Large Language Models as Computational Linguistics Tools: A Comparative Analysis of ChatGPT and Google Machine Translations

Mohammad Awad AlAfnan

Liberal Arts Department, American University of the Middle East, Kuwait

Corresponding Author: Mohammad Awad AlAfnan, Email: mohammad.al-afnan@aum.edu.kw

Abstract

This study investigates the effectiveness of Large Language Models (LLMs), specifically ChatGPT, in machine translation and compares them with traditional tools like Google Translate. The research focused on translating speeches by King Abdullah II of Jordan, delivered in Arabic and English at significant international events in 2023. The study evaluated the translations based on meaning, functional and textual adequacy, target language mechanics, style, register, and idiomaticity. The analysis revealed that Google Translate's Arabic-English translations were deficient, with contextual accuracy and meaning issues necessitating major revisions. The English-Arabic translations by Google Translate also required significant edits due to literal translation practices and inadequacies in several areas. Contrariwise, ChatGPT's Arabic-English translations were rated as acceptable, needing only minor edits, and offered more natural-sounding translations. The English-Arabic translations by ChatGPT, while better than Google Translate, still showed some deficiencies but were deemed acceptable with minor adjustments. The study underscores the irreplaceability of human translators in ensuring accurate and contextually rich translations. However, the study also highlights the potential of LLMs, such as ChatGPT, to significantly enhance the translation process. Developers are encouraged to enhance LLMs' contextual understanding and natural language processing capabilities. This involves expanding training datasets to include diverse and context-rich examples and improving the models' ability to handle different registers, styles, and idiomatic expressions. The study strongly advocates for a collaborative model in the translation industry that integrates machine translation with human expertise to enhance efficiency while maintaining quality. These insights are crucial for driving advancements in developing and applying machine translation tools, ensuring they complement rather than replace human translators.

Keywords: large language models; machine translation; ChatGPT; google translate; computational linguistics

Introduction

Computational linguistics emerged due to the explosion of human data on the internet and the rapid advancement of computational power. These advancements enabled researchers to analyze language and behavior on a large scale with detailed observations. Initially, statistical text analysis dating back to content dictionaries facilitated computational linguistics research by structuring

non-numeric data [1-3]. The advent of large language models (LLMs) may revolutionize computational linguistics by offering these capabilities without the need for custom training data [4]

This study aims to evaluate the extent to which LLMs can reshape computational linguistics. Robust computational techniques are crucial for analyzing textual data and understanding various social phenomena across disciplines. Current computational linguistics methodologies often rely on supervised text classification and generation to extend manual labeling efforts to new texts, a process known as coding. However, dependable supervised methods typically require substantial amounts of human-annotated training data. On the other hand, unsupervised methods can be executed without such data but may produce less interpretable results. Presently, the availability of data resources limits the theories and subjects computational linguistics can explore.

LLMs have the potential to eliminate these limitations. Recent LLMs have showcased remarkable capabilities in accurately classifying text, summarizing documents, answering questions, and generating understandable explanations across various domains, sometimes surpassing human performance without supervision. If LLMs can similarly offer reliable labels and summary codes through zero-shot prompting, computational linguistics research can expand beyond the limitations of available tools and data resources [5-7]. To effectively utilize LLMs, behavioral researchers need to comprehend the advantages and disadvantages of different modeling decisions (model selection), as well as how these decisions intersect with their specific fields of expertise (domain utility) and intended applications (functionality). By assessing LLMs across a wide range of computational linguistics tasks, this study provides insights into the following research questions:

Can LLMs be used as computational linguistics tools?

Did MT tools develop to overcome the weaknesses listed in previous research?

Do LLMs provide a better alternative than the traditional MT tools?

This exploratory study compares Arabic < > English human translation, machine translation, and LLMs translation to answer these research questions.

What are Large Language Models?

LLMs burst onto the scene in late 2022 and early 2023 and fascinated academic researchers and the general public. Models like ChatGPT and GPT-4 are highly skilled at holding natural conversations on various topics, leading even cautious research teams to suggest they display hints of artificial general intelligence [8]. Consequently, there is considerable interest in understanding the capabilities and constraints of these models and how they might reshape society.

LLMs have been under extensive examination even before their recent surge in popularity, with several years dedicated to active research to assess how their abilities compare with those of humans carefully. This literature encompasses various task families, focusing on linguistic abilities and others targeting commonsense knowledge and logical reasoning. Recent LLM advancements have demonstrated remarkable zero-shot capabilities across text-generation tasks based on natural language instructions [9]. Nevertheless, there is high user expectation for text rewriting, and any unintended edits by the model can diminish user satisfaction.

Recent research has shown that LLMs excel in various natural language tasks, including automatic summarization (creating a condensed version of the text), machine translation (translating text between languages), and question answering (developing systems that answer questions based on the text) [10, 11]. LLMs have succeeded in these tasks due to two key factors. Firstly, LLMs are built upon the transformer [12], a cutting-edge neural network architecture with many parameters. The key innovation of the transformer lies in its self-attention mechanisms, which enable the model to comprehend better the relationships between different elements of the input [13].

