in ChatGPT and evaluate its capacity to serve as a substitute for human translators. The research focuses on 12 selected excerpts from *Mawsim Al-HijraIla Al-Shamal* (1966) by TayebSalih, comparing ChatGPT's English translations with the established version by Denys Johnson-Davies (*Season of Migration to the North*, 1969). A mixed-method approach, integrating both qualitative and quantitative analysis, was employed using three key dimensions of the Multidimensional Quality Metrics (MQM) framework: accuracy, fluency, and design. To ensure consistency and objectivity, the MQM scoring model was applied. The results indicate that while human translation demonstrates superior accuracy (94.5% versus ChatGPT's 77.9%), ChatGPT performs well in terms of fluency, scoring slightly higher than HT (97.2% vs. 96.6%). Nonetheless, ChatGPT exhibits notable shortcomings in design-related aspects and frequently introduces extraneous content. The study concludes that although ChatGPT offers fluent outputs, it lacks the reliability and cultural nuance required for the translation of Arabic literature, underscoring the continued necessity of skilled human translators like Denys Johnson-Davies. The researcher use MT as AI generated translation as one type of translation.

2. Research methods
2.1. Data collection tools

This study employs a qualitative approach. The researcher chose The Forty Rules of Love as the sample of study. It is written in English by the Turkish-British author, Elif Shafak and it consists of 229 pages. It was later translated into Turkish as *Aşk*. The book is also translated into Arabic by Khaled Al Jubilee and it consists of 512 pages named قواعد العشق الاربعون. *The Forty Rules of Love* by Elif Shafak is rich with poignant, spiritually charged, and emotionally resonant moments. The novel alternates between two timelines; a contemporary story about Ella Rubenstein and a historical tale about Rumi and Shams of Tabriz; and its most powerful moments usually come from the "rules" that express profound Sufi teachings about love, ego, patience, and the divine. It addresses a set of critical questions about identity, alterity, and the silent codes embedded in the narrative. Literary translation is widely regarded as the most complex and demanding form of translation due to its intricate interplay of language, culture, and artistic expression (Rankawat, 2017; Yousef, 2012). It has also been described as *"an art with many difficulties"* (Jun, 2019, p. 119), underscoring the nuanced challenges it presents beyond mere linguistic conversion.

Five excerpts were purposefully selected for their figurative language, idioms, cultural references, or emotional nuance to be studied in the three Arabic translations. Google translation an chatgpt translation are compared to the human translation made by Khaled Al Jubilee.

2.2.   Data analysis

The Multidimensional Quality Metrics (MQM) is a framework for analytic *Translation Quality Evaluation* (TQE). It can be applied to both human translation, machine translation and AI-generated translation. Noted for its flexibility, MQM can be tailored to suit the specific requirements of individual research studies (Fakih et al., 2024). Its error typology consists of eight main dimensions, only the following are studied in this research.

Table 1: Multidimensional Quality Metrics dimensions

Dimension		Purpose
1	Accuracy	It addresses errors where the translation does not accurately convey the intended meaning of the source text.
2	Fluency	It refers to issues in the linguistic structure of a text, independent of its status as a translation.
3	Terminology	It focuses on the incorrect use of domain-specific terms.
4	Locale Conventions	It concerns the failure to meet cultural or regional expectations of the target audience.
5	Style	It includes errors related to stylistic or register choices inappropriate for the text type
6	Verity	Verity refers to extra-linguistic truth correspondence in texts.

The MQM framework is employed in this study due to its hierarchical and systematic classification of error types, which makes it particularly suitable for evaluating literary translations. Additionally, the framework's issue typology was developed using a comprehensive set of MT quality measures, aligning well with the objectives of this study (Lommel et al., 2013).

3.   Findings and discussion

An essential aspect to maintain in any translation is cohesion. This concept refers to "the ways in which the components of the textual world, i.e., the configuration of concepts and relations which underlie the surface text, are mutually accessible and relevant" (Beaugrande and Dressler, 1981, p. 3). Cohesion plays a crucial role in ensuring that the translated text reads as naturally and fluently as the original. In our context, accuracy is similar to cohesion. The first part of analysis is to ensure the message is conveyed accurately from the ST to the TT.

In this section we will discuss each excerpt from the English version and it counterpart in the Arabic translation and compare it to the translation made by both Google translate and CHATGPT-4. Each excerpt is analyzed for six dimensions.

