Discover Artificial Intelligence

Research

Utilizing ChatGPT for subtitling movies: a comparative analysis of accuracy and fluency with human translators

Noureldin Abdelaal[1]

Received: 11 December 2024 / Accepted: 13 May 2025
Published online: 06 October 2025

Abstract
This study conducts a comprehensive examination of English-to-Arabic translations, comparing outputs from human translators with those from ChatGPT 3.5. A corpus was compiled using the "Our Cartoon President" series and its Arabic translations, aligning source texts with target translations. Thirty-two randomly selected examples underwent comparison between ChatGPT 3.5 outputs and human translations sourced from the Subscene website, where the latter were crafted by professional translators. To assess the translations, a survey was administered to English-Arabic translation professors, who rated fluency and accuracy on a 1 to 5 scale. Collected data underwent SPSS analysis, including a T-test to determine statistically significant differences between human and ChatGPT translations. Employing a mixed-methods approach, qualitative insights were drawn from professors' evaluations, while quantitative data stemmed from their ratings. Results indicate that human translation slightly outperformed ChatGPT translation, as demonstrated by both qualitative and quantitative analyses. Notably, the study underscores the continued necessity of post-editing, even with the advent of AI translation tools like ChatGPT, to ensure the highest quality and accuracy in translated content. This study sheds light on the comparative effectiveness of human and ChatGPT translations in English-to-Arabic contexts, emphasizing ChatGPT's potential utility as a subtitling tool to expedite the translation process.

Keywords Subtitling · ChatGPT · Human translation · Accuracy of translation · Fluency of translation

1 Introduction

In the dynamic landscape of translation studies, the advent of artificial intelligence (AI) has introduced new possibilities and challenges. Machine translation systems, powered by sophisticated algorithms and vast language datasets, have revolutionized the translation process, offering rapid and accessible translations across diverse languages. Among these systems, ChatGPT 3.5, developed by OpenAI, stands out as a cutting-edge language model renowned for its advanced natural language processing capabilities. However, despite the rapid advancements in AI-driven translation technologies, questions persist regarding their comparative effectiveness with human translators, particularly in nuanced contexts such as cultural references and idiomatic expressions. While machine translation systems excel in processing large volumes of text and generating translations with remarkable speed, they often struggle to capture the subtleties of human language, including cultural nuances and contextual meanings.

In the realm of English-to-Arabic translation, this debate takes on added significance, given the complexities of translating between languages with distinct linguistic structures and cultural contexts. Arabic, with its rich literary heritage and

✉ Noureldin Abdelaal, nourabdelal@yahoo.com | [1]October University for Modern Sciences and Arts, Faculty of Languages 6 October City, Giza, Egypt.

diverse dialects, poses unique challenges for translation, requiring a deep understanding of both language and culture to convey meaning accurately and fluently. Every day, we encounter a multitude of new movies spanning from American English to Arabic, necessitating translation between the two languages. The question arises as to whether AI, particularly ChatGPT, can effectively aid in this process and enhance subtitling quality. This study delves into the potential of ChatGPT as a valuable tool in subtitling, considering the complexities inherent in translating subtitles, which often involve cultural nuances, colloquialisms, and slang language. Against this backdrop, this study seeks to provide a comprehensive examination of English-to-Arabic translations, employing a comparative analysis between human translators and ChatGPT 3.5. The focus is on evaluating the fluency and accuracy of translations across both human and AI-generated outputs, using a corpus comprising the "Our Cartoon President" series and its Arabic translations.

By leveraging both qualitative and quantitative methodologies, this research aims to offer insights into the strengths and limitations of human and machine-generated translations in the context of English-to-Arabic translation. Through an analysis of translation quality, informed by the expertise of translation professors, and statistical comparisons using tools such as SPSS, the study seeks to contribute to ongoing discussions surrounding the role of AI in translation and potential avenues for improvement in machine translation systems.

Furthermore, by uncovering specific challenges faced by ChatGPT 3.5, such as the translation of cultural references, the study provides valuable considerations for the development of future AI-driven language models. Ultimately, by exploring the complex interplay between language, culture, and technology, this research aims to advance our understanding of the evolving landscape of translation studies in the era of artificial intelligence.

1.1 Literature review

Subtitles, a fundamental aspect of audiovisual translation (AVT), condense the original dialogue into written form, typically displayed at the bottom of the screen and synchronized with the corresponding segments of the conversation. According to Diaz-Cintas (2019), AVT encompasses the localization of audiovisual content through diverse translation techniques, considering both verbal and nonverbal elements conveyed through acoustic and visual channels.

Diaz-Cintas and Remael [1] categorize subtitles into three primary types based on linguistic distinctions: intralingual, interlingual, and bilingual. Intralingual subtitles involve the transition from oral to written communication within the same language, while interlingual subtitles entail translation between languages. Bilingual subtitles, displaying translations in two languages simultaneously, are commonly employed in bilingual regions or international film festivals to broaden viewership.

Subtitling presents numerous challenges, including linguistic and cultural aspects. Cultural disparities between source and target texts necessitate strategies such as foreignization and domestication to bridge gaps in cultural understanding. Additionally, linguistic challenges encompass lexical choice, syntax, pronunciation, and accent variations, all of which require careful consideration during the subtitling process. Formal constraints, technical issues, and the need for intentional omissions further complicate the subtitling process, demanding meticulous attention to detail to ensure coherence and clarity for viewers.

Moreover, humor shows and culturally specific expressions present unique challenges, requiring subtitlers to navigate cultural nuances effectively. Issues such as cursing and forbidden phrases necessitate decisions regarding inclusion or omission based on cultural acceptability in the target audience. Technical considerations, such as subtitle duration and synchronization with visual cuts, are crucial for maintaining a seamless viewing experience. In light of these complexities, the integration of AI, particularly ChatGPT, in subtitling processes offers promising solutions for enhancing efficiency and accuracy. However, the extent to which AI can address the intricate challenges of subtitling, particularly in multilingual and culturally diverse contexts, warrants further exploration.

1.2 Natural language processing (NLP and machine learning

Natural Language Processing (NLP) is a dynamic research domain focused on developing computer algorithms capable of comprehending and processing human language. Recent strides in NLP, epitomized by sophisticated deep learning models such as GPT-3, have demonstrated remarkable potential in generating text that closely resembles human speech (Hendy et al., [2]). The emergence of these language models, including NLP and ChatGPT, has ushered in a new era of possibilities for businesses, governments, and organizations globally, enabling machines to analyze and interpret natural language data in unprecedented ways (Tan et al., [3]). Consequently, NLP and ChatGPT have become indispensable tools for entities aiming to leverage natural language data to optimize their operations.

Within the realm of translation, machine learning has garnered increasing traction, particularly in the context of NLP. Unlike conventional rule-based machine translation systems that rely on manually crafted rules, machine learning-based approaches harness vast datasets to discern patterns and rules, significantly augmenting translation accuracy (Banat & Adla, [4]). Through the adaptation to language variations and context, machine learning-based translation methods have demonstrated substantial enhancements in translation quality (Tan et al., [3]).

Neural Machine Translation (NMT), a prominent machine learning-based approach, has gained widespread acclaim for its exceptional performance across diverse language pairs [5]. Employing artificial neural networks to learn mappings between source and target languages, NMT systems leverage multiple interconnected layers to progressively process input data and generate desired outputs [6]. One of NMT's notable advantages lies in its ability to address word variations within sentences, surmounting the limitations of previous rule-based and statistical approaches, particularly with regard to handling lengthy and intricate sentences. By adeptly capturing the intricacies and idiomatic expressions of human language, NMT facilitates effective translations at both the sentence and contextual levels (Jiao et al., [7]). Its adaptability is further underscored by customizable neural network architectures tailored to specific languages, domains, and tasks, thereby positioning NMT as a contemporary solution for diverse translation requirements. Leveraging extensive parallel corpora for training, NMT models continually refine their translation capabilities, resulting in enhanced translation fidelity [8].

The advent of large-scale pre-trained language models like GPT-3 has revolutionized NLP, particularly in the domain of machine translation. Trained on vast textual data, GPT-3 exhibits remarkable proficiency in generating human-like text with exceptional accuracy and fluency. A notable advantage of GPT-3 lies in its adaptability, as these models can be fine-tuned for specific tasks, such as translation, with minimal task-specific training data, underscoring the efficacy of transfer learning (Hendy et al., [2]). GPT-3's formidable performance in various language tasks, including translation, can be attributed to its extensive transformer-based architecture, which excels in processing and generating text with precision and coherence, even in the absence of explicit rules or prompts (Jiao et al., [7]).

Despite the substantial advancements facilitated by pre-trained language models like GPT-3, challenges persist. The computational intensity associated with training these models may necessitate specialized hardware for optimal performance. Moreover, the quality of translation outputs is intricately linked to the quantity and quality of training data, as well as the chosen fine-tuning strategy (Tan et al., [3]). Additionally, concerns surrounding the interpretability of machine learning-based translation systems, often perceived as opaque "black boxes," underscore the importance of transparency and interpretability in translation decisions [8].

In conclusion, machine learning-based translation, epitomized by approaches such as NMT and pre-trained language models like GPT-3, has significantly mitigated the limitations of traditional rule-based methods, culminating in substantial improvements in translation quality. However, ongoing efforts are warranted to address challenges related to dataset availability, system interpretability, and computational intensity, thereby ensuring the continued advancement and efficacy of machine learning-based translation methodologies.

1.3 ChatGPT: a powerful language model for translation and conversation

ChatGPT, a cutting-edge language model developed by OpenAI based on the GPT-3.5 architecture, has garnered significant attention for its ability to generate natural and coherent responses across a wide array of inputs, including text prompts, questions, and conversations. Rooted in a transformer architecture, particularly well-suited for language processing tasks (Shyam et al., [9]), GPT-3 stands out for its vast scale, boasting over 175 billion parameters in its largest iteration. Trained on a diverse corpus of text data encompassing books, articles, and websites, the model learns intricate patterns and relationships inherent in language.

At its core, the GPT-3 architecture comprises a series of transformer blocks interconnected in a feedforward network. Each block consists of two sub-layers: a multi-head self-attention mechanism and a position-wise fully connected feedforward network. The self-attention mechanism enables the model to focus on different segments of input text, while the feedforward network applies non-linear transformations to input features [6].