Secondly, LLMs employ a two-stage training process to learn from data efficiently. In the initial pretraining stage, LLMs utilize a self-supervised learning approach, allowing them to learn from vast amounts of unannotated data without manual annotation. This capability provides a significant advantage over traditional fully supervised deep learning models, as it eliminates the need for extensive manual annotation and enhances scalability. In the subsequent fine-tuning stage, LLMs are trained on small, task-specific, annotated datasets to leverage the knowledge acquired during the pretraining stage to perform specific tasks intended by end users. Consequently, LLMs achieve high accuracy on various tasks with minimal human-provided labels [13]

Despite their strengths, LLMs, in general, and ChatGPT, as a specific example, have several limitations. One drawback is that they may generate plausible but incorrect responses, such as inventing terms it should be familiar with. This phenomenon, known as the hallucination effect, is common in many natural language processing models [14-15]. ChatGPT often follows instructions rather than engaging in genuine interaction. For example, when users provide insufficient information, ChatGPT tends to make assumptions about what the user wants to hear rather than asking clarifying questions [13].

What is Computational Linguistics?

Computational linguistics focuses on understanding written and spoken language from a computational standpoint and developing tools and systems that can effectively process and generate language in large-scale applications or interactive dialogues. By exploring language computationally, we gain insights into human thinking and intelligence, as language reflects the mind. Additionally, given that language is the most natural and flexible mode of communication for humans, computers with linguistic capabilities could significantly enhance human-computer interaction.

The theoretical objectives of computational linguistics include creating grammatical and semantic frameworks to describe languages in ways that facilitate computational analysis of syntax and meaning. This field's theoretical and practical research draws upon various disciplines, such as theoretical linguistics, philosophical logic, cognitive science (particularly psycholinguistics), and computer science. However, early work from the mid-1950s to around 1970 prioritized practical applications like Machine Translation (MT) and simple Question Answering (QA) rather than theoretical considerations. In MT, key concerns revolved around lexical structures, domain-specific sublanguages (e.g., weather reports), and language translation processes using graph transformation or transfer grammars. For QA, the focus was on understanding question patterns within specific domains and how these patterns related to the formats in which answers might be stored, such as in relational databases.

Computational sociolinguists employ computational techniques to gauge the interplay between society and language, encompassing the stylistic and structural attributes that differentiate speakers [16]. Language variation is intricately tied to social identity [17], spanning group affiliation [18], geographic location [19], and socioeconomic status [20-21], as well as individual characteristics such as age and gender [22].

The examination of themes, settings [23], and narratives [24-25] is fundamental in literary studies [26], with themes analyzed by methods such as topic modeling. In contrast, settings are often identified using named entity recognition and toponym resolution. These techniques, already addressed by models like GPT 4 Turbo, offer solutions. Our approach concentrates on narrative analysis through natural language processing (NLP), parsing narratives into chains involving agents, their relationships, and the events they partake in. Our focus extends to social role labeling and event extraction, as well as studying agents in terms of power dynamics and emotions. We also delve into figurative language and humor classification, evaluating these aspects within our study of literary devices.

What is Machine Translation?

As mentioned earlier, computational linguistics looked into theoretical linguistics and applied applications. One of the applications studied in computational linguistics is machine translation. Machine Translation (MT) was one of the first applications that computers were thought to be able to solve. As a result, various approaches were developed to address the MT problem, the most well-known being statistical machine translation (SMT), where much work was done on creating parallel datasets (also known as bitext) and researching new MT techniques.

A breakthrough with encouraging results was made in 2013 with the introduction of end-to-end neural encoder-decoder-based machine translation (MT) systems, quickly gaining popularity as neural machine translation (NMT). Currently, NMT is the most widely used method in the community. However, it was not long before it was understood that these early NMT systems needed enormous amounts of parallel data to provide results that were on par with SMT [27]. Dataset size is not an issue for high-resource language pairings (e.g., English and French) because academics have produced several parallel corpora. However, for many of the more than 7,000

languages in use globally, the need to have vast volumes of parallel data is not a feasible assumption. For low-resource languages (LRLs), it is consequently seen as a significant difficulty [27]. It is helpful to automatically translate between most of these LRLs for social and economic reasons, primarily for nations with several official languages. Consequently, there has been a discernible upsurge in NMT research on LRL pairings conducted by academia and business in recent years.

According to [28], MT refers to computer-based activities related to translation. More specifically, [29] states that computer-aided translation can involve both human-aided MT and machine-aided human translation. However, MT focuses on automating the entire translation process and is associated with computerized systems that produce translations, excluding "computer-based translation tools which support translators by providing access to online dictionaries, remote terminology databanks, transmission and reception of texts, etc." (p. 431). MT has evolved from its inception right after the Second World War, utilizing various approaches [28]. Neural MT has emerged as a popular method based on deep learning technology and large artificial neural networks with powerful algorithms [30-37].

Digitalization and globalization, along with advancements in computational linguistics and the availability of Machine Translation (MT) tools like Babylon, DeepL, Google Translate, Microsoft Translator, Systran, and Yandex Translate, have made it possible to translate a wide range of text types into different languages [38]. While translated texts often achieve a proficiency level of B2, according to the Common European Framework of Reference for Languages [36], MT still has limitations that prevent it from reaching similar quality standards to human-mediated translation processes [31, 38-40]. Besides linguistic constraints, MT also suffers from sociolinguistic and pragmatic inadequacies [41, 42]. Another weakness of MT is the requirement for large parallel datasets, many of which are limited to specific domains and languages [43, 44].

MT is widely used in commerce, tourism, and education [37, 45]. In educational settings for foreign language learning, there are mixed attitudes among instructors and learners toward automated translation [46, 47]. Despite the everyday use of technological devices and internet access in multimodal learning environments, MT is sometimes restricted or prohibited [48]. Concerns have been raised regarding MT's ethicality and accuracy [46]. However, recent research suggests correcting mistakes in automatically translated texts can enhance second language acquisition in advanced learners and improve their translation skills [34, 36, 49]. Integrating MT into the learning process requires critical reflection [47] and should also involve pre-editing source texts [50].