3.1.   Excerpt one- First Rule of Love

How we see God is a direct reflection of how we see ourselves. If God brings to mind mostly fear and blame, it means there is too much fear and blame welled inside us. If we see God as full of love and compassion, so are we. Page 28

"إن الطريقة التي نرى فيها الله ما هي إلا انعكاس للطريقة التي نرى فيها أنفسنا. فإذا لم يكن الله يجلب إلى عقولنا سوى الخوف والملامة، فهذا يعني أن قدرًا كبيرًا من الخوف والملامة يتدفّق في نفوسنا. أما إذا رأينا الله مفعمًا بالمحبة والرحمة، فإننا نكون كذلك" ص48

Google translate	CHATGPT-4
كيف نرى الله هو انعكاس مباشر لكيفية رؤيتنا لنفسك. إذا كان الله كثر في أذهاننا الخوف واللوم في الغالب، فهذا يعني أن هناك الكثير من الخوف واللوم يتراكمون بداخلنا. وإذا رأيناه مليئا بالحب والرحمة، فنحن كذلك.	كيف نرى الله هو انعكاس مباشر لكيفية رؤيتنا لأنفسنا. إذا كان الله يثير في أذهاننا الخوف واللوم في الغالب، فهذا يعني أن هناك الكثير من الخوف واللوم متراكمين بداخلنا. وإذا رأيناه مليئًا بالحب والرحمة، فنحن كذلك.

In assessing the accuracy of the three translations using the MQM framework, the HT by Khaled Al Jubilee demonstrates a high level of fidelity to the ST. It effectively conveys the intended meaning with fluent and idiomatic Arabic, with no major errors identified, making it the most accurate translation. The GPT-4 translation also seems accurate, maintaining the overall meaning and coherence of the ST. However, it contains a few minor issues, such as word choices like متراكمين, while this word is grammatically acceptable, it is more loaded a word than intended by Elif. Nonetheless, these do not significantly distort the meaning. In comparison to the Google Translate version, there appears notable inaccuracies. It includes grammatical and lexical errors such as رؤيتنا لنفسك (our view of *yourself*) instead of the correct reflexive pronoun, and the awkward use of كثر, which distorts the intended meaning. These issues result in major to critical errors that compromise the translation's accuracy and reliability.

When it comes to fluency, the HT by Khaled Al Jubilee demonstrates superior linguistic quality. It consistently uses grammatically correct structures, appropriate register, and natural phrasing that aligns well with native Arabic norms. Spelling, punctuation, and typography are accurately handled, contributing to smooth readability. The CHATGPT-4 translation, on the other hand, performs well in terms of fluency but occasionally exhibits less natural phrasing. For instance, some sentence such as يثير في أذهاننا الخوف واللوم في الغالب, could be improved for a more native-like one. However, this is a minor error and does not hinder comprehension. In contrast, the Google Translate output shows multiple fluency-related problems. It contains frequent grammatical errors, awkward phrasing, inconsistent register, and occasional spelling or typographic mistakes. These factors disrupt the reading experience and make the text feel less polished and less natural to native speakers. Consequently, Google Translate's fluency errors are more severe, often classified as major or critical under the MQM framework.

In terms of terminology, the HT exhibits precise and contextually appropriate use of terms, reflecting a deep understanding of the source material and its cultural nuances. The specialized vocabulary and key concepts are accurately rendered, ensuring clarity and consistency throughout the text. The GPT-4 translation generally handles terminology well, correctly translating most key terms and maintaining coherence. However, it occasionally chooses less precise or slightly off context words, which may reduce terminological accuracy in subtle ways but typically do not obscure meaning. On the contrary, the Google Translate version struggles with consistent and accurate terminology. It often mistranslates or substitutes terms with incorrect or vague alternatives, leading to confusion or distortion of the ST Such as يتراكمون بداخلنا (incorrect plural form), when it should be متراكمينor متراكم. These terminological inaccuracies significantly affect the quality of the translation, reflecting major to critical errors under the MQM framework.