Employing a technique known as "unsupervised pre-training," GPT-3 learns language patterns and relationships by predicting the next word in a sequence based on preceding words (Brown et al., [10]). This approach enables the model to understand intricate linguistic structures and nuances. To employ ChatGPT for translation tasks, input text undergoes tokenization and embedding, with each token representing its meaning through an embedding vector. These embeddings are then processed through the model's encoder, which generates a hidden representation of the text through

multiple layers of self-attention and feedforward networks. Subsequently, the decoder generates the output text in the target language, leveraging learned patterns to produce coherent translations [11].

ChatGPT's effectiveness in translation tasks stems from its ability to handle diverse inputs and generate contextually relevant translations. Its transformer architecture, particularly the self-attention mechanism, facilitates the capture of intricate linguistic patterns [12]. While excelling in generating fluent translations, the model's performance may vary based on text complexity and domain specificity. Notably, ChatGPT's adaptability and ability to learn from diverse datasets distinguish it from traditional rule-based translation systems, making it a versatile tool for various language tasks (Brown et al., [10]).

Despite its prowess, evaluating ChatGPT's translations for accuracy, especially in domains requiring precision and cultural sensitivity, remains paramount. Continuous refinement and fine-tuning based on user feedback contribute to enhancing the model's performance over time. As such, ChatGPT represents a significant advancement in language modeling, underscoring the potential of large-scale transformer architectures for translation and conversational applications [6], Brown et al., [10]).

1.4 AI role in subtitling

AI is increasingly employed to provide intralingual subtitles for movies, particularly to aid individuals with hearing impairments. Subtitles also play a pivotal role in news broadcasts, especially during live transmissions, where real-time subtitling is indispensable [13]. However, achieving proficiency in real-time subtitling traditionally requires years of practice, and despite efforts, quality often falls short. In recent years, there has been a shift towards using AI-generated subtitles, leveraging Automatic Speech Recognition (ASR) techniques to automate the subtitling process [14].

Nevertheless, AI-generated subtitles encounter challenges, and their quality can be deficient. ASR systems may struggle with accurately transcribing spoken words due to factors such as variations in speech, dialects, casual speech, noisy environments, and background sounds [14, 15]). ASR systems rely on probability measurements, introducing uncertainties in transcriptions, which can be problematic, especially in the context of live news where conveying clear, credible, and non-misleading information is paramount [16].

To illustrate potential risks, consider a hypothetical scenario: during a news report on the ongoing pandemic, the news reporter states, "Quarantine is not over, you are not allowed to go out," while the AI-generated subtitle reads, "Quarantine is now over, you are now allowed to go out" [16]. If a viewer, like Sara with the sound muted, relies solely on the subtitles, misinformation and misinterpretation can occur, underscoring the critical nature of transcription accuracy.

Human cognitive processes play a vital role in understanding information, capable of correcting mistakes, building possible meanings, and predicting subsequent words (Holt et al., [17]). Acknowledging transcription uncertainties and presenting them to users could potentially prevent misinterpretations [16]. Previous studies suggest that visualizing uncertainty aids users in data interpretation and decision-making [16, 18]). However, in the case of AI-generated subtitles, the complexity of uncertain data visualization may either enhance understanding or lead to confusion, impacting trust [16]. The effectiveness of uncertainty visualization in this context remains unclear and requires further exploration (Kay et al., 2016).

The integration of artificial intelligence (AI) in language-related tasks has gained significant attention in recent research, particularly in translation and educational contexts. Studies have examined AI's potential to enhance learning, facilitate language acquisition, and improve efficiency in translation tasks, including subtitling.

Mohamed et al. (2024) explored AI's impact on student motivation and learning across various cultural contexts. Their findings indicate that AI tools, including ChatGPT, can foster critical thinking, autonomy, and engagement. While the study primarily focuses on education, its implications extend to AI-assisted translation, where maintaining fluency and coherence is essential. Similarly, Mohamed [19, 20]) investigated the role of ChatGPT in English as a Foreign Language (EFL) instruction, highlighting its ability to provide rapid, contextually relevant responses. However, concerns remain about AI's potential to reinforce biases and hinder users' critical thinking—issues that are particularly relevant to AI-driven subtitling, where preserving the intended meaning and avoiding distortions is crucial.

Further, Abdelaal and Al Sawi [21] examined university professors' perceptions of AI integration in education, revealing both enthusiasm and apprehension. While AI is recognized as a valuable tool, concerns persist regarding its autonomy, financial constraints, and the potential dehumanization of teaching. These findings align with broader discussions on AI's role in translation, where balancing technological efficiency with human expertise remains a challenge.

Together, these studies underscore AI's growing influence in language processing and translation. While AI-driven tools like ChatGPT enhance accessibility and efficiency, questions about accuracy, fluency, and ethical considerations persist. This discussion provides a foundation for analyzing ChatGPT's effectiveness in movie subtitling, particularly in comparison to human translators regarding linguistic precision, contextual adaptation, and cultural nuance.

While the initial literature review provides an overview of key contributions to the field, a more detailed synthesis and critical contrast of existing research can offer greater insight into the current state of knowledge and help identify research gaps. Several studies within the field share common ground, particularly regarding the effectiveness of AI in translation and subtitling tasks. For instance, Díaz-Cintas and Remael [22] emphasize the importance of context and cultural nuance in audiovisual translation, a view echoed by House [29], who also stresses the critical role of communication across languages. These studies align in their understanding that translation, whether manual or automated, must balance linguistic accuracy with cultural relevance.However, there are notable disagreements between studies in their evaluation of AI's capacity for subtitling and translation. For example, Jiao et al. [7] raise concerns about the limitations of AI, particularly in dealing with complex idiomatic expressions and culturally bound terms. This contrast underscores the ongoing debate regarding the limitations of AI in capturing the subtle cultural and contextual factors critical to high-quality translation, especially in audiovisual contexts. Additionally, although there is a growing consensus that AI models like GPT-4 show promise for professional translation tasks [8], conflicting views are still evident regarding their ability to handle nuances such as humor, irony, and cultural references, which are often crucial in audiovisual translation. Furthermore, studies on the perception of AI in translation have highlighted divergent views. While some scholars, like Mohamed et al. [20], argue that AI tools are welcomed by educators and translators for their efficiency, others, such as Pym (2012), express skepticism, cautioning that AI might undermine the role of human translators in maintaining quality and cultural sensitivity. This disagreement between the optimism of AI proponents and the caution of those focused on human expertise indicates a gap in research that could be explored further. More research could address how AI and human translators can work synergistically rather than in competition, particularly in specialized domains like audiovisual translation.In terms of the future of AI in subtitling and translation, there is a lack of consensus on the specific challenges that need to be addressed. For instance, while Tan et al. [3] identify machine translation's dependency on large corpora and data-driven models as a key limitation, others, like Vaswani et al. [6], suggest that new developments in neural machine translation are paving the way for more adaptive systems capable of handling a variety of linguistic complexities. This points to a gap in the literature regarding the potential for integrating emerging machine learning approaches with traditional human practices.By synthesizing these differing viewpoints, this study aims to not only clarify the existing state of knowledge but also to pinpoint areas where current research is insufficient or contradictory. In particular, it seeks to explore the intersections between AI's technological potential and the cultural nuances that human translators typically navigate, proposing a framework for how both can coexist and complement each other in subtitling practices. Through a more comprehensive understanding of these contrasting perspectives, this paper intends to contribute to the ongoing discussion and offer new insights into the evolving relationship between AI and human translation efforts.

In the context of this paper, while some studies have discussed subtitling quality (e.g., [23, 24]), no research has been conducted to examine the feasibility of using AI, especially ChatGPT, in subtitling movies. Therefore, the current study aims to assess the extent to which ChatGPT can serve as a valuable tool in subtitling movies, taking into account the challenges and potential risks associated with AI-generated subtitles.

1.5 Research questions

1. To what extent are the ChatGPT translation of subtitles are accurate and fluent?
2. Is there statistically significant difference between the human translation and ChatGPT translation of subtitles?

1.6 Methodology

A corpus consisting of the Our Cartoon President" series and its Arabic translation was established, aligning source texts (STs) with their target translations (TTS). Subsequently, thirty-two (32) examples were randomly selected from the "Our Cartoon President" series and presented to ChatGPT 3.5 for comparison with human translations sourced from the Subscene website. The human translations available on Subscene were crafted by professional translators.

The choice of 'Our Cartoon President' for this study was primarily driven by its satirical and politically charged content, which provides a rich context for examining the challenges and opportunities presented in subtitling and translation. This animated show is particularly well-suited for an exploration of how subtitling strategies are employed to balance humor,

cultural references, and political commentary, making it an ideal case study for understanding subtitling dynamics in the context of complex cultural materials.

The 32 examples selected for this study were chosen based on their relevance to key subtitling challenges, including the presence of culturally bound terms, humor, taboo language, and the nuances of political satire. These examples were specifically drawn from *Our Cartoon President* to ensure that a variety of subtitling techniques and strategies were represented, offering a comprehensive overview of the subtitling process in a complex animated political context. Additionally, the examples were selected to illustrate a balance of both linguistic and cultural complexity, ensuring a diverse and representative sample.

In selecting a sample size of 32, a number of studies support the use of an adequate number of examples to cover the subtitling challenges identified above. For instance, Fink [25] suggests that a sample size between 10 and 30 is often sufficient for expert panels in specialized fields, particularly when examining specific phenomena such as humor and culturally bound terms in subtitling. Creswell [26] further supports this, noting that the sample size for expert surveys should reflect the research goals and the need for detailed insights, with smaller groups being suitable for qualitative investigations. Additionally, Blaikie [27] and Sargeant et al. [28] argue that smaller sample sizes, such as the 32 examples used in this study, are appropriate for studies that focus on expert evaluation, particularly in cases where linguistic and cultural complexity are involved.

Thus, the 32 examples in this study not only serve to address the main challenges in subtitling but also align with best practices for qualitative research in specialized fields. The strategic selection of these examples ensures that both linguistic and cultural challenges are covered in a way that provides valuable insights into the subtitling techniques employed in *Our Cartoon President*, offering a comprehensive view of how political satire, humor, and cultural nuances are handled in the subtitling process.

The outputs generated by ChatGPT and the human translations were compiled into a survey format and submitted to English-Arabic translation professors for evaluation.

The professors were asked to assess the fluency and accuracy of each translation using a scale ranging from 1 to 5. The collected survey data was analyzed using SPSS, incorporating a T-test to examine whether there was a statistically significant difference between human and ChatGPT translations.

The survey is designed to assess the fluency and accuracy of translations in subtitling, with a focus on evaluating translations of political satire from *Our Cartoon President*. Respondents are asked to rate the fluency of translations on a scale of 1 to 5, where:

- 1 represents "Not Fluent at All," with incoherence and numerous errors,
- 2 represents "Limited Fluency," with some coherence but noticeable errors,
- 3 represents "Moderate Fluency," with good coherence and occasional errors,
- 4 represents "Fluent," with smooth flow and minor errors,
- 5 represents "Highly Fluent," with flawless coherence and minimal to no errors.