These mixed reactions towards MT make it worth investigating, especially during the recent developments in artificial intelligence, to answer the following research questions: Can LLMs be used as computational linguistics tools? Did MT tools develop to overcome the weaknesses listed in previous research? Do LLMs provide a better alternative than the traditional existing MT tools?

Methods

This paper investigates the possibilities of using LLMs as computational linguistics and machine translation tools. Specifically, it strives to answer the following research questions.

Can LLMs be used as computational linguistics tools?

Did MT tools develop to overcome the weaknesses listed in previous research?

Do LLMs provide a better alternative than the traditional MT tools?

A couple of Arabic and English speeches delivered by the King of Jordan, King Abdullah II, were selected to answer these questions. The speeches are provided in Arabic and English on the official website of King Abdullah II (https://kingabdullah.jo).

The first speech (address) was delivered in English on the 13th of December, 2023, at the Global Refugee Forum in Geneva. The speech was delivered in seven minutes. The official website of King Abdullah provides the speech in English as the source language (https://kingabdullah.jo/en/speeches/global-refugee-forum-geneva). The translation of the speech is also provided in Arabic (the target language) on the same website (كلمة-جلالة-الملك-عبدالله-الثاني-في-المنتدى-العالمي-للاجئين-بجنيف/https://kingabdullah.jo/ar/speeches). This translation is considered the official human translation of the speech as the source on the official website provides it.

King Abdullah II delivered the second speech (remarks) in the joint Arab-Islamic Extraordinary Summit on Gaza in Riyadh on the 11th of November, 2023. The official website of King Abdullah provides the speech in the source language, which is Arabic (كلمة-جلالة-الملك-عبدالله-الثاني-في-القمة العربية الإسلامية المشتركة غير/https://kingabdullah.jo/ar/speeches العادية حول غزة). The translation of the speech is also provided in English (the target language) on the website as well (https://kingabdullah.jo/en/speeches/joint-arab-islamic-extraordinary-summit-gaza-riyadh). This translation is also considered the official human-generated translation of the speech as the source on the official website provides it.

As the two speeches are provided in the source and the target languages by the official source, LLMs and MT translations are generated and compared to the speech in the source languages and the official translations in the target languages. The comparison and the analysis will look into a number of categories: (1) meaning, (2) functional and textual adequacy, (3) target language mechanics, (4) style, register, and idiomaticity, (5) overall rating (see figure 1 for the detailed evaluation rubric of translations). The evaluation rubric is a functional rubric with four levels for every category. The levels are (1) excellent, (2) acceptable, (3) deficient, and (4) unacceptable.

	Meaning	Functional and Textual Adequacy	Target Language Mechanics	Style, Register and Idiomaticity	Overall Rating
Excellent	☐ The translation accurately reflects the content of the source text. Slight nuances and shades of meaning have been rendered adequately. The translation contains few or no transfer errors, and those present have only a minor effect on meaning.	☐ The translation accurately addresses the target audience by fully taking its needs and characteristics into consideration. It also fully accomplishes the intended communicative function. Minor or no edits needed.	☐ The translation contains few or no violations of the rules and conventions of target language mechanics (grammar, spelling, punctuation, etc.). It reads similarly to texts originally written in the target language to the same purpose.	☐ Style and register are appropriate for the text type in the target language and for the target audience. The translation reads smoothly and wording is idiomatic. Any errors have only a minor effect on meaning.	☐ Publish and/or use as is.
Acceptable	☐ The translation contains occasional and/or minor transfer errors that slightly obscure or change meaning.	☐ The translation approximates to the intended communicative function and to the needs and characteristics of the target audience. Can be repaired with edits.	☐ The translation contains occasional errors in target language mechanics.	☐ The translation contains occasional and/or minor inappropriate style/register choices or occasional unidiomatic wording. Such errors may slightly obscure meaning.	☐ Minor edits needed before publishing.
Deficient	☐ The translation shows some misunderstanding of the original and/or the translation brief. It contains more than occasional changes in meaning and/or unwarranted omissions/additions.	☐ The translation gives some consideration to the intended communicative function and the needs of the target audience, but misses some important aspect(s) of it. Repair requires considerable effort.	☐ The translation is hard to comprehend because it contains frequent and/or obvious errors in target language mechanics. The structure of the source language shows up in the translation and affects its readability.	☐ The translation contains frequent inappropriate style/register choices or frequent unidiomatic wording. Such errors may obscure or change meaning.	☐ Major revision needed before publishing.
Unacceptable	☐ The translation reflects or contains important unwarranted deviations from the original. It contains inaccurate renditions and/or important omissions or additions that cannot be justified by the translation brief. Very defective comprehension of the original text.	☐ The translation disregards for the intended communicative function and fails to address the target audience. Cannot be repaired with revisions.	☐ The text is extremely difficult to read, bordering on being incomprehensible. The translation reveals serious language proficiency issues. Ungrammatical use of the target language.	☐ The translation contains excessive inappropriate style/register choices or disruptive unidiomatic wording. Such errors obscure or change meaning.	☐ Redo translation.

Figure 1: Translation evaluation rubric (adopted from Ustaszewski (2014))

The LLMs translation in this study is meticulously crafted by ChatGPT, a leading AI Chatpot in the current era, as noted by [51]. Renowned for its ability to generate human-like texts nearly indistinguishable from original content, ChatGPT has been the subject of extensive research to uncover its educational and public benefits [52, 53].