In locale conventions, the HT clearly stands out for its natural flow, rhetorical richness, and adherence to the stylistic norms of Arabic. It employs culturally appropriate expressions, maintains consistent

voice and person, and captures the emotional depth of the original text through idiomatic phrasing such as يتدفّق في نفوسنا. ChatGPT-4, while grammatically accurate and semantically clear, lacks the stylistic elegance and cultural resonance expected in Arabic literary writing. Its expressions are more literal and less expressive, making the translation feel somewhat mechanical. Google Translate shows significant shortcomings in that it contains grammatical errors, misuses pronouns like رؤيتنا لنفسك, and includes non-idiomatic phrases that break the natural rhythm of Arabic, indicating a lack of awareness of both linguistic and cultural conventions.

When it comes to style, HT once again excels, capturing the philosophical tone, emotional depth, and rhythmic elegance of the original English passage. It uses stylistically rich, literary Arabic with varied sentence structure, appropriate emphasis, and emotive vocabulary, such as يتدفّق في and ما هي إلا انعكاس نفوسنا, which add both depth and flow to the text. This aligns with traditional Arabic stylistic norms. ChatGPT-4's translation, while stylistically cleaner than Google Translate's, leans toward a neutral and informative tone, more functional than artistic. The vocabulary is accurate, and sentence structure is clear, but it lacks the figurative richness and rhetorical devices found in high-quality Arabic prose. The result is a readable but relatively flat text that does not fully reflect the reflective and expressive style of the ST. Google Translate's TT fares the worst stylistically. It reads like a literal, machine-generated rendering, with awkward phrases كثر في أذهاننا and grammatical errors that disrupt the flow and coherence. The translation lacks both fluency and stylistic sensitivity, making it unsuitable for literary or emotionally nuanced texts.

3.2.   Excerpt two

You can study God through everything and everyone in the universe, because God is not confined in a mosque, synagogue, or church. But if you are still in need of knowing where exactly His abode is, there is only one place to look for Him: in the heart of a true lover. Page 43

تقول إحدى القواعد إنه يمكنك أن تدرس الله من خلال كل شيء وكل شخص في هذا الكون، لأن وجود الله لا ينحصر في المسجد، أو في الكنيسة أو في الكنيس. لكنك إذا كنت لا تزال تريد أن تعرف أين يقع عرشه بالتحديد يوجد مكان واحد فقط تستطيع أن تبحث فيه عنه، وهو قلب عاشق حقيقي. فلم يعش أحد بعد رؤيته، ولم يمت أحد بعد رؤيته . فمن يجده يبقى معه إلى الأبد . ص 89

Google translate	CHATGPT-4
يمكنك دراسة الله من خلال كل شيء وكل شخص في الكون، لأن الله ليس محصورًا في مسجد أو معبد أو كنيسة. ولكن إن كنت لا تزال بحاجة إلى معرفة مكانه بالضبط، فليس هناك سوى مكان واحد للبحث عنه: في قلب عاشق حقيقي.	يمكنك أن تدرس الله من خلال كل شيء وكل شخص في هذا الكون، لأن الله لا يُحصر في مسجد أو كنيس أو كنيسة. ولكن إن كنت لا تزال بحاجة إلى معرفة مكان سكنه بالضبط، فليس هناك سوى مكان واحد لتبحث فيه عنه: في قلب المُحب الصادق

The HT of the passage demonstrates a high level of accuracy, preserving not just the literal meaning but also the deeper spiritual connotations of the ST. It captures the intended message that 'God transcends religious buildings and can be found in the heart of a true lover'. Particularly, the phrase أين يقع عرشه بالتحديد (where exactly His throne is) adds symbolic depth that resonates with Arabic religious and mystical traditions, enriching the original meaning. The fluency of the translation is excellent, with smooth, cohesive sentence structures and natural transitions. It reads like an authentic piece of literary Arabic, using idiomatic expressions and elegant phrasing that reflect both the content and the tone of the source. The style is deeply spiritual, emotive, and aligned with Arabic literary and Sufi traditions, making it highly effective in conveying the essence of the original.

In contrast, the ChatGPT-4 translation is accurate and grammatically correct but leans toward a more literal and modern rendering. It maintains the core message and does not distort meaning, yet it lacks the symbolic and emotional richness found in the HT. The phrase مكان سكنه is a technically correct

equivalent of 'abode,' but it is less poetic and spiritually resonant than عرشه. While the translation is fluent and follows Arabic conventions, the style remains neutral and lacks the rhetorical flair expected in literary or spiritual writing. Thus, although ChatGPT-4 provides a clear and accessible translation, it does not evoke the same emotional or cultural impact.