This scale is designed to help evaluate how well the translations capture the nuances of the original content while maintaining clarity, readability, and coherence in the target language. The survey results will provide insights into the effectiveness of subtitling techniques and strategies, particularly in handling linguistic and cultural challenges in political satire.

This study embraces a mixed-methods approach, incorporating both qualitative and quantitative components. Qualitative data was collected and analyzed through the evaluation provided by translation professors, while quantitative data was obtained from the professors' ratings on fluency and accuracy. The combination of these methods offers a comprehensive understanding of the comparative performance of human and ChatGPT translations in the context of English-to-Arabic translation of the "Our Cartoon President" series.

2 Results and discussion

2.1 Qualitative results

This section presents the qualitative findings of the study. Table 1 shows selected subtitles extracted from the English Cartoon Series, along with both the human-generated translation and the ChatGPT-generated translation. In this section, ChatGPT and the human translations translations will be assessed qualitatively for accuracy and fluency.

Table 1 English subtitles alongside human and ChatGPT translations

Example no	ST	HT	CHATGPT
1	00:00:14,889 – > 00:00:17,459 I'm Chuck Todd, and no one remembers hiring me	00:00:13,626 – > 00:00:16,042 (أدعى (تشاك تود ولا أحد يذكر توظيفي 3	00:00:14,889 – > 00:00:17,459 أنا تشاك تود، ولا يتذكر أحد تعييني
2	4 00:00:17,484 – > 00:00:19,990 The New York Times has obtained ten years of figures	00:00:16,167 – > 00:00:18,542 (استحصلت صحيفة (نيويورك تايمز على مستحقات عشر سنوات	4 00:00:17,484 – > 00:00:19,990 حصلت صحيفة نيويورك تايمز على معلومات تمتد لعشر سنوات
3	5 00:00:20,015 – > 00:00:21,693 from President Trump's tax returns,	4 00:00:18,667 – > 00:00:20,334 من إقرارات (ترامب) الضريبية	5 00:00:20,015 – > 00:00:21,693 من إقرارات ضريبة الرئيس ترامب
4	6 00:00:21,718 – > 00:00:24,052 and for many of those years, he lost more money	5 00:00:20,459 – > 00:00:22,792 واتضح أنه في أغلبية هذه السنوات خسر المال	6 00:00:21,718 – > 00:00:24,052 وخلال العديد من تلك السنوات، فقد أكثر من المال

2.2 Example 1: pragmatic failure and literal translation

In this instance, the translation produced by ChatGPT, "ولا يتذكر أحد تعييني," fails to capture the intended pragmatic meaning of the source text. The phrase inaccurately conveys a lack of memory rather than the intended implication that nobody wants to hire Chuck Todd. Pragmatic failures in machine translation (MT) are well-documented in translation studies [29, 30], particularly in relation to implicature (Grice, 1975). In this case, the incorrect selection of "يتذكر" *(remembers)* instead of "يريد" *(wants)* reflects a common shortcoming of AI-driven translation models, which often struggle with non-literal or context-dependent meanings (Venuti, 2018) (Table 2, 3, 4, 5, 6, 7).

Furthermore, fluency is another concern in this translation. According to Toury's [31] notion of acceptability in target language norms, a translation should sound natural in the TL while preserving the ST's meaning. Neither the MT nor the human translation (HT) achieves this, as the HT maintains a highly literal structure, which may not align with Arabic syntactic and pragmatic norms. This aligns with the findings of Pym (2012), who highlights the trade-off between literal accuracy and fluency in translation.

2.3 Example 2: lexical choices and pragmatic meaning

A significant discrepancy exists between the human translation and ChatGPT's output in this case. The human translator employs an uncommon Arabic word, "استحصلت," to translate "obtain," which, while accurate, is not commonly used in everyday Arabic. On the other hand, ChatGPT's translation is unexpectedly more idiomatic, though it still fails to fully capture the contextual pragmatic meaning. This highlights the limitations of MT in handling lexical variation, as discussed by Newmark [32], who argues that word-for-word translation often leads to unnatural phrasing in the TL.

Pragmatically, context plays a crucial role in lexical selection, as emphasized by Hatim and Mason [33]. ChatGPT's choice may be idiomatic but lacks sensitivity to the nuanced implications embedded in the original text. This aligns with Reiss and Vermeer's [34] Skopos theory, which suggests that translations should consider the intended function of the text rather than relying solely on word-level equivalence.

2.4 Example 3: syntactic issues and word order

The human translation in this case appears more accurate, effectively conveying the source text's meaning while preserving the syntactic norms of Arabic. In contrast, ChatGPT's translation, although conceptually correct, exhibits a structural flaw related to word order, specifically with "ضريبة." Placing "ضريبة" at the end of the phrase would improve clarity, aligning with Arabic's syntactic structure and ensuring the correct modification of "إقرارات ترامب."

Word order errors in MT have been widely studied, with researchers such as Koehn [35] emphasizing how neural machine translation (NMT) systems tend to struggle with target languages that have flexible word order, like Arabic. Additionally, research by Dorr et al. [36] suggests that syntactic divergences between English and Arabic, particularly in adjective-noun placement and verb-subject-object order, contribute to such translation issues.

2.5 Example 4: idiomatic fluency and naturalness

In this case, the human translation successfully conveys the intended meaning in an idiomatic and fluent manner. ChatGPT's output, although conceptually accurate, lacks idiomatic naturalness, demonstrating a tendency to adhere closely to the source text at the expense of readability. This issue is consistent with findings by Munday [37], who notes that AI-driven translations often prioritize lexical and syntactic correspondence over idiomatic fluency.

The problem can also be analyzed through the lens of Nida's [38] concept of dynamic equivalence, which stresses that translation should reproduce the effect of the source text rather than adhere rigidly to its structure. Since Arabic and English exhibit significant differences in idiomatic expression, a literal approach often results in awkward phrasing, reinforcing the need for post-editing when using AI-generated translations (Toral & Way, [39]).

Table 2 ChatGPT and HT

No		HT	ChatGPT
5	14 00:00:38,555 –> 00:00:41,289 Who am I kidding, Rudy?! I'm so frickin' embarrassed!	00:00:37,250 –> 00:00:39,959 (لا فائدة من إنكار الأمر يا (رودي !يا له من إحراج	14 00:00:38,555 –> 00:00:41,289 !من أنا أمازح، رودي؟! أنا محرج جدًا
6	15 00:00:41,376 –> 00:00:43,968 Be honest. Do you think being exposed as a broke fraud	13 00:00:40,125 –> 00:00:41,834 !كن صريحاً معي هل تعتقد أن فضحي	15 00:00:41,376 –> 00:00:43,968 كن صادقاً. هل تعتقد أن الكشف عني كنصاب مفلس
7	16 00:00:43,993 –> 00:00:46,538 will affect people's perception that I'm not a broke fraud?	14 00:00:42,000 –> 00:00:45,167 كوني منافقاً مفلساً سيؤثر في تصور الآخرين بأنني لست كذلك؟	16 00:00:43,993 –> 00:00:46,538 سيؤثر في انطباع الناس بأنني لست نصابًا مفلسًا؟
8	23 00:00:57,196 –> 00:00:58,718 We gotta get out of this news cycle	19 00:00:55,959 –> 00:00:57,375 يجب أن تنسانا وسائل الإعلام	23 00:00:57,196 –> 00:00:58,718 يجب أن نتجاوز دورة الأخبار هذه
9	24 00:00:58,743 –> 00:01:00,585 I need you to FaceTime Brian Williams,	20 00:00:57,501 –> 00:01:00,834 اتصل مرئياً بـ(براين ويليامز	24 00:00:58,743 –> 00:01:00,585 أحتاج منك أن تتصل ببرايان وليامز عبر FaceTime
10	25 00:01:00,610 –> 00:01:01,944 and I'm gonna show him my junk!	) لأعرض عليه منطقتي الحساسة	، 25 00:01:00,610 –> 00:01:01,944 !وسأظهر له جزءًا من جسمي

Table 3 ChatGPT and HT translations

No	ST	HT	ChatGPT
11	77 00:03:09,008 – > 00:03:11,766 Dad, that Mueller report stirred up some old feelings 78 00:03:11,791 – > 00:03:13,726 for Trump Tower Moscow, huh?	62 00:03:07,751 – > 00:03:11,834 حرّك تقرير (مولر) بعض المشاعر (في (موسكو (في قلب برج (ترامب 63 00:03:11,959 – > 00:03:12,959 أليس كذلك؟	77 00:03:09,008 – > 00:03:11,766 أبي، هذا تقرير مولر أثار بعض المشاعر القديمة 78 00:03:11,791 – > 00:03:13,726 حول برج ترامب في موسكو، صحيح؟
12	85 00:03:30,587 – > 00:03:34,623 Vladimir Putin has sold the land reserved for Trump Tower Moscow	70 00:03:29,292 – > 00:03:33,250 باع (فلادمير بوتن) الأرض المخصصة (لبناء برج (ترامب) في (موسكو	85 00:03:30,587 – > 00:03:34,623 قد باع فلاديمير بوتين الأرض المخصصة لبرج ترامب في موسكو
13	86 00:03:34,648 – > 00:03:37,447 to Amazon CEO Jeff Bezos,	71 00:03:33,375 – > 00:03:36,083 (لمدير (أمازون) التنفيذي (جيف بيزوس	86 00:03:34,648 – > 00:03:37,447 للرئيس التنفيذي لأمازون جيف بيزوس،
14	87 00:03:37,449 – > 00:03:39,449 a man whose head shot and dick pic 88 00:03:39,451 – > 00:03:41,385 are indistinguishable to the human eye	72 00:03:36,209 – > 00:03:40,083 الرجل الذي لا يمكن تفرقة صورة عن رأسه لصورة عن منطقته الحساسة	87 00:03:37,449 – > 00:03:39,449 00:03:39,451 88 رجل تكنولوجيا يصعب التمييز بين صورة رأسه 00:03:41,385 – > وصورة خاصته الجنسية في عين الإنسان

Table 4 ChatGPT and HT

No	ST	HT	ChatGPT
15	139 00:05:33,531 – > 00:05:36,696 I can't believe we have the majority in the House	113 00:05:32,250 – > 00:05:35,417 لا أصدق أننا حصلنا على الأغلبية في مجلس النواب	139 00:05:33,531 – > 00:05:36,696 لا أستطيع أن أصدق أن لدينا الأغلبية في مجلس النواب
16	140 00:05:36,822 – > 00:05:38,719 And the minority in the Senate,	114 00:05:35,542 – > 00:05:37,459 كما حصلنا على الأقلية في مجلس الشورى	140 00:05:36,822 – > 00:05:38,719 والأقلية في مجلس الشيوخ،
17	141 00:05:38,790 – > 00:05:40,977 which is even more powerful than the majority	115 00:05:37,584 – > 00:05:39,667 وهذا أقوى من الأغلبية	141 00:05:38,790 – > 00:05:40,977 والتي هي أقوى حتى من الأغلبية

Table 5 ChatGPT and HT translations

No	ST	HT	ChatGPT
18	181 00:07:40,304 – > 00:07:42,109 I don't need a hotel 182 00:07:42,133 – > 00:07:44,507 with 80 Russian hookers roaming the hallways	151 00:07:40,125 – > 00:07:43,000 لست بحاجة إلى فندق تجوب ردهاته 80 بائعة هوى روسية	181 00:07:40,304 – > 00:07:42,109 لا أحتاج إلى فندق 182 00:07:42,133 – > 00:07:44,507 مع 80 عاهرة روسية تجوب الأروقة
19	194 00:08:10,337 – > 00:08:12,476 I will kick your ass!	160 00:08:09,042 – > 00:08:12,125 فسأضربك شر ضرب إلنفعل ذلك	194 00:08:10,337 – > 00:08:12,476 سأركل مؤخرتك! .