Google Translate generates MT in this study. According to [54], Google Translate is the best machine translation. After listing down a number of elements that need to be considered with machine translation (i.e., quality, consistency, confidentiality, security, customization, adaptability, ongoing support, and improvement), [54] concluded that Google Translate uses machine learning and neural networks to translate text and documents accurately.

Human Translation of the Speeches

The human translation of the two speeches is a testament to the precision and quality that can only be achieved through human involvement. It demonstrates accuracy in meaning, functional and textual adequacy, target language mechanics, style, register, and idiomaticity, earning it the 'publish and use as is' overall rating.

For the Arabic speech (414 words- 2102 characters without spaces (2496 characters with spaces)) translated into English, the meaning has been accurately transferred and reflects the source text. We have slight nuances and shades with meaning, and we do not have transfer errors. The functional and textual adequacy is accurate. It addresses the audience in a way that takes into consideration the characteristics. We do not need edits, maybe except in the part where the king

addressed the audience using الإخوة القادة, and it was translated as 'brothers.' It could have been translated as '*dear leaders.*' The literal translation of the expression is '*brother leaders.*' As '*brother leaders*' is not idiomatic in English, the translator focused on '*brothers.*' The translator could have focused on '*leaders*' as decision-making is needed in this context. For target language mechanics, the translation does not contain any violations of target language mechanics in terms of spelling, punctuation, and grammar. The speech also reads as texts that originated in the target language. Style, register, and idiomaticity are appropriate for the same level in the target language. The translation is smooth, and the wording is idiomatic.

For the English speech (740 words- 3887 characters with no spaces (4599 characters with spaces)) translated to Arabic, it can also be said that the meaning has been accurately transferred and reflects the source text. We have slight nuances and shades with meaning, and we do not have transfer errors. The functional and textual adequacy is accurate. It addresses the audience in a way that takes into consideration the characteristics. It is also evident that the translator used Arabic rhetoric to enhance understanding. We do not need edits. For the point raised above in the Arabic-English translation, it is noted that King Abdullah II used 'dear brothers' to address the audience. The translator literally translated it to ايه الاخوه, which is accurate. For target language mechanics, the translation does not contain any violations of target language mechanics in terms of spelling, punctuation, and grammar. The speech also reads as texts that originated in the target language. Style, register, and idiomaticity are appropriate for the same level in the target language. The translation is smooth, and the wording is idiomatic. For clarity, when the king used the abbreviation '*UNRWA,*' the Arabic translator included the abbreviation in *Arabic (الأونروا)* and added وكاله الامم المتحدة لاغاثة وتشغيل الاجئين, the name in full.

This reveals that the human translations of Arabic to English and English to Arabic speeches are excellent based on the rubric used. Both translations are ready to be published without changes. In fact, they are published on King Abdullah II's official website.

Machine Translation of the Speeches

As mentioned above, Google Translate was used for machine translation. The first weakness noticed in Google Translate is that it does not allow translation that is longer than 5000 characters at a time. If a text is longer than this number, the copy-pasting process must be done more than once.

Arabic-English Google Translation

An overview of the Arabic-English Google Translate translation reveals misunderstandings, textual and contextual inaccuracies, and biases. This does not mean that Google Translation is 'unacceptable.' It is 'deficient' as the translation needs significant revisions before publication. This is the case as the translated text is a King's speech that needs to be accurate and precise to avoid misunderstanding.

Regarding meaning, the first category in the evaluation rubric, it is noticed that, in general terms, Google Translate provides a relatively satisfactory translation, but with some deviations, contextual misunderstanding, and biases. For example, 'الحضور الكرام' was translated as '*distinguished attendees.*' The attendees at the summit, in addition to the state members, are 'guests' who were invited to attend the summit. The human translation of 'الحضور الكرام' is '*distinguished guests,*' which is more accurate. In addition, 'للقتل والتدمير' was translated by Google Translate as '*being killed and destroyed.*' The human translation is '*face death and destruction.*' The human translation obviously provides better expression than Google Translate, making the utterance sound more potent than the literal translation *being killed and destroyed*. The literal understanding of the source text in the machine translation can also be noticed in example 1. The word 'انحيازا' was literally translated as 'a bias.' In this context, the intended meaning was 'victory,' as the human translation shows. King Abdullah II intended to say that 'human values won as the United Nations General Assembly's resolution on Gaza' was 'fair.' As such, it can be concluded that in relation to meaning, the machine translation can be considered 'deficient 'the translation shows some misunderstanding of the original and the translation brief. It contains more than occasional changes in meaning and unwarranted omissions/additions'.

EX 1:	SL	لقد كان قرار الجمعية العامة للأمم المتحدة بشأن غزة انتصارا للقيم الإنسانية، وانحيازا للحق في الحياة والسلام
	TL-MT	The United Nations General Assembly's resolution on Gaza was a victory for human values, a bias towards the right to life and peace.
	TL-HT	The United Nations General Assembly's decision on Gaza was a victory for humanitarian values and for the right to life and peace.