Google Translate, on the other hand, performs noticeably weaker across several dimensions. While it conveys the general meaning, it introduces inaccuracies such as using معبد (temple) instead of كنيسة (synagogue), thereby changing the religious reference and affecting the cultural specificity of the translation. The fluency is passable but often reads as mechanical, with phrases like مكانه بالضبط sounding awkward and unnatural in Arabic. The translation also lacks the cultural and figurative sensitivity necessary for such a spiritually nuanced text. Stylistically, it is flat and utilitarian, offering no poetic or rhetorical embellishment. Therefore, while useful for basic understanding, Google Translate falls short in delivering a translation that is culturally, stylistically, and spiritually faithful to the original.

3.3. Excerpt three

Try not to resist the changes that come your way. Instead, let life live through you. And do not worry that your life is turning upside down. How do you know that the side you are used to is better than the one to come? page 68

لا تحاول أن تقاوم التغييرات التي تعترض سبيلك، بل دع الحياة تعيش فيك. ولا تقلق إذا قلبت حياتك رأساً على عقب. فكيف يمكنك أن تعرف أن الجانب الذي اعتدت عليه أفضل من الجانب الذي سيأتي؟ ص 148

Google translate	CHATGPT-4
حاول ألا تقاوم التغييرات التي تعترض طريقك. بل دع الحياة تعيش فيك. ولا تقلق من أن حياتك قد تنقلب رأسًا على عقب. كيف تعرف أن الجانب الذي اعتدت عليه أفضل من الجانب الذي سيأتي؟	حاول ألا تقاوم التغييرات التي تأتي في طريقك. بدلاً من ذلك، دع الحياة تعبر من خلالك. ولا تقلق إذا شعرت أن حياتك تنقلب رأساً على عقب، فكيف لك أن تعرف أن الجانب الذي اعتدت عليه أفضل من الجانب القادم؟

Accuracy in this verse's HT is highly faithful to the ST. It preserves the intended message and metaphoric structure of the original without distortion or omission. CHATGPT-4 also delivers a mostly accurate translation, but it introduces a slight semantic deviation with the phrase دع الحياة تعبر من خلالك (let life pass through you) rather than "let life live through you," which slightly alters the depth and poetic nature of the ST metaphor. Google Translate, while understandable, suffers from minor phrasing inconsistencies; such as لا تقلق من أن حياتك قد تنقلب (do not worry that your life may turn upside down) which, although grammatically correct, sound unnatural and detract from the ST's tone.

In evaluating fluency, Al Jubilee's translation exhibits smooth, natural Arabic with fluid syntax and appropriate rhythm. His version reads like authentic literary Arabic. GPT-4 produces a fluent and coherent translation, though its style is a bit more mechanical and less elegant, and some phrases such as تعبر من خلالك, feel less idiomatic. Google Translate lags in fluency, with a flatter sentence structure and minor punctuation or phrasing that breaks the flow of the text, such as a disjointed tone in the final rhetorical question.

From a style and register perspective, Al Jubilee's version most accurately reflects the contemplative, philosophical tone of the original English. The language is formal, expressive, and culturally aligned with Arabic literary norms. GPT-4 attempts to match this tone and mostly succeeds, though some stylistic choices lack the poetic richness and emotional resonance found in the human translation. Google Translate fails to capture the literary and reflective style; its tone remains neutral and mechanical, often reading like a direct rendering rather than a thoughtful adaptation.

Finally, regarding locale conventions, Khaled Al Jubilee's translation adheres strictly to Arabic norms, including punctuation, text direction, and idiomatic usage. GPT-4 generally respects these conventions as well, with minor deviations. Google Translate, however, sometimes misplaces punctuation and tends to use non-idiomatic expressions, indicating weaker alignment with Arabic writing norms.