2.6 Example 5: literal translation and pragmatic nuances

In Example 5, ChatGPT's translation deviates significantly from the intended meaning by offering a literal rendition that fails to capture the pragmatic and contextual nuances. This reflects a common challenge in machine translation, where AI models often rely on direct, surface-level word mapping without considering the deeper, intended meaning of the text [30]. The human translation, although not fully capturing the tone, provides a more accurate representation of the source text, highlighting the tension between accuracy and tonal fidelity. However, the human translator misinterprets the question's point by rendering it as "no point to dent it," which, although more contextually appropriate than ChatGPT's output, still falls short. This error demonstrates the human tendency to prioritize fluency over accuracy in certain contexts, a phenomenon well-discussed by Nida [38], who emphasizes the balance between dynamic equivalence and formal equivalence in translation (Table 2).

Despite the human error, the human translation remains superior to ChatGPT's literal approach, underscoring a key distinction between human and machine translation: humans are better at understanding and interpreting subtleties of tone and context (Hatim & Mason, [33]).

2.7 Example 6: technical accuracy and colloquial language

In Example 6, ChatGPT outperforms the human translation by more effectively conveying the intended meaning and technicalities of the source text. The human translation attempts to convey the subtitle's meaning but results in an incomplete sentence, highlighting the difficulty of maintaining both fluency and completeness in machine translation [35]. However, a suggestion is made to replace the colloquial term "كنصاب" (fraudster) with the more formal "محتال"

Discover Artificial Intelligence (2025) 5:257 | https://doi.org/10.1007/s44163-025-00310-z

Table 6 ChatGPT and HT translations

No	ST	HT	ChatGPT
20	294 00:12:18,561 – > 00:12:22,101 If there's a hell, I'm pretty sure we're gonna go	231 00:12:17,375 – > 00:12:20,709 أنا واثق من أن مصيرنا الجحيم إن كان له وجود	294 00:12:18,561 – > 00:12:22,101 إذا كان هناك جحيم، فأنا متأكد تمامًا أننا سنذهب
21	295 00:12:22,126 – > 00:12:24,695 Oh, Jesus. Enough with the hell talk, Eric	232 00:12:20,834 – > 00:12:23,334 إيا إلهي (كفاك كلاماً عن الجحيم يا (إريك	295 00:12:22,126 – > 00:12:24,695 يا إلهي. كفاية مع حديث الجحيم، إيريك
22	296 00:12:24,720 – > 00:12:26,500 Everything's gonna be fine,	233 00:12:23,459 – > 00:12:24,918 سيكون كل شيء بخير	296 00:12:24,720 – > 00:12:26,500 ،كل شيء سيكون على ما يُرام
23	297 00:12:26,525 – > 00:12:29,226 after Giuliani turns in another flawless performance	234 00:12:25,000 – > 00:12:27,834 (بعد أن يقدم (جولياني أداء آخر خالياً من الشوائب	297 00:12:26,525 – > 00:12:29,226 بعد أن يقدم جولياني أداءً خاليًا من العيوب آخر
24	298 00:12:29,505 – > 00:12:32,864 Isn't this hotel an obvious scam designed to enrich the President?	235 00:12:28,375 – > 00:12:31,459 أليس هذا الفندق خدعة جديدة مصممة لإثراء الرئيس؟	298 00:12:29,505 – > 00:12:32,864 أليس هذا الفندق عبارة عن احتيال واضح مصمم لإثراء الرئيس؟

Table 7 ChatGPT and HT translations

No	ST	HT	ChatGPT
25	311 00:12:58,008 – > 00:13:00,297 - Wow. Ok… Rings a bell - … fucking asshole!	246 00:12:56,876 – > 00:12:58,999 - حسناً… قرع ناقوس الخطر - أيها الأحمق اللعين!	311 00:12:58,008 – > 00:13:00,297 • واو. حسنًا… يثير ذلك ذكريات • إيا عاهر …
26	318 00:13:13,000 – > 00:13:16,888 Join the new "I'm with Blue, how about you?" campaign	252 00:13:11,792 – > 00:13:15,542 انضموا إلى حملة "أنا مع الحزب الديمقراطي فماذا عنك؟"	318 00:13:13,000 – > 00:13:16,888 "انضم إلى حملة "أنا مع الأزرق، ماذا عنك؟
27	324 00:13:26,648 – > 00:13:28,341 Low-key, people are tired 325 00:13:28,366 – > 00:13:30,693 of scripted, focus-grouped messages	257 00:13:25,417 – > 00:13:29,250 الكلام في سرك ولكن الشعب سئم من الرسائل المنهجية الموجهة لفئة معينة	324 00:13:26,648 – > 00:13:28,341 00:13:28,366 325 بشكل هادئ، الناس متعبة – > 00:13:30,693 من الرسائل المكتوبة ومركبة من المجموعات المستهدفة
28	290 00:12:08,892 – > 00:12:10,826 Maybe get off your rich, bony asses	227 00:12:07,626 – > 00:12:09,626 لماذا لا تتحركان أيها الثريان	290 00:12:08,892 – > 00:12:10,826 ربما انزلقوا من على مؤخرتيكم الثرية والعظمية

(swindler), which would improve the translation's overall register. This aligns with Reiss and Vermeer's [34] Skopos theory, which advocates for adapting the translation based on the intended function and audience, especially in technical or formal contexts.

2.8 Example 7: conciseness and display suitability

For Example 7, both the human and machine translations are accurate, but ChatGPT's brevity makes its translation more suitable for on-screen display. This highlights an important aspect of audiovisual translation, where space limitations necessitate concise yet effective renditions (Diaz Cintas & Remael, [1]). ChatGPT's translation ensures ease of presentation without sacrificing meaning, supporting the view that brevity does not necessarily compromise accuracy. As Pym (2012) notes, MT can excel in situations where conciseness is prioritized, especially in subtitling tasks that require the compression of information.

2.9 Example 8: idiomatic and contextual accuracy

In Example 8, both the human and machine translations are inaccurate. The human translation attempts to use an idiomatic expression but fails to accurately convey the source text's intended meaning, while ChatGPT's use of "دورة" (cycle) is less idiomatic and awkward in context. A suggested revision, "دوامة الأخبار" (news spiral), would be more idiomatic and contextually appropriate. This example underscores a frequent challenge in both human and machine translations—rendering idiomatic expressions accurately. According to Munday [37], idiomatic expressions often present translation difficulties due to their reliance on culturally specific nuances, which AI models may struggle to grasp.

2.10 Example 9: visual interaction and terminology

In Example 9, both translations are appropriate. The human translator opts for "مرئياً" (visually), while ChatGPT retains *"FaceTime,"* which also conveys the notion of visual interaction effectively. This example illustrates the flexibility of machine translation when translating proper nouns or widely recognized terms. The use of *"FaceTime"* by ChatGPT is an instance of what Newmark [32] refers to as "transference," where the source term is kept unchanged due to its universal recognition. Both renditions are contextually relevant, demonstrating that certain terms do not require localization, particularly when they are globally understood (Venuti, 2018).

2.11 Example 10: euphemism and specificity

In Example 10, both the human and machine translators offer euphemistic translations. ChatGPT uses a more general expression, "جزءًا من جسمي" (a part of my body), while the human translator is more specific with "منطقتي الحساسة" (my private parts). The human translation is considered more accurate due to its specificity. This example highlights the challenge of maintaining appropriateness in sensitive contexts. According to Toury [31], the translator's decision in such situations is often guided by cultural norms and expectations, especially when rendering euphemisms or terms related to the body. The human translator's choice of a more precise term aligns with these cultural expectations, while ChatGPT's generalization may fail to convey the full intended impact of the source text (Table 3).

2.12 Example 11: faithfulness vs. idiomatic expression

In Example 11, both the human and ChatGPT translations seem appropriate but exhibit different approaches to conveying the intended meaning. The human translation omits the word "old" from the target text and instead uses the phrase "حرّك بعض المشاعر" (stirred some feelings), which represents a more generalized expression. This approach sacrifices some precision for emotional resonance, aiming to evoke a broader sentiment rather than sticking to the literal words of the source. On the other hand, ChatGPT remains more faithful to the source text, rendering "أثار بعض المشاعر القديمة" (stirred some old feelings), which is a more literal translation. However, this translation does not fully capture the idiomatic nature of the source text. Both translations convey the essential meaning but lack the idiomatic fluency often necessary for effective communication in a target language.

This phenomenon reflects a long-standing challenge in translation studies: the tension between "faithfulness" to the source text and the production of a fluent, idiomatic target text (Venuti, 2018). The human translator's decision to omit "old" can be seen as an attempt to prioritize naturalness in the target language, while ChatGPT's more literal approach risks a stilted rendering. Studies have suggested that achieving a balance between both aspects is key to successful translation [31, 38]. In this case, ChatGPT's translation, while still somewhat literal, manages to retain greater fidelity to the original expression while introducing an idiomatic touch. This aligns with Reiss and Vermeer's [34] Skopos theory, which emphasizes that translations should be adapted according to the communicative function and the target audience.