SL: Source language; **TL-MT**: Target language-Machine Translation; **TL-HT**: Target language-Human translation

Regarding functional and textual adequacy, machine translation provides some consideration to the intended communicative function and the needs of the target audience but misses some important aspects. Repair requires considerable effort'. That is, the functional and textual adequacy of Google Translate is 'Deficient.' For functional adequacy, Google Translate ensures that the translated text achieves the same communicative effect as the original text as it considers the target audience, context, and the desired outcome of the translation. It generally conveys the original text's meaning, tone, and style appropriately to the target audience. For textual adequacy, Google Translate ensures that the translation preserves the original text's meaning, nuances, and organization as closely as possible. It involves maintaining the source text's coherence, cohesion, and stylistic features in the target language. A textually adequate translation accurately represents the source text without adding or omitting significant information. The only possible functional and textual adequacy issue is related to some needed linguistic changes, as example 2 shows. As example 2 shows, 'على استضافة المملكة العربية السعودية لهذه القمة العربية الإسلامية' was translated as '*for hosting the Kingdom of Saudi Arabia for this Arab-Islamic summit*.' This textual inadequacy can be edited as '*for hosting this Arab-Islamic summit in the Kingdom of Saudi Arabia*'.

EX 2: SL أشكر أخي خادم الحرمين الشريفين، وأخي سمو الأمير محمد بن سلمان، على استضافة المملكة العربية السعودية لهذه القمة العربية الإسلامية

TL-MT I thank my brother, the Custodian of the Two Holy Mosques, and my brother, His Highness Prince Mohammed bin Salman, for hosting the Kingdom of Saudi Arabia for this Arab-Islamic summit.

TL-HT I would like to thank my brother, the Custodian of the Two Holy Mosques, and His Royal Highness Crown Prince Mohammed bin Salman, for hosting this Arab-Islamic summit in the Kingdom of Saudi Arabia.

SL: Source language; **TL-MT**: Target language-Machine Translation; **TL-HT**: Target language-Human translation

This functional and textual inadequacy can also be noticed in example 3. The use of 'the mentality of the citadel' in the place of '*fortress mentality*' for 'عقلية القلعة' and the use of '*isolation walls*' in the place of '*separation walls*' for 'وجدران العزل' which are not idiomatic, and the use of 'sanctities and rights prevailed' for 'المقدسات والحقوق,' which shall be translated as '*holy sites and rights,*' are examples of functional and textual inadequacy in Google translate. This shows that Google Translate provides a literal translation that does not provide accurate context-based and functional-level translation.

EX 3: SL هذا الظلم لم يبدأ قبل شهر، بل هو امتداد لأكثر من سبعة عقود سادت فيها عقلية القلعة وجدران العزل والاعتداء على المقدسات والحقوق، وغالبية ضحاياها المدنيون الأبرياء

TL-MT This injustice did not begin a month ago. Instead, it is an extension of more than seven decades in which the mentality of the citadel, isolation walls, and attacks on sanctities and rights prevailed, with the majority of its victims being innocent civilians.

TL-HT This injustice did not begin a month ago. It is a continuation of over seven decades dominated by a fortress mentality of separation walls and violations against holy sites and rights, the majority of whose victims are innocent civilians.

SL: Source language; **TL-MT**: Target language-Machine Translation; **TL-HT**: Target language-Human translation

For target language mechanics (i.e., grammar, punctuation, and spelling), the machine translation 'contains few or no violations of the rules and conventions of the target language mechanics (i.e., grammar, punctuation, and spelling). It reads similarly to texts originally written in the target language regarding grammar, punctuation, and spelling. However, in terms of style, register, and idiomaticity, the text is far from being read, similar to texts originally written in the target language. As Table 2 shows, the style and register of the translated text are lower than the human-translated text. For example, 'سيدنا محمد' was translated as '*Master Muhammad*'. Muslims do not address the Prophet of Islam as the '*Master.*' Muslims address the Prophet of Islam as '*Prophet Muhammad.*' In addition, 'صدام' was translated as '*clash*' by Google Translate, whereas it was translated as '*conflict* by the human translator. 'نجتمع' was translated as '*gather*' by Google Translate, whereas

it was translated as '*convene*' by the human translation as the translator realized that it was a summit. In addition, 'امتداد' was translated as '*extension*' by Google Translate, whereas the human translator translated it as '*continuation*' as the translator knows that the speech is about a war. Moreover, 'بؤرة' was translated as '*hotbed*' by Google. In contrast, it was translated as '*source*' by the human translators, which reflects a higher register and an idiomatic expression by the human translator to match the register of the original speech (AlAfnan, 2018).

Table 1: Style, register, and idiomaticity in MT Arabic-English translation

Source Language	Machine Translation	Human Translation
سيدنا محمد	Master Muhammad,	Prophet Muhammad
نجتمع اليوم	We gather today	We convene today
صدام كبير	major clash	major conflict
قد تصل	may reach	will spiral into
هو امتداد لأكثر	it is an extension of	it is a continuation of
تخنق الحياة	stifle life	suffocates life
ولا يمكن أن نقبل أن تتحول قضيتنا الشرعية العادلة إلى بؤرة تشعل الصراع بين الأديان.	We cannot accept that our just, legitimate cause becomes a *hotbed that ignites inter-religious conflict.*	We cannot allow for our just and legitimate cause to be turned into *a source of fomenting conflict between religions.*

Based on the above, Google Translate's overall rating of the Arabic-English machine translation is 'Deficicint.' That is, the translation needs significant revisions before publishing.

English-Arabic Google Translation

The English-Arabic Google translation also shows some literal translation practices. An overview of the translation would give the impression that the translation is 'acceptable.' However, a detailed investigation into meaning, functional and textual adequacy, target language mechanics, style, register, and idiomaticity would reveal that the transition needs many edits.