3.4.   Excerpt four

When the ney abated, my father appeared onstage. With measured, soft steps, he approached and greeted the audience. Six dervishes followed him, all my father's disciples, all wearing long white garments with large skirts. They crossed their hands on their chests, bowing in front of my father to get his blessing. Then the music started, and, one by one, the dervishes began to spin, first slowly, then with breathtaking speed, their skirts opening up like lotus flowers. Page 174

عندما خفت صوت الناي، تقدم أبي إلى المنصة، وبخطوات مدروسة، متأنية، رقيقة، اقترب وحيا الحاضرين. ثم تبعه ستة دراويش، كلهم من مريدي أبي وكانوا يرتدون أثواباً بيضاً طويلة واسعة من الأسفل، وقد عقدوا أيديهم على صدورهم، ثم انحنوا أمام أبي لنيل بركته؛ ثمّ عزفت الموسيقى، وبدأ الدراويش، الواحد تلو الآخر، يدورون حول أنفسهم، ببطء في البدء، ثم بسرعة أكبر، وتفتحت أذيال أثوابهم مثل أزهار اللوتس. ص 388

Google translate	CHATGPT-4
عندما أُطلقت النار على الناي، صعد والدي على المسرح. بخطوات هادئة وواثقة، اقترب من الجمهور ورحّب بهم. تبعه ستة دراويش، جميعهم من تلاميذ والدي، يرتدون جميعًا ثيابًا بيضاء طويلة وتنانير واسعة. شبكوا أيديهم على صدورهم، وانحنوا أمام والدي طلبًا لبركاته. ثم بدأت الموسيقى، وبدأ الدراويش بالدوران، واحدًا تلو الآخر، ببطء في البداية، ثم بسرعة مذهلة، وتنفتح تنانيرهم كزهور اللوتس.	عندما خفت صوت الناي، ظهر والدي على خشبة المسرح. بخطى هادئة ومتزنة، اقترب من الجمهور وحيّاهم. تبعه ستة دراويش، جميعهم من تلاميذ والدي، وكانوا يرتدون ثيابًا بيضاء طويلة وتنانير واسعة. شبكوا أيديهم على صدورهم، وانحنوا أمام والدي طلبًا لبركته. ثم بدأت الموسيقى، وبدأ الدراويش بالدوران، واحدًا تلو الآخر، ببطء في البداية، ثم بسرعة تخطف الأنفاس، وتفتّحت تنانيرهم مثل زهور اللوتس.

The ChatGPT-4 translation appears to be accurate and stylistically elegant, preserving key details such as the disciples' reverence and the dervishes' spinning. Minor nuances like omitting one of the adjectives describing the father's steps, are the only slight errors from the ST. In contrast, the Google Translate version contains serious flaws, most notably mistranslating عندما خفت صوت الناي as 'when the ney was shot,' which distorts the meaning entirely. It also inconsistently shifts tenses and lacks the poetic resonance of the source. The HT aligns very closely with the English, suggesting the English was translated with care, while the ChatGPT version offers a strong back-translation, and Google Translate fails to capture the passage's nuance and tone.

The HT is naturally fluent, rich in literary tone, and stylistically polished. It flows with rhythmic elegance, using phrases like تفتحت أذيال أثوابهم مثل أزهار اللوتس and بخطوات مدروسة، متأنية، رقيقة that reflect traditional Arabic narrative aesthetics. Its sentence structure is cohesive, with precise use of conjunctions and transitions. This fluency is not only grammatical but also poetic, fitting for the spiritual and performative scene it describes.

The ChatGPT-4 translation is quite fluent and natural, however; it is slightly more neutral in tone compared to the ST, it lacks the human touch. It nonetheless, maintains smooth sentence flow and appropriate word choices. Phrases such as بسرعة تخطف الأنفاس and بخطى هادئة ومتزنة demonstrate a good command of idiomatic Arabic. Although it simplifies some descriptions (e.g., omitting one adjective or using ظهر instead of تقدّم), the overall narrative remains cohesive and elegant. Google Translate, on the other hand, suffers from noticeable fluency issues. The glaring mistranslation in the opening عندما أُطلقت النار على الناي introduces confusion immediately, and other phrases feel mechanical or awkward, such as تنفتح تنانيرهم in present tense within a past-tense narrative.