2.13 Example 12: explicitation and clarity

In Example 12, both translations—human and ChatGPT—are accurate, but the human translation stands out due to its greater precision. By incorporating the word "لبناء" (for building), the human translator uses an explicitation strategy that clarifies the meaning in a more explicit manner than ChatGPT's translation. This choice enhances the clarity of the translation and makes it more accessible to the audience, especially in the context of subtitles, where brevity and precision are crucial [40].

Explicitation is a common strategy in translation when there is a need to add information not explicitly stated in the source text. According to Klaudy [41], explicitation often improves the transparency of the target text, making it easier for the audience to understand. In this case, the human translator's decision to include "لبناء" ensures that the target audience has a clear understanding of the source text's purpose. This highlights the translator's role in enhancing readability and ensuring the meaning is communicated effectively, particularly when dealing with technical or complex language.

2.14 Example 13: minor syntactic variations

In Example 13, both the human translation and ChatGPT's translation are highly accurate. The simplicity and clarity of the source text result in translations that are almost identical, with only a minor difference in syntactic structure. This discrepancy, while present, is considered negligible in terms of meaning, as both translations effectively convey the intended message. Such minor syntactic variations are common in translation and often result from the inherent flexibility of language structures across cultures [32].

According to Gile [42], such small differences typically do not affect the overall quality of the translation and may even reflect the translator's individual style or the specific syntactic rules of the target language. In this example, both translations effectively achieve their communicative purpose, suggesting that small syntactic differences do not always hinder the clarity of the message.

2.15 Example 14: euphemism and cultural sensitivity

Example 14 presents a more complex translation challenge, particularly due to the sensitive cultural nature of the content. The human translation is accurate, but there is an issue with the preposition choice. The translator uses "عن" (about), which may not convey the intended sense of belongingness. A more suitable preposition, "ل" (for), could be used to better reflect the source text's nuance. This minor issue exemplifies the delicate decisions translators must make when working with cultural content, especially when dealing with euphemisms or sensitive topics [30].

In this instance, the human translator uses the euphemistic phrase "منطقته الحساسة" (his sensitive area), which offers a softer, less explicit translation of a term related to male anatomy. This strategy aligns with strategies of indirectness and avoidance in translation (Beaugrande & Dressler, [43]). A suggestion is made to replace "منطقته الحساسة" with "عضو ذكري" (his genitals), which is a more direct translation that maintains a refined tone. This shift reflects a balancing act in translation: how to preserve the intended meaning while adhering to cultural norms of decency.

ChatGPT's translation introduces words not present in the source text, such as *"technology,"* and uses a different phrase to translate *"dick"* as "خاصته الجنسية" (his sexual organ). While this translation is somewhat accurate, it deviates from the directness of the source text. The introduction of additional words may be an attempt by the AI to fill gaps in understanding, but it leads to an overly descriptive and somewhat awkward translation [35]. Despite these differences, both translations align in their goal of conveying sensitive content while maintaining a certain degree of appropriateness, which highlights the challenges of handling euphemisms in translation (Hatim & Mason, [33]).

2.16 Example 15: accuracy and conciseness in subtitling

In Example 15, both the human and ChatGPT translations demonstrate accuracy, with the translations reflecting the direct, simple nature of the source subtitle. The human translation is notably more concise than the ChatGPT translation, which aligns with best practices in subtitling. Conciseness is essential in subtitling, as it ensures that the translation remains easily readable while preserving the essence of the source text. As studies in audiovisual translation have indicated, brevity is a crucial aspect of effective subtitling, ensuring that viewers can read the subtitles within the time frame of each scene without compromising comprehension (Diaz Cintas & Remael, [1]) (Table 4).

Moreover, both translations make the same choice in rendering "the House" as "مجلس النواب" (the House of Representatives), a translation that accurately captures the political institution referred to in the source text. The decision to translate this term with a capital "H" is a clear example of maintaining consistency and aligning with the cultural context of the target language. This consistency highlights a shared understanding between the human translator and ChatGPT, ensuring that the translation is appropriate and accurate. As Pym [44] notes, translation decisions like these often stem from an awareness of context and the importance of maintaining fidelity to the source while adapting to the target language.

2.17 Example 16: domestication vs. foreignization

In Example 16, the human translator opts for a domestication strategy, rendering "the Senate" as "مجلس الشورى" (Consultative Council), a term that is commonly used in some Arab countries, including Saudi Arabia. This choice of terminology reflects the human translator's awareness of cultural and regional context, adapting the translation to resonate with the target audience's expectations. As Venuti [45] describes, domestication involves making the translation familiar to the target culture, which can make it more relatable and culturally relevant.

On the other hand, ChatGPT uses "مجلس الشيوخ" (the Senate), a term that is more faithful to the source text and more widely understood across various Arabic-speaking countries. This translation choice aligns with the foreignization strategy, which seeks to retain elements of the source culture in the target text. Foreignization can help preserve the original context and cultural specificity of the source, but it may also reduce accessibility for a target audience with different cultural references [45].

Given these two translation strategies, the ChatGPT translation may be considered more accurate in terms of fidelity to the source text. However, as House [29] notes, the choice of translation strategy often depends on the intended function of the text and the preferences of the target audience. If the subtitle is intended for a broader Arabic-speaking audience, ChatGPT's choice of "مجلس الشيوخ" may be preferable due to its wider recognition. However, the human translator's use of "مجلس الشورى" might be more suitable for a specific regional audience, making the translation culturally relevant. This example underscores the complexity of translation choices and the need for context-sensitive decisions in subtitle translation.

2.18 Example 17: shortness and accuracy in subtitles

In Example 17, both the human and ChatGPT translations are equally accurate, with both translations effectively conveying the meaning of the source text. However, the human translation is shorter, which is an important aspect of subtitling. As mentioned earlier, conciseness is key in subtitling, as it ensures that viewers can read the subtitles quickly and efficiently. Studies have shown that shorter subtitles improve readability and viewer engagement, as they fit within the limited reading time available during each scene (Gottlieb, 2001). The importance of this principle is underscored by the human translator's decision to keep the translation brief without sacrificing accuracy.

This highlights an important aspect of subtitling translation: achieving a balance between accuracy and brevity. While both translations are correct, the human translation is more streamlined and adheres to the general conventions of subtitle translation. This trade-off between accuracy and conciseness is often necessary in audiovisual translation, where the visual content and the speed of the dialogue must be taken into account [30].

2.19 Example 18: euphemism and language style in translation

In Example 18, both the human translation (HT) and ChatGPT translation are accurate in rendering the source text (ST), but they differ significantly in terms of fluency and tone. The HT translation uses the term "بائعة هوى" (literally "seller of desire"), which is a euphemistic expression commonly used in Arabic to refer to a prostitute. In contrast, ChatGPT uses the term "عاهرة" (a more direct and somewhat less euphemistic term for prostitute). Both terms are technically correct, but the HT's choice is more appropriate due to its milder, more culturally acceptable nature. Euphemisms, as noted by Krenn [46], are frequently employed in translation to adapt sensitive topics in a way that aligns with the cultural norms and sensibilities of the target audience. This translation strategy helps maintain a level of politeness and avoids offending the viewer, especially in contexts where explicit language might be deemed inappropriate (Table 5).

In terms of fluency, the HT translation is more elegant and elevated compared to ChatGPT's more literal approach. This is consistent with findings by Gottlieb (2001), who highlights the importance of ensuring that subtitles not only convey meaning but also reflect the stylistic conventions of the target language. The choice of language style—especially in a film or TV context—can significantly affect how a message resonates with an audience. In this instance, the HT's euphemistic style is a better fit for a subtitling context, where preserving the tone and cultural appropriateness is critical.

Regarding the second part of the subtitle about roaming the hallways, the HT translation is more accurate and sophisticated. It uses a more elevated style of language compared to ChatGPT, which does not clearly specify the hotel in question. This failure to use a clear referent for the location diminishes the clarity of the translation. As Pym [44] suggests, ambiguity in translation can hinder comprehension and lead to a less effective conveyance of meaning, particularly in subtitles, where brevity and clarity are essential.

2.20 Example 19: euphemism and fidelity to the source text

In Example 19, the HT and ChatGPT translations diverge more significantly, particularly in their cultural and linguistic approaches. The HT translation employs euphemistic expressions, which are designed to align with the cultural norms and moral sensitivities of the target audience. This approach involves strategic use of cognates and mitigated expressions to ensure that the translation is culturally appropriate. However, this strategy of domestication, as described by Venuti [45], may sacrifice some degree of fidelity to the source text. By adapting the language to fit the target culture's norms, the HT translation risks losing some of the original text's boldness and directness, which could be seen as a loss of cultural specificity.

Conversely, ChatGPT's translation prioritizes fidelity by offering a more literal rendition of the ST, preserving its raw, colloquial tone. While this approach aligns with the principles of foreignization [45], it might not be as well-suited to the target audience's cultural expectations. The use of anatomical terms related to the posterior, for example, could be seen as jarring or culturally insensitive in the TL, where such language is typically avoided in everyday conversation. However, as noted by House [29], such terms might still appear in moments of heightened emotion, such as anger or frustration, where colloquial language or even taboo expressions are used for emphasis. This reflects the tension between faithfulness to the source text and the necessity of adapting to the cultural norms of the target audience.

The HT's strategy of euphemism is supported by studies in audiovisual translation, where adapting a translation to the target culture can improve both comprehension and acceptance by viewers (Diaz Cintas & Remael, [1]). Despite this, ChatGPT's approach, by sticking closely to the source text, may appeal to audiences who value directness and authenticity, as emphasized in Pym's [44] work on translation strategies. Both translations, therefore, maintain accuracy but diverge in their approach to cultural sensitivity and tone. While ChatGPT's translation may be more faithful to the source, the HT excels in fluency, as it aligns more closely with the formal, euphemistic style commonly found in subtitle translation.

2.21 Example 20: fluency and accuracy in translation

In Example 20, both the human translation (HT) and ChatGPT translation exhibit a comparable level of accuracy, effectively capturing the essence of the source text (ST). However, the HT translation slightly outperforms the ChatGPT translation by emphasizing crucial elements of the original subtitle. As noted by Nida [38], effective translation involves balancing both accuracy and the preservation of key concepts, and the HT succeeds in doing so by focusing on the most important aspects of the ST. In subtitling, fluency is equally significant because it ensures that the translation reads

naturally and smoothly in the target language, avoiding any awkward or forced expressions. The HT achieves this fluency slightly more successfully than ChatGPT, which, while accurate, may at times struggle with rendering certain expressions in a manner that feels entirely natural in the target language (Baker, [47]) (Table 6).