Regarding meaning, the machine translation can be considered 'deficient as 'the translation shows some misunderstanding of the original and the translation brief. It contains more than occasional changes in meaning and unwarranted omissions/additions'. For example, '*Jordan's national identity*,' in example 4, was literally translated to 'الهوية الوطنية الأردنية.' The proper translation shall be 'المبادئ الوطنية الأردنية' as the discussion here is not literally about the national identity (الهوية); it is about principles and values (المبادئ). In example 4, the translation of the 'because that is who we are' was also literally word-for-word translated to 'لأن هذا هو ما نحن عليه.' This should have been translated as 'لأن ذلك يتنافى مع صميم هويتنا.' The ideal translation in Arabic means that ' *as this (turning our back to refugees) is against our values and principles'*.

EX 4:   SL   In an increasingly volatile region, welcoming refugees has become an indelible part of Jordan's national identity. We cannot turn our backs on refugees, because that is who we are.

     TL-MT   وفي منطقة مضطربة بشكل متزايد، أصبح الترحيب باللاجئين جزءاً لا يمحى من الهوية الوطنية الأردنية لا يمكننا أن ندير ظهورنا للاجئين، لأن هذا هو ما نحن عليه

     TL-HT   ان منح الملاذ الآمن للاجئين جزء لا يتجزأ من المبادئ الوطنية الأردنية، خصوصا في هذه المنطقة المضطربة، فلا يمكننا أن ندير ظهورنا لهم لأن ذلك يتنافى مع صميم هويتنا.

<p align="center">**SL**: Source language; **TL-MT**: Target language-Machine Translation; **TL-HT**: Target language-Human translation</p>

The discussion on the accuracy of the transfer of meaning can also take us to the functional and textual adequacy of the translation. In example 5, the machine translation has totally distorted the meaning of the source text. '*But Jordanians have been increasingly feeling that the world is turning its backs on them, as refugee hosts*' was translated as 'لكن الأردنيين يشعرون بشكل متزايد بأن العالم يدير ظهره لهم، باعتبارهم مضيفين للاجئين'. If the machine translation were translated back to English, it would mean, '*But Jordanians have been increasingly feeling that the world is turning its backs on them because they host refugees.*' That is, Jordanians shall not host refugees to avoid being left behind. This is apparent distortion. The ideal translation of King Abdullah II's speech excerpt shall be 'لكن الأردنيين يشعرون بشكل متزايد بأن العالم يدير ظهره لهم، ويتجاهل جهودهم كمستضيفين للاجئين' which if translated back to English would means that the world ignores the efforts of Jordan to host refugees and does not provide enough eid.

EX 5:   SL   But Jordanians have been increasingly feeling that the world is turning its backs on them, as refugee hosts.

     TL-MT   لكن الأردنيين يشعرون بشكل متزايد بأن العالم يدير ظهره لهم، باعتبارهم مضيفين للاجئين

     TL-HT   لكن الأردنيين يشعرون بشكل متزايد بأن العالم يدير ظهره لهم، ويتجاهل جهودهم كمستضيفين للاجئين

<p align="center">**SL**: Source language; **TL-MT**: Target language-Machine Translation; **TL-HT**: Target language-Human translation</p>

The functional and textual inadequacy in machine translation is also evident in a number of other occurrences. In example 6, '*as serious crises compete for international attention*' was translated as 'وبينما تتنافس الأزمات الخطيرة على جذب الاهتمام الدولي.' The use of the word 'تتنافس' to translate '*compete*' is literal and does not provide the functional meaning. It gives the impression that '*crises strive to achieve international attention.*' The intended meaning is that there are several international crises. As such, they '*crowd*' to achieve attention. In Arabic, the translation shall be 'تتزاحم' as it gives a functional meaning rather than a literal meaning for the word use. In addition, in example 6, '*the plight of refugees and their host countries has taken a backseat*' was translated by Google Translate as 'تراجعت محنة اللاجئين والدول المضيفة لهم' If translated back to English, this sentence means '*the plight of refugees and their host countries have eased.*' That is, the plight of refugees and host countries has been resolved. This is also a distortion of the original text as the ideal transaction

shall be 'يتراجع التركيز على محنة اللاجئين والبلدان المستضيفة لهم' to provide functional and textual adequacy and accuracy.

EX 6:	SL	As severe crises compete for international attention, the plight of refugees and their host countries has taken a backseat. But this is a lapse that the international community can ill afford.
	TL-MT	وبينما تتنافس الأزمات الخطيرة على جذب الاهتمام الدولي، تراجعت محنة اللاجئين والدول المضيفة لهم ولكن هذه هفوة لا يستطيع المجتمع الدولي أن يتحملها
	TL-HT	وبينما تتزاحم الأزمات الخطيرة لتستحوذ على الاهتمام الدولي، يتراجع التركيز على محنة اللاجئين والبلدان المستضيفة لهم، إلا أن المجتمع الدولي لا يملك ترف تجاهل هذه القضية.

SL: Source language; **TL-MT**: Target language-Machine Translation; **TL-HT**: Target language-Human translation

Regarding target language mechanics (i.e., spelling, grammar, and punctuation). The target text is almost free of any spelling and grammar errors. However, regarding punctuation, the target text follows the punctuation marks used in the source text. However, according to Al Qinai (2008), "Arabic has its conventions of punctuation which are not strictly governed by the same rules applicable to English" (p. 5). Therefore, placing punctuation marks in the target text in place of punctuation marks in the source text disturbs the smooth flow of ideas in the target text. The use of punctuation marks in machine translation shall follow the language mechanics of the target language.