The HT follows culturally appropriate Arabic literary conventions. It employs a rhythmic, descriptive style typical of Arabic narrative prose, especially in spiritual or ceremonial contexts. The use of triplet

adjectives رقيقة، متأنية، مدروسةreflects a stylistic norm in Arabic for emphasis and rhythm. The phrasing is rich and lyrical, echoing Sufi and mystical themes, which resonate well with Arab readers familiar with such imagery. Its tone is respectful, poetic, and deeply rooted in Arabic literary tradition, with metaphorical language مثل أزهار اللوتس and reverent structure fitting the cultural context of a whirling dervish performance. ChatGPT-4 translation largely respects Arabic stylistic and locale norms. It uses appropriate vocabulary and maintains a neutral to slightly elevated literary register. While it streamlines some of the original's embellishments such as reducing three adjectives to two, it avoids awkward literalism and chooses phrasing that sounds natural to native speakers. Cultural conventions such as reverence toward the father figure and the dervishes, are handled with care, using expressions like بسرعة تخطف الأنفاس and طلبًا لبركته, which are idiomatically and culturally appropriate. While slightly more modern in tone than the ST, it aligns well with contemporary Arabic literary translations.

Google Translate, however, diverges significantly from Arabic stylistic and cultural norms. Its phrase choices are often literal or mistranslated, leading to unnatural expressions like عندما أُطلقت النار على الناي, which, not only distorts the meaning, but also ignores the poetic and respectful tone of the original. Moreover; it uses overly modern or robotic structures and sometimes mismatches verb tenses like mixing past and present, disrupting the expected narrative flow. Additionally, certain expressions like طلبًا لبركته(plural) subtly miss the tone of humility and singular reverence found in the HT.

3.5. Excerpt Five – rule number 41

A life without love is of no account. Don't ask yourself what kind of love you should seek—spiritual or material, divine or mundane, Eastern or Western. Divisions only lead to more divisions. Love has no labels, no definitions. It is what it is, pure and simple. Page 225

لا قيمة للحياة من دون عشق. لا تسأل نفسك ما نوع العشق الذي تريده، روحي أم مادي، إلهي أم دنيوي غربي أم شرقي ... فالانقسامات لا تؤدي إلا إلى مزيد من الانقسامات. ليس للعشق تسميات ولا علامات ولا تعاريف. إنه كما هو، نقي وبسيط. ص 500

Google translate	CHATGPT-4
الحياة بلا حب لا قيمة لها. لا تسأل نفسك أي نوع من الحب تبحث عنه - روحي أم مادي، إلهي أم دنيوي، شرقي أم غربي. الانقسامات لا تؤدي إلا إلى المزيد من الانقسامات. الحب لا مسميات له ولا تعريفات. هو كما هو، خالص وبسيط.	الحياة بلا حب لا قيمة لها. لا تسأل نفسك عن نوع الحب الذي يجب أن تبحث عنه ـــ هل هو روحي أم مادي، إلهي أم دنيوي، شرقي أم غربي. الانقسامات لا تؤدي إلا إلى مزيد من الانقسامات. الحب لا يحمل تسميات ولا تعريفات. هو كما هو، نقي وبسيط

In terms of accuracy, the HT is the most faithful to the ST, capturing not only the literal meaning but also the emotional and spiritual depth conveyed in phrases like "a life without love is of no account" and using the word عشق to emphasize a profound, possibly mystical form of love. The ChatGPT-4 translation is also highly accurate, maintaining the logical structure and preserving the contrasts (spiritual vs. material, divine vs. mundane, and so on), although it goes for the more general term حب, which slightly softens the intensity of the message. Stylistically, it remains clear and elegant. The Google Translate translation, although generally accurate in meaning, lacks the idiomatic sophistication of the other two. Phrases like ولا مسميات له and خالص وبسيطare technically correct but feel less natural or poetic in Arabic, making the overall tone flat.

In terms of fluency, the HT stands out as the most natural. It flows smoothly, using refined, rhythmic phrasing and literary devices characteristic of high-quality Arabic prose. Expressions such as لا تسأل نفسك ما نوع العشق الذي تريده and إنه كما هو، نقي وبسيطfeel organic and stylistically coherent, matching the contemplative, philosophical tone of the passage. ChatGPT-4 translation is similarly fluent, with grammatically correct and idiomatic structures. Sentences like الحب لا يحمل تسميات ولا تعريفات and الانقسامات reflect a strong command of MSA. While slightly more straightforward than لا تؤدي إلا إلى مزيد من الانقسامات

the ST, it remains elegant and easy to read. On the contrary, the Google Translate translation which is noticeably less fluent. Though largely comprehensible, some phrasing feels

awkward or mechanical; for example, الحب لا مسميات له ولا تعريفات/sounds less natural than alternatives, and خالص وبسيط is less stylistically suitable than نقي وبسيط.