This finding aligns with the arguments put forth by Baker [47], who emphasizes the importance of achieving both accuracy and fluency in translation. While accuracy ensures that the message is faithfully conveyed, fluency ensures that the translation is accessible and comfortable for the audience to consume. The HT's ability to achieve a balance between these elements makes it a more suitable choice in this instance.

2.22 Example 21: capturing cultural nuances

In Example 21, the HT translation excels in capturing the cultural nuances of the ST subtitle, ensuring both accuracy and fluency. This reflects the translator's ability to apply the concept of dynamic equivalence [38], which involves adapting the translation to ensure that the same effect is achieved in the target culture as in the source culture. The HT translation's ability to maintain fluency and naturalness while conveying cultural context is crucial in subtitling, where the audience's understanding of cultural references and idiomatic expressions must be carefully considered.

On the other hand, the ChatGPT translation adopts a literal approach that leads to issues with both accuracy and fluency. This type of direct translation can often result in expressions that are less appropriate within the context of the target culture. As noted by Newmark [32], literal translation may sometimes compromise meaning by failing to account for cultural and contextual subtleties. In this case, ChatGPT's attempt at literal translation leads to a rendition that, while accurate, does not resonate with the target audience's expectations or linguistic norms.

2.23 Example 22: fluency and language refinement

In Example 22, the ChatGPT translation demonstrates a higher level of fluency and uses standard Arabic more effectively than the HT. The ChatGPT translation is able to deliver a more polished and refined expression, showcasing its linguistic versatility and capability to generate sophisticated language outputs. This highlights the strengths of machine translation, particularly in handling more formal language registers and ensuring that the translation adheres to the syntactic and lexical conventions of the target language. ChatGPT's ability to maintain accuracy while producing standard Arabic demonstrates its potential to perform well in contexts where a refined and formal translation is required.

However, the human translation still maintains an edge in terms of contextual appropriateness, especially when dealing with more colloquial or culturally sensitive expressions. As demonstrated by Pym [44], human translators often have an advantage in interpreting contextual nuances, cultural references, and tone, making them more suited for tasks where the translation's cultural alignment is paramount.

2.24 Example 23: fluency vs. syntactic order in translation

In Example 23, the HT translation is mostly satisfactory but encounters an issue with the phrase "أداء آخر خالياً من الشوائب" (another flawless performance), where the term "الشوائب" (impurities) might not be the most commonly used expression in this context. This issue illustrates the importance of choosing words that align not only with the meaning but also with the target audience's expectations [29]. While "الشوائب" is technically accurate, a more familiar term like "خالي من العيوب" (free of flaws) may have been a better choice in this case, ensuring greater fluency and comprehension.

On the other hand, ChatGPT's translation faces syntactic issues, particularly with word order. Although it provides a more natural rendering of the term "flawless" compared to the HT, the syntactic order of the translation causes a drop in fluency. This reflects a common challenge in machine translation: while the model may excel at lexical accuracy, it can sometimes falter when handling the subtleties of syntactic structures in the target language [48]. The HT, in contrast, maintains a more accurate syntactic order and results in a higher level of fluency despite the slight lexical awkwardness of "الشوائب."

2.25 Example 24: accuracy and fluency in professional translation

In Example 24, the HT translation outperforms the ChatGPT translation in both accuracy and fluency. The HT translation is precise, with a natural flow and adherence to linguistic conventions, reflecting the human translator's higher level of proficiency. This demonstrates the advantage of human expertise in rendering complex or nuanced texts that require

both linguistic accuracy and cultural sensitivity. As Biau [49] suggests, human translators bring an invaluable level of intuition, allowing them to make informed decisions on word choice, idiomatic expression, and syntactic structure that align with the purpose of the translation.

Conversely, the ChatGPT translation, though generally accurate, lacks the fluidity and naturalness achieved by the HT. While machine translation has made significant strides in recent years, it still faces challenges in matching the subtleties of human expression, particularly when the translation involves complex or specialized content (Baker, [47]).

2.26 Example 25: idiomatic and euphemistic translation

In Example 25, the HT translation demonstrates both idiomatic and euphemistic qualities that enhance its fluency and cultural appropriateness. The use of the euphemistic phrase "الأحمق اللعين" (stupid) to replace the more vulgar *"fucking asshole!"* showcases the translator's sensitivity to the cultural norms and expectations of the target audience. Euphemism in translation, as noted by Baker [47], allows the translator to convey the meaning of the source text while adhering to social and cultural conventions that may be more tolerant of less offensive language. By choosing a less offensive expression, the HT translation avoids potential cultural discomfort while preserving the message's essence (Table 7).

In contrast, the ChatGPT translation fails to produce a meaningful or appropriate rendition, making it significantly less effective in this context. This highlights a common challenge for machine translation systems—handling emotionally charged or taboo language. According to Pym [44], machine translation systems often struggle with slang, idiomatic expressions, or taboo language due to a lack of cultural and contextual sensitivity. The HT translation, by contrast, is both accurate and culturally sensitive, demonstrating the advantage of human expertise in managing such delicate translation issues.

2.27 Example 26: idiomatic accuracy and cultural references

In Example 26, the HT translation excels in capturing the idiomatic expression *'Blue,'* which refers to the Democratic Party. This type of idiomatic translation is an area where human translators typically outperform machine translation systems, as they are better able to recognize cultural and political references. As House [29] emphasizes, dynamic equivalence, which requires adaptation of idiomatic and culturally-specific terms, is a key component of effective translation. The HT correctly identified the political reference and provided an appropriate translation, maintaining both accuracy and contextual relevance.

On the other hand, the ChatGPT translation falls short by failing to recognize the idiomatic nature of *'Blue'* and offering an inappropriate rendering that doesn't capture the intended meaning. This demonstrates one of the primary challenges faced by machine translation, especially in dealing with terms that rely heavily on cultural context. As Baker [47] notes, idiomatic expressions require a deep understanding of both the source and target cultures, which machine translation systems often lack. The HT translation, which successfully navigates this cultural reference, highlights the human translator's ability to manage such nuances effectively.

2.28 Example 27: challenges with idiomatic translation

Example 27 illustrates a similar issue, where ChatGPT struggles with translating a subtitle that contains an idiomatic expression. This further emphasizes the limitations of machine translation in accurately capturing the meaning of nuanced, culture-specific expressions. The inability to properly translate idioms can lead to loss of meaning or awkward phrasing, which could undermine the translation's fluency and effectiveness. According to Newmark [32], idiomatic expressions are often untranslatable through literal methods, and translating them requires cultural adaptation, which human translators are typically more adept at handling than machine systems.

2.29 Example 28: literal translation and its limitations

In Example 28, ChatGPT's reliance on a literal translation again results in a failure to accurately convey the intended meaning of the ST subtitle. This points to a recurring limitation of machine translation, where literal renditions often miss the nuances of meaning and tone that human translators are better equipped to interpret. As noted by Bowker and Pearson [48], machine translations frequently rely on word-for-word translations, which can work for straightforward texts but falter when dealing with complex or idiomatic expressions. In this case, the lack of contextual sensitivity in the

ChatGPT translation makes it less effective than the HT translation, which likely considered both the surface meaning and the deeper contextual implications.

2.30 Conclusion: idiomatic and cultural sensitivity in translation

These examples illustrate the ongoing challenges of machine translation, particularly when it comes to idiomatic expressions and culturally specific references. While machine translation systems like ChatGPT have made significant strides, they still face limitations in providing idiomatic and contextually sensitive translations. Human translators, on the other hand, can draw on their understanding of cultural nuances, idiomatic usage, and the subtleties of language to produce more accurate and culturally appropriate translations.

Both Baker [47] and Pym [44] emphasize that while machine translation is useful for many purposes, it often struggles with translating content that requires cultural sensitivity, emotional nuance, and idiomatic understanding. The HT's success in navigating these challenges underscores the importance of human expertise in translation, especially when dealing with complex, nuanced, or culturally sensitive content.

The analysis of the translation examples highlights the complex interplay between accuracy, fluency, and cultural sensitivity in both human and machine-generated translations. While both human and machine translations often provide a close approximation of the source text's meaning, human translators consistently outperform machine systems in conveying the nuanced and culturally-specific aspects of the original content. Through the careful handling of idiomatic expressions, cultural references, and euphemistic language, human translators can produce translations that are both contextually appropriate and idiomatically fluent. On the other hand, machine translation systems like ChatGPT, despite their technological advancements, still struggle with idiomatic and culturally sensitive elements, often resulting in translations that lack the required depth and fluency. This underscores the enduring value of human expertise in translation, particularly when dealing with complex, nuanced texts where cultural context and emotional tone are crucial. As machine translation continues to evolve, its ability to handle these subtleties will improve, but for now, human translators remain essential for achieving the highest levels of accuracy and fluency in translation.

2.31 Quantitative results

To assess the fluency of human translators in comparison to ChatGPT's translation, a paired samples T-test was conducted. The results are presented below.

3 Fluency

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Fluency (Chatgpt)	3.88	32	1.314	0.232
	Fluency (Human)	4.44	32	0.801	0.142

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Fluency (Chatgpt) – Fluency (human)	−0.563	1.162	0.205	−0.982	−0.143	−2.738	31	0.010

The paired samples T-test revealed a statistically significant difference in fluency between ChatGPT and human translators

(t = -2.738, df = 31, p = 0.010). The mean difference was -0.563, suggesting that, on average, ChatGPT exhibited a slightly lower fluency compared to human translators. The 95% confidence interval ranged from -0.982 to -0.143.

4 Accuracy

To assess the accuracy of human translators in comparison to ChatGPT's translation, a paired samples T-test was conducted. The results are presented below.

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Accuracy (ChatGpt)	3.84	32	1.273	0.225
	Accuracy (Human)	4.41	32	0.756	0.134

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Accuracy (ChatGpt) – Accuracy (Human)	−0.563	1.162	0.205	−0.982	−0.143	−2.738	31	0.010

The paired samples T-test indicates a statistically significant difference in accuracy between ChatGPT and human translators (t = −2.738, df = 31, p = 0.010). The mean difference was -0.563, signifying that, on average, ChatGPT exhibited a slightly lower accuracy compared to human translators. The 95% confidence interval ranged from −0.982 to −0.143.