Table 2: Style, register, and idiomaticity in MT English-Arabic translation

Source Language	Machine Translation	Human Translation
And now, as we speak	والآن، بينما نتحدث	والآن، بينما نجتمع
With all eyes on Gaza	ومع كل الأنظار نحو غزة	ومع توجه كل الأنظار نحو غزة
global crises demand long-term responsibility-sharing	وأن الأزمات العالمية تتطلب تقاسم المسؤولية على المدى الطويل	وأن الأزمات العالمية تستوجب التشارك في تحمل المسؤولية على المدى الطويل
Jordan is pushing for a more coordinated humanitarian response in Gaza.	ويدفع الأردن من أجل استجابة إنسانية أكثر تنسيقا في غزة	بضغط الأردن باتجاه استجابة إنسانية أكثر تنسيقا في غزة
we recognize that this is a long-term commitment that we are undertaking on behalf of the international community	ولكننا ندرك أن هذا الالتزام طويل الأمد نتعهد به بالنيابة عن المجتمع الدولي.	لكننا نعي أن هذا الالتزام طويل المدى نتحمله بالنيابة عن المجتمع الدولي
Let's make this forum count.	دعونا نجعل هذا المنتدى مهمًّا	فلنعمل معا لنجاح هذا المنتدى

Concerning style, register, and idiomaticity, as Table 2 shows, there are several occurrences of amateurish style and lower register. The context of the speech is the Global Refugee Forum in Geneva. The attendees are heads of state and international delegates. The style of the original

speech is formal and professional. In the machine translation, '*as we speak*' was literally translated as 'بينما نتحدث' which gives the impression of an informal context. The ideal translation shall be 'بينما نجتمع' to maintain the style and the formality (register) of the speech. In addition, we also have idiomaticity inaccuracy, as we see in the translation of 'responsibility-sharing,' which was translated literally to 'تقاسم المسؤولية.' The ideal and idiomatic translation shall be ' التشارك في تحمل المسؤولية ' تحمل المسؤولية is idiomatic in Arabic, whereas 'تقاسم المسؤولية' is a literal translation that is not used in the Arabic language. The translation of '*let us make this forum count*' also provides a clear example of the deficient translation provided by Google Translate as it translated it as ' دعونا نجعل هذا المنتدى مهمًا,' which literally means 'let us make this forum important.' The accurate and ideal translation for this sentence shall be 'فلنعمل معا لنجاح هذا المنتدى' *which means 'let us work together for the success of this forum.'*

Google Translate's English-Arabic machine translation is rated as 'Deficient.' That is, it needs significant revisions before publishing.

Large Language Models Translation

As mentioned above, ChatGPT 4 was used for the Artificial Intelligence-Large Language Model (LLM) transaction. Chat GPT 4 is available for free on the OpenAI website. To translate the Arabic to English speech, the 'translate this to English: (pasted the text of the speech)' prompt was used. The 'translate this to Arabic: (pasted the text of the speech)' prompt was used to translate the English to Arabic speech. It took ChatGPT 4 almost 4 seconds to start translating the texts.

Arabic-English Large Language Model Translation

An overview of the Arabic-English ChatGPT 4 translation reveals that it is more accurate than Google Translate machine translation but less accurate than human translation.

With meaning, the Arabic-English translation provided a decent transfer of meaning to the target language. The points discussed in the Arabic-English Google Translate also exist in the LLM translation but with a lesser impact. For example, the 'الحضور الكرام' was translated as '*distinguished attendees,*' which carries the same connotation as discussed earlier. This is the case as the attendees are officials and heads of state invited to the summit. The more accurate translation should have been '*distinguished guests.*' In addition, 'يتعرضون للقتل والتدمير' was translated as 'subjected to killing and destruction.' This is better than the translation provided by Google Translate (*being killed and destroyed*), but the use of '*face death and destruction*' would have provided a better description of the event. Similarly, the description of the United Nations General Assembly's resolution on Gaza as *'bias'* was not a successful translation as it is described as '*right*. This translation comes as a result of the literal translation of the word 'وانحيازا' which can also be translated as '*right*. The word comes in the context of 'وانحيازا للحق في الحياة والسلام' which could have been translated as '*victory for humanitarian values and for the right to life and peace*' (see example 7).

EX 7: SL لقد كان قرار الجمعية العامة للأمم المتحدة بشأن غزة انتصارا للقيم الإنسانية، وانحيازا للحق في الحياة والسلام

TL-LLM The decision of the United Nations General Assembly regarding Gaza was a triumph for human values, a bias towards the right to life and peace.

TL-HT The United Nations General Assembly's decision on Gaza was a victory for humanitarian values and for the right to life and peace.

<center>SL: Source language; **TL-MT**: Target language-Machine Translation; **TL-HT**: Target language-Human translation</center>

In relation to functional and textual adequacy, the LLM-translated text seems and sounds more natural than the text translated by Google Translate. As example 8 shows, ' على استضافة المملكة العربية السعودية لهذه القمة العربية الإسلامية' was translated as '*for hosting this Arab-Islamic summit in the Kingdom of Saudi Arabia.*' This translation differs from the Google Translate-provided translation (for hosting the Kingdom of Saudi Arabia for this Arab-Islamic summit). It is precisely the same as the translation provided by the human translation. However, it is noticed that ChatGPT follows the English language punctuation style. In Arabic, if there is a list, the letter 'و,' which means 'and,' is used, not a comma, to separate the items in the list. In the ChatGPT translation, commas are used, and no 'و' is added following the English language punctuation style.