In terms of style, locale conventions, and register, the HT demonstrates the highest level of literary quality and cultural authenticity. Its rhythm and rhetorical structure, specifically in phrases like لا تسميات ولا علامات ولا تعاريف, reflect traditional Arabic eloquence and are well-suited to the philosophical nature of the content. The ChatGPT-4 translation, while slightly modernized and less rich, maintains fluency and respect for Arabic stylistic norms. It uses clear, natural expressions and proper sentence structure, though the choice of the more general حب softens the intensity of the original message. Nevertheless, it remains culturally appropriate and accessible. The Google Translate version, though accurate in basic meaning, lacks stylistic polish and often fails to adhere to Arabic literary conventions. Its phrasing is more mechanical, and word choices like خالص instead of نقي feel less idiomatic.

4.   Discussion

In order to answer the research question on whether can AI-Powered Machine Translation Match the Quality of Human Translation, the translation of the literary work of the forty rules of love by Elif Shafak was analysed using the Multidimensional Quality Metrics (MQM) as a framework. Five excerpts from the book are translated by Google Translate and CHATGPT -4 and then compared to the human translation of Khaled Al Jubillee's قواعد العشق الأربعون. The study revealed that while CHATGPT -4 uses clear, natural expressions and proper sentence structure, it chooses general language which softens the intensity of the ST. therefore, failing to mimic the poetic nature of Arabic. It performs well in terms of fluency but occasionally exhibits less natural phrasing. Google translate on the other hand, contains severe flaws, most notably mistranslating the ST. furthermore, Google Translate shows multiple fluency-related problems. It contains frequent grammatical errors, awkward phrasing, inconsistent register, and occasional spelling or typographic mistakes.

In fact, across all four excerpts, the HT consistently showed the greatest accuracy, not only in rendering literal meanings but also in preserving the deeper philosophical, emotional, and spiritual layers of the ST. Al Jubilee's word choices reflect a deep awareness of both the surface meaning and its culturally embedded resonances. These choices demonstrate a deliberate interpretive effort characteristic of human literary translation.

On the other hand, ChatGPT-4 was largely accurate in conveying core meanings, with only minor deviations or slightly less expressive equivalents. Its translations remain semantically faithful, but they often fall short of capturing metaphorical depth or symbolic richness. This indicates strong structural comprehension but more limited sensitivity to literary nuance. Google Translate showed acceptable accuracy for general meaning but produces frequent errors in pronouns, verb forms, and religious terminology. In several cases, GT introduces severe distortions. These critical errors suggest that while MT can grasp literal lexical patterns, it struggles with context, figurative meanings, and narrative coherence.

Moreover, the fluency of the HT is consistently superior. Its phrasing mirrors the stylistic expectations of educated Arabic readers, maintaining smooth syntactic flow, coherent rhythm, and appropriate punctuation. The HT's sentences unfold with a natural cadence that enhances readability and supports the reflective tone of the ST.

The findings also reveal that AI technology struggles to reliably translate cultural terms, often resulting in significant alterations to the original meaning. Furthermore, this AI model tends to translate certain words literally, disregarding the broader context and the crucial role of diacritical marks in shaping meaning. Such shortcomings highlight the challenges AI faces in capturing the cultural and linguistic intricacies that are essential for faithful and nuanced translations.

The findings of the study align with Al Rousan, Jaradat and Malkawi (2025) in the sense that both find that the HT is more reliable for translating Arabic literature than AI technology for different reasons. They argue that AI tool committed more accuracy errors than the HT, failing to generate an accurate translation. Similarly, Zhao et al, (2023) highlight the need for further investigation into whether CHATGPT translation quality could be enhanced through the integration of external knowledge sources since, using CHATGPT for domain-specific translation tasks is not advisable.