In summary, both qualitative and quantitative findings indicate that human translators outperformed ChatGPT in terms of accuracy and fluency. Despite this, the results suggest that ChatGPT can be a valuable tool for efficiently translating movie subtitles, potentially expediting the overall translation process. It is alos noticed that in spite of the significant difference between ChatGPT translations and HT translations, the difference is slight. This suggests that ChatGPT is a very useful in subtitling. However, it is crucial to note that the output generated by ChatGPT should be carefully reviewed and revised, as it demonstrated limitations in accurately handling cultural references. Specifically, ChatGPT struggled to provide contextually appropriate translations for cultural nuances.

4.1 Limitations

Sample Size and Scope: The study focused on a limited number of translation examples, which may not fully capture the diversity of translation challenges across different genres or contexts. A larger, more varied sample would provide a broader perspective on the strengths and weaknesses of both human and machine translations.

Language Pair: The study was conducted on translations between English and Arabic. The findings may not be directly applicable to other language pairs, as different languages present unique syntactic, semantic, and cultural challenges. Future studies could expand to include a wider variety of languages.

Technological Constraints: While ChatGPT represents an advanced AI-based translation tool, its performance is influenced by the underlying technology, which may not always accurately capture the nuances of certain idiomatic expressions, slang, or culturally sensitive content. The results may not fully reflect the capabilities of future AI models, as translation technology is rapidly evolving.

Subjectivity in Evaluation: While the study attempts to provide an objective evaluation of the translations, the assessment of fluency and accuracy remains somewhat subjective. Different evaluators may interpret and rate translations differently, leading to potential biases in the results. Future research could involve multiple evaluators to enhance the objectivity and reliability of the findings.

4.2 Recommendations

Expansion of Data Set: Future research should involve a larger and more diverse set of translation examples, including various genres (e.g., legal, technical, literary) and contexts. This would provide a more comprehensive understanding of how human and machine translations perform in different scenarios.
Incorporating Human Feedback into AI Models: Given the limitations of machine translation in handling idiomatic and culturally sensitive content, it is recommended that AI systems like ChatGPT be further trained with human feedback. This could improve the system's ability to produce more accurate and contextually appropriate translations, especially for non-literal or culturally nuanced expressions.
Cross-Linguistic Studies: It would be valuable to extend this study to include other language pairs, particularly those with distinct linguistic structures or cultural contexts. This would allow for a broader comparison of the translation capabilities of human versus machine systems across different languages.
Development of Specialized Translation Systems: Future AI models could benefit from being trained on specific domains (e.g., medical, legal, literary) to improve their performance in these specialized areas. This approach would help AI systems produce more accurate and contextually relevant translations in domains that require precision and domain-specific knowledge.

4.3 Pedagogical implications

AI in Translation Education: The findings highlight the growing role of AI in language learning and translation education. Pedagogically, this suggests that AI tools like ChatGPT could serve as valuable aids in teaching translation, providing students with an initial draft that they can analyze and refine. However, educators should emphasize the importance of critical thinking and cultural awareness when evaluating machine-generated translations, as AI systems are still limited in handling idiomatic and culturally sensitive content.
Enhancing Cultural Awareness: Given that AI systems may struggle with cultural nuances, translation programs should incorporate more in-depth lessons on cultural sensitivity and contextual factors. Students should be taught not only the mechanics of translation but also how to navigate the cultural and emotional underpinnings of the source and target languages.
Balancing Human and Machine Contributions: While AI tools can be useful in translation workflows, educators should emphasize that human expertise is irreplaceable, particularly when it comes to conveying subtle meanings, humor, and cultural references. Pedagogical frameworks could integrate both human and machine translation work, encouraging students to use AI as a starting point but to rely on their own linguistic and cultural knowledge for refinement.
Critical Evaluation Skills: The study underscores the need for students to develop strong evaluation skills when working with machine translations. Students should be trained to critically assess machine-generated texts, identifying areas where AI may fall short in terms of accuracy, fluency, and cultural appropriateness. This will empower students to produce higher-quality translations, even when utilizing AI tools.

4.4 Future AI advancements and the potential role of post-editing in improving subtitle quality

The landscape of machine translation (MT) has been rapidly evolving, with AI-driven systems like GPT-3.5 and its successors, including GPT-4, bringing substantial advancements in translation quality. As these AI systems continue to evolve, their role in subtitling, especially in complex fields such as political satire and cultural content, will expand. However, the need for post-editing to ensure the highest quality in subtitles will likely remain significant due to several inherent limitations of current AI systems.

5 Advancements in AI: GPT-4 and Beyond

GPT-4, the latest iteration of OpenAI's language model, introduces several key advancements over its predecessor, GPT-3.5. These include:

- Improved Contextual Understanding: GPT-4 boasts enhanced contextual awareness, which enables it to better understand nuanced content, such as humor, irony, and complex political themes, that are often challenging for

previous AI models. This improved understanding can lead to more accurate and contextually appropriate translations in subtitling tasks.
- Multilingual Proficiency: GPT-4 is trained on an even broader range of languages, improving its ability to handle non-English languages, including Arabic, with greater fluency. This makes it an appealing tool for subtitling in various languages, especially those with complex syntax and cultural context like Arabic.
- More Natural Output: With improvements in natural language generation, GPT-4 produces text that is more coherent, fluent, and contextually appropriate. For subtitling, this means fewer awkward phrases, better flow of language, and greater alignment with the visual context, potentially enhancing viewer comprehension and enjoyment.

6 Limitations of GPT-4 and the Need for Post-Editing

Despite the advancements in GPT-4, AI-generated translations still face several challenges that necessitate post-editing:

- Cultural Sensitivity: AI models like GPT-4 still struggle with translating culturally bound terms, idioms, and references, which are critical in subtitling, especially for content with rich political or cultural satire. While GPT-4 may produce more accurate translations, it may still miss subtle cultural nuances or misinterpret culturally-specific humor, requiring human post-editing to ensure cultural appropriateness.
- Humor and Sarcasm: Political satire, humor, and sarcasm are particularly difficult for AI to handle accurately, even with the advancements of GPT-4. AI may fail to fully capture the tone or intent behind certain jokes, leading to translations that lack the same humor or may even confuse the audience. Human translators can intervene during the post-editing process to refine these elements and ensure the intended meaning is conveyed.
- Consistency in Terminology: AI can sometimes produce inconsistencies in terminology, especially with long-form content like series and movies. For instance, a character's name or a recurring phrase might be translated inconsistently throughout a series, which may confuse viewers. Post-editing can resolve these issues by ensuring consistent terminology and phrasing across all episodes or scenes.
- Technical Quality: While GPT-4 may produce grammatically correct translations, issues related to sentence structure, readability, and pacing in subtitled content may still arise. For example, AI-generated translations may be too lengthy or too short to fit the allotted time for subtitles, which could affect viewer experience. Post-editing can help adjust the translations to match the required length and ensure that subtitles flow in sync with the dialogue.

7 The role of post-editing

Post-editing will continue to play a critical role in improving subtitle quality, even with the advancements of AI technologies like GPT-4. Post-editing will primarily focus on the following areas:

- Ensuring Cultural Accuracy: Human editors can identify and address cultural discrepancies that AI may overlook. By refining translations of idiomatic expressions, cultural references, and humor, post-editing ensures that subtitles resonate with the target audience and align with the source material's intent.
- Enhancing Contextual Relevance: AI models, even advanced ones like GPT-4, may still misinterpret specific contextual cues, particularly in content such as political satire or humor. Post-editing allows human translators to make adjustments to better capture the tone, intent, and cultural undertones of the content, improving the overall viewer experience.
- Improving Readability: AI-generated translations may sometimes result in subtitles that are difficult to read or that do not align with the pacing of the original content. Post-editing allows editors to optimize sentence structure, word choice, and subtitle timing to ensure readability and synchronization with the visuals.
- Maintaining Consistency: Human editors are crucial for ensuring that key terminology, names, and expressions remain consistent throughout the translation. This is especially important in series like *Our Cartoon President*, where recurring characters and phrases need to be accurately and consistently translated across multiple episodes.

8 AI as a tool, not a replacement

While GPT-4 and future AI advancements hold great promise in automating subtitling tasks and improving efficiency, they should be viewed as tools to assist human translators rather than replacements. The future of subtitling lies in a hybrid model where AI generates initial translations, and human post-editors refine them for cultural and contextual accuracy.

In the long term, advancements in AI, particularly those focused on enhancing its understanding of cultural context and humor, could reduce the need for extensive post-editing. However, for the foreseeable future, human expertise will remain indispensable in ensuring that subtitles are both accurate and culturally appropriate, enhancing the viewing experience for audiences worldwide.

8.1 Actionable solutions for integrating AI into movie subtitling workflows

AI-driven solutions can significantly enhance the efficiency and accuracy of movie subtitling while maintaining linguistic and cultural nuances. Below are key strategies for integrating AI into movie subtitling workflows:

9 AI-assisted transcription & time-coding

- Use AI-powered speech recognition tools (e.g., Whisper, Rev.ai, Aegisub) to generate initial transcriptions.
- Implement AI-driven time-coding to automatically segment subtitles based on speech patterns and scene changes.
- Use voice activity detection (VAD) to refine subtitle synchronization with dialogue pacing.

10 AI-enhanced subtitle translation & localization

- Utilize Neural Machine Translation (NMT) (e.g., DeepL, Google Translate, ModernMT) for initial translations.
- Train AI models on movie-specific corpora to enhance accuracy in translating idioms, slang, and cultural references.
- Integrate translation memory (TM) and glossaries to ensure consistency in names, recurring phrases, and franchise-specific terminology.

11 AI-driven subtitle adaptation & readability

- Implement AI-based tools to adjust subtitle length and reading speed according to industry standards (e.g., Netflix TT, BBC guidelines).
- Use AI to identify and split long sentences while maintaining natural readability.
- Leverage AI to suggest line breaks and prevent awkward phrasing in translated subtitles.

12 AI-assisted quality control & post-editing

- Employ AI-powered spell-checkers and grammar correction tools (e.g., Grammarly, LanguageTool) to refine accuracy.
- Use AI-driven style checkers to ensure consistency in tone, register, and formatting.
- Implement AI-based error detection models to flag mistranslations, missing subtitles, or incorrect timecodes.

13 Automation of compliance with industry standards

- Configure AI to automatically apply subtitle constraints (e.g., characters per line, reading speed limits).
- Ensure AI-generated subtitles comply with accessibility requirements (e.g., SDH for hearing-impaired audiences).
- Use AI-driven compliance tools to adhere to platform-specific standards (e.g., Netflix Timed Text Style Guide).

14 AI for contextual & cultural adaptation

- Train AI models to recognize cultural nuances, idioms, and humor, preventing literal translations that may confuse audiences.
- Use context-aware AI tools to refine subtitle choices based on genre-specific expectations (e.g., comedy, drama, sci-fi).
- Implement AI-assisted sentiment analysis to detect and preserve emotional tone in translations.