EX 8: SL أشكر أخي خادم الحرمين الشريفين، وأخي سمو الأمير محمد بن سلمان، على استضافة المملكة العربية السعودية لهذه القمة العربية الإسلامية

TL-LLM I thank my brother, the Custodian of the Two Holy Mosques, and my brother, His Royal Highness Prince Mohammed bin Salman, for hosting this Arab-Islamic summit in the Kingdom of Saudi Arabia.

TL-HT I would like to thank my brother, the Custodian of the Two Holy Mosques, and His Royal Highness Crown Prince Mohammed bin Salman, for hosting this Arab-Islamic summit in the Kingdom of Saudi Arabia.

<center>SL: Source language; **TL-MT**: Target language-Machine Translation; **TL-HT**: Target language-Human translation</center>

The functional and textual inadequacy in LLM translation can also be noticed in example 9. The use of 'mentality of siege' in the place of '*fortress mentality'* for 'عقلية القلعة' and the use of '*isolation walls'* in the place of '*separation walls'* for 'وجدران العزل,' which are not idiomatic, and the use of '*sanctities and rights prevailed'* for 'المقدسات والحقوق,' which shall be translated as '*holy sites and rights,'* are examples of functional and textual inadequacy in LLM translation. This shows that, like Google Translate, LLM also provides literal translations that do not provide accurate context-based and functional-level translation.

EX 9: SL هذا الظلم لم يبدأ قبل شهر، بل هو امتداد لأكثر من سبعة عقود سادت فيها عقلية القلعة وجدران العزل والاعتداء على المقدسات والحقوق، وغالبية ضحاياها المدنيون الأبرياء

TL-LLM	This injustice did not begin a month ago; instead, it is an extension of more than seven decades during which the mentality of siege, walls of isolation, and attacks on sanctities and rights prevailed, with the majority of its victims being innocent civilians.
TL-HT	This injustice did not begin a month ago. It is a continuation of over seven decades dominated by a fortress mentality of separation walls and violations against holy sites and rights, the majority of whose victims are innocent civilians.

SL: Source language; **TL-MT**: Target language-Machine Translation; **TL-HT**: Target language-Human translation

LLM translation provides the target text with accurate spelling and grammar for target language mechanics (i.e., grammar, punctuation, and spelling). However, it uses English language punctuation techniques, as mentioned earlier. Using commas to separate lists without using the word 'و' which means '*and*' in English, is wrong as it does not follow the mechanics of the target language (Arabic). However, in terms of style, register, and idiomaticity, the text is produced more accurately than MT. As Table 2 shows, 'سيدنا محمد' was translated as Prophet Muhammad,' which is accurate. Google Translate translated it as '*Master Muhammad.*' In addition, 'صدام' was translated as 'conflict,' which is also correct. Google Translate translated it as '*clash,*' which does not represent the level of the issue. It has also been noticed that the LLM translation sometimes does not provide accurate words in the target language. For example, 'نجتمع' was translated as '*gather*' by ChatGPT and Google Translate, whereas it was translated as '*convene*' by the human translation as the translator realized that it was a summit. In addition, 'امتداد' was translated as '*extension*' by ChatGPT and Google Translate. In contrast, the human translator translated it as '*continuation*' as the translator knows that the speech is about a war. Moreover, 'بؤرة' was translated as '*hotbed*' by ChatGPT and Google. In contrast, it was translated as '*source*' by the human translators, which reflects a higher register and an idiomatic expression by the human translator to match the register of the original speech.

Table 3: Style, register, and idiomaticity in LLM Arabic-English translation

Source Language	LLM Trsnaltion	Machine Translation	Human Translation
سيدنا محمد	Prophet Muhammad	Master Muhammad,	Prophet Muhammad
نجتمع اليوم	Today, we gather	We gather today	We convene today
صدام كبير	major conflict	major clash	major conflict
قد تصل	may reach	may reach	will spiral into
هو امتداد لأكثر	It is an extension of	it is an extension of	it is a continuation of
تخنق الحياة	stifle life	stifle life	suffocates life
ولا يمكن أن نقبل أن تتحول قضيتنا الشرعية العادلة إلى بؤرة تشعل الصراع بين الأديان.	We cannot accept that our just, legitimate cause becomes a hotbed that ignites inter-religious conflict.	We cannot accept that our just, legitimate cause becomes a *hotbed that ignites inter-religious conflict.*	We cannot allow for our just and legitimate cause to be turned into *a source of fomenting conflict between religions.*

Based on the above, the overall rating of the Arabic-English ChatGPT translation provided is 'Acceptable. ' That is, the translation needs minor edits before publishing.

English-Arabic LLM Translation

The English-Arabic LLM translation can be considered 'acceptable'. A detailed investigation into meaning, functional and textual adequacy, target language mechanics, style, register, and idiomaticity reveals that the translation needs fewer edits than Google Translate but can be published with minor edits.

Regarding meaning, as |Google Translate, the LLM translation can be considered 'deficient as 'the translation shows some misunderstanding of the original and the translation brief. It contains more than occasional changes in meaning and unwarranted omissions/additions'. For example, '*Jordan's national identity*,' in example 10, was literally translated to 'الهوية الوطنية الأردنية'. The proper translation shall be 'المبادئ الوطنية الأردنية' as the discussion here is not literally about the national identity (الهوية); it is about principles and values (المبادئ). In example 10, the translation of the '*because that is who we are*' was also literally word-for-word translated to 'لأن هذا هو من نحن'. This should have been translated as 'لأن ذلك يتنافى مع صميم هويتنا'. The ideal