5.   Conclusion

It can be concluded from this research that the comparison between Google Translate, ChatGPT-4, and Khaled Al-Jubillee's HT demonstrates that current AI systems still fall short of achieving the linguistic, cultural, and aesthetic depth of the human brain required in Arabic literary translation. While ChatGPT-4 showed comparatively stronger performance in; producing clear, coherent sentences and maintaining a high level of fluency, it often relied on general or softened language that diluted the stylistic intensity of the source text. Its inability to consistently reproduce the poetic and emotive quality of Arabic literature reflects a broader limitation in capturing literary nuance. Google Translate, by distinction, exhibited more serious issues, including mistranslations, grammatical errors, inconsistent register, and fluency problems, making its output unreliable for literary contexts.

In general, the findings reinforce that while AI-powered translation tools continue to advance and offer useful support, they are not yet capable of matching the cultural sensitivity, contextual ,understanding, and stylistic precision achieved by skilled human translators, especially in the domain of literary translation.

This study offers several important implications and contributes to ongoing discussions in translation studies, machine translation (MT), and the emerging role of AI-powered language models. By systematically applying the MQM framework to compare Human Translation (HT), ChatGPT-4, and Google Translate across multiple literary excerpts, the study provides meaningful insights into the strengths, limitations, and future potential of automated translation systems in handling literary texts.

References

Al Rousan, R., Jaradat, R., & Malkawi, M. (2025). ChatGPT translation vs. human translation: an examination of a literary text. *Media & Communication Studies*, *11*(1). https://doi.org/https://doi.org/10.1080/23311886.2025.2472916

Bhatt, C., Kumar, I., Vijayakumar, V., Singh, K. U., & Kumar, A. (2021). The state of the art of deep learning models in medical science and their challenges. *Multimedia Systems*, 27(4), 599–613. https://doi.org/DOI:10.1007/s00530-020-00694

Beldjenna, A. and Bekaddour, S. (2024). Formation and Semantic Behavior of Verbs Of Color In Algerian Arabic. ALTRALANG Journal 6 (2), 272-284.

Fakih, A., Ghassemiazghandi, M., Fakih, A.-H. A., & Singh, M. K. M. (2024). Evaluation of Instagram's Neural Machine Translation for Literary Texts: An MQM-Based Analysis. *GEMA Online Journal of Language Studies* , *24*(1), 213–233. https://doi.org/DOI:10.17576/gema-2024-2401-13

Farghal, M. (2012). *Advanced Issues in Arabic English Translation Studies*. Kuwait University Press.

Hutchins, J. (1995). MACHINE TRANSLATION: A BRIEF HISTORY. In *Concise history of the language sciences: from the Sumerians to the cognitivists*. Oxford: Pergamon Press.

Jun, L. (2019). *Literary translation and its challenges: Between art and accuracy*. Beijing: Academic Publishing House.

Rankawat, M. (2017). Literary translation: The most challenging form of translation. *Journal of Literature and Translation Studies*, 5(2), 112–122.

Richards, J. C. and Schmidt, R. W. (2013). *Longman Dictionary of Language Teaching and Applied Linguistics*. Routledge. https://doi.org/10.4324/9781315833835

Shafak, E. (2010). *The forty rules of love*. New York: Viking.

Shafak, E. (2011). قواعد العشق الأربعون (K. Al Jubilee, Trans.). Beirut: Dar Al-Adab. (Original work published 2010)

Turchaninov, R., & Stefurak , O. (2024). AI-based machine translation vs. human translation of French literary works. Presented at the Scientific Horizon in the Context of Social CrisesAt, Tokyo, Japan.

Yousef, T. (2012). Literary Translation: Old and New Challenges. International Journal of Arabic-English Studies, 13(1), 49–64. https://doi.org/DOI:10.33806/ijaes2000.13.1.4

Yulianto, A., & Supriatnaningsih, R. (2021). Google Translate vs. DeepL: A quantitative evaluation of close-language pair translation (French to English). *AJELP The Asian Journal of English Language and Pedagogy*, *9*(2), 109–127. https://doi.org/DOI:10.37134/ajelp.vol9.2.9.2021

Yusuf, R. & Sivanadhan, I. (2025). The Role of Artificial Intelligence in Enhancing Language Translation and Cross Cultural Communication. MJSSH, 9(4), 245–256. https://doi.org/10.33306/mjssh/379

Zhao, Y., Zhang, M., Chen, X., Deng, Y., Geng, A., Liu, L., Zhang, Z. (2023). Human evaluation for translation quality of ChatGPT: A preliminary study. 282–287. https://doi.org/10.26615/issn.2683-0078.2023_023