15 Ethical considerations & human oversight

- Maintain human involvement in reviewing and refining AI-generated subtitles, especially for culturally sensitive content.
- Use on-premise AI solutions to ensure compliance with copyright and data privacy laws.
- Develop AI models that minimize bias in translation, particularly in films featuring multilingual or dialect-heavy dialogue.

By implementing these AI-driven solutions, movie subtitling can become more efficient, cost-effective, and high-quality, while still preserving artistic and cultural integrity. Let me know if you need more tailored suggestions!

16 Conclusion

This research offers a comprehensive investigation into English-to-Arabic translations, employing a comparative analysis between human translators and ChatGPT 3.5. The study utilized a corpus comprising episodes from the "Our Cartoon President" series and their corresponding Arabic translations, aligning source texts with target translations. Thirty-two randomly selected instances were subjected to comparison between ChatGPT 3.5 and human translations obtained from the Subscene website, where the latter were crafted by professional translators. To assess the quality of translations, a survey was designed and presented to English-Arabic translation professors. These experts were tasked with rating the fluency and accuracy of translations on a scale ranging from 1 to 5. The collected data underwent analysis using SPSS, including a T-test to ascertain statistically significant differences between human and ChatGPT translations. The research adopts a mixed-methods approach, integrating qualitative insights derived from the evaluations provided by translation professors, alongside quantitative data gathered from their ratings on fluency and accuracy.

The study's outcomes provide valuable insights into the comparative effectiveness of human and ChatGPT translations in the realm of English-to-Arabic translation. The emphasis is on discerning potential applications and identifying areas for improvement in machine translation systems. Notably, the research underscores a limitation of ChatGPT, revealing challenges in translating cultural references. This nuanced understanding contributes to delineating the strengths and weaknesses of ChatGPT in various translation scenarios.

In summary, the HT translation consistently outperforms the ChatGPT counterpart in terms of both fluency and accuracy, particularly when dealing with cultural references and sensitive expressions such as swearing and taboo words. The HT demonstrates a nuanced understanding of idiomatic language, successfully preserving cultural connotations and delivering more contextually appropriate translations. This reinforces the notion that while ChatGPT may provide literal translations, it often struggles to capture the subtleties and cultural nuances present in the source text, emphasizing the proficiency of human translators in handling complex linguistic and cultural elements. This comparison underscores the notable disparities between HT and ChatGPT, emphasizing that AI, represented by ChatGPT, is not poised to replace human translators in the near future. The nuanced understanding, cultural sensitivity, and contextual appropriateness demonstrated by the HT highlight the complex and inherently human aspects of translation that AI systems currently struggle to replicate effectively. The importance of cultural references and the handling of sensitive language reaffirm the indispensable role of human translators in navigating the intricacies of language and culture.

Author contributions It is written by one author only.

Funding Not applicable

Data Availability No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate The need for ethical approval was waived off by the ethical committee of October University for Modern Sciences and Arts because of the nature of the study.

Informed consent Informed consent was obtained from all individual participants included in the study.

Competing interests The authors declare no competing interests.

References

1. Díaz Cintas J, Remael A. Audiovisual Translation: Subtitling. St. Jerome Publishing. (2007).
2. Hendy A, Al-Moslmi T, Aloraini N, Bouamor D. How good are GPT models at machine translation? A comprehensive evaluation. (2023). https://arxiv.org/abs/2302.09210.
3. Tan L, Xing D, Chang CH, Li H, Li Y, Ren B. Changes in genome architecture and transcriptional dynamics progress independently of sensory experience during postnatal brain development. Cell, (2020);180(4):729–744.e17. https://doi.org/10.1016/j.cell.2020.01.015
4. Banat M, Abu Adla Y. Exploring the effectiveness of GPT-3 in translating specialized religious text from Arabic to English: A comparative study with human translation. Theor Pract Lang Stud. 2023;13(3):786–793. https://doi.org/10.17507/tpls.1303.23.
5. Zhou, L., Hu, W., Zhang, J., & Zong, C. (2017). Neural system combination for machine translation. ArXiv preprint. https://arxiv.org/abs/1704.06393
6. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Polosukhin I. Attention is all you need. Adv Neural Inf Process Syst. 2017;30:1.
7. Jiao W, Zeng X, Wang Y, Wang H, Liu Q. ParroT: Translating during chat using large language models tuned with human translation and feedback. In Findings of the Association for Computational Linguistics: EMNLP 2023 (pp. 15009–15020). https://aclanthology.org/2023.findings-emnlp.1001/
8. Liu, X., Zheng, Y., Du, Z., Ding, M., Qian, Y., Yang, Z., & Tang, J. (2021). GPT understands, too. ArXiv preprint arXiv:2103.10385. https://arxiv.org/abs/2103.10385
9. Shyam P, Santoro A, Singh S, Pineau J. Meta-learning with implicit gradients. In Advances in Neural Information Processing Systems (2020);33:113–123. https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf
10. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, Amodei D. Language models are few-shot learners. (2020). preprint arxiv:2005.14165
11. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. Language models are unsupervised multitask learners. OpenAI. 2019
12. Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., & Zettlemoyer, L. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020; 7871–7880.
13. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol. 2006;3(2):77–101. https://doi.org/10.1191/1478088706qp063oa.
14. Deng L, Huang X. Challenges in adopting speech recognition. Commun ACM. 2004;47(1):69–75. https://doi.org/10.1145/962081.962108.
15. O'Shaughnessy D. Invited paper: automatic speech recognition: History, methods, and challenges. Pattern Recogn. 2008;41(10):2965–79. https://doi.org/10.1016/j.patcog.2008.05.008.
16. Greis, M., Joshi, A., Singer, K., Schmidt, A., & Machulla, T. (2018). Uncertainty visualization influences how humans aggregate discrepant information. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (pp. 1–12). https://doi.org/10.1145/3173574.3174079
17. Holt BG, Lessard JP, Borregaard MK, Fritz SA, Araújo MB, Dimitrov D, Rahbek C. An update of Wallace's zoogeographic regions of the world. Science (2012); 339(6115):74–78. https://doi.org/10.1126/science.1228282
18. Greis, M., Hullman, J., Correll, M., Kay, M., & Shaer, O. (2017). Designing for uncertainty in HCI: When does uncertainty help? In Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (pp. 593–600). https://doi.org/10.1145/3027063.3027091
19. Mohamed A. Exploring the potential of an AI-based chatbot (ChatGPT) in enhancing English as a foreign language (EFL) teaching: perceptions of EFL faculty members. Educ Inf Technol. 2024. https://doi.org/10.1007/s10639-023-11917-z.

20. Mohamed A, et al. Empowering the faculty of education students: applying AI's potential for motivating and enhancing learning. Innov Higher Educ. 2024. https://doi.org/10.1007/s10755-024-09747-z.
21. Abdelaal, N., & Al Sawi, M. (2024). Perceptions of the use of AI in education: Challenges and implications. Asian Journal of Applied Linguistics. https://www.castledown.com/journals/ajal/article/view/1309
22. Díaz-Cintas J, Remael A. Audiovisual translation: subtitling. England: Routledge; 2007.
23. Abdelaal NM. Subtitling of culture-bound terms: strategies and quality assessment. Heliyon. 2019;5(4): e01318. https://doi.org/10.1016/j.heliyon.2019.e01318.
24. Abdelaal NM, Al Sarhani A. Subtitling strategies of swear words and taboo expressions in the movie *Training Day*. Heliyon. 2021;7(7): e07351. https://doi.org/10.1016/j.heliyon.2021.e07351.
25. Fink A. How to conduct surveys: a step-by-step guide. 5th ed. Thousand Oaks: SAGE Publications; 2013.
26. Creswell JW. Research design: qualitative, quantitative, and mixed methods approaches. 4th ed. Thousand Oaks: SAGE Publications; 2014.
27. Blaikie N. Designing social research: a guide for the confused. 2nd ed. Cambridge: Polity Press; 2010.
28. Sargeant J, Loney P, Chin R. Assessing the accuracy of survey responses: a comparison of two techniques for reducing bias. Health Serv Res. 2011;46(1):324–32. https://doi.org/10.1111/j.1475-6773.2010.01193.x.
29. House J. Translation as communication across languages and cultures. England: Routledge; 2015.
30. Baker M. In other words: a coursebook on translation. 3rd ed. England: Routledge; 2018.
31. Toury G. Descriptive translation studies and beyond. Amsterdam: John Benjamins; 1995.
32. Newmark P. A textbook of translation. Prentice Hall; 1988.
33. Hatim B, Mason I. The Translator as Communicator. Routledge. (1997).
34. Reiss, K., & Vermeer, H. J. Grundlegung einer allgemeinen Translationstheorie. Niemeyer. 1984
35. Koehn P. Neural machine translation. Cambridge: Cambridge University Press; 2020.
36. Dorr BJ, Garman C, Monz C. Machine translation evaluation and optimization. Springer; 2010.
37. Munday J. Introducing translation studies: theories and applications. 4th ed. England: Routledge; 2016.
38. Nida, E. A. Toward a science of translating: With special reference to principles and procedures involved in Bible translating. E.J. Brill. 1964
39. Toral A, Way A. Attaining the unattainable? Reassessing claims of human parity in neural machine translation. In Proceedings of the Third Conference on Machine Translation: Research Papers (2018), (pp. 113–123). https://aclanthology.org/W18-6312/
40. Katan D. Translation as intercultural communication: Selected papers from the EST Congress, 2007. England: Routledge; 2014.
41. Klaudy K. The role of explicitation in translation. In: Károly K, editor. The translator's handbook. Routledge; 2009. p. 183–204.
42. Gile D. Basic concepts and models for interpreter and translator training. Amsterdam: John Benjamins; 2009.
43. Beaugrande R, de Dressler W. Introduction to Text Linguistics. Longman (1981).
44. Pym A. Exploring translation theories. England: Routledge; 2010.
45. Venuti L. The translator's invisibility: a history of translation. England: Routledge; 1995.
46. Krenn, K. Euphemism and translation: An investigation into the role of euphemistic strategies in subtitles. University of Vienna. 2016
47. Baker, M. In Other Words: A Coursebook on Translation (2nd ed.). Routledge. (2011).
48. Bowker L, Pearson J. Working with specialized language: a practical guide to using corpora. England: Routledge; 2002.
49. Biau A. The art of translation in context: theory and practice. Cambridge: Cambridge Scholars Publishing; 2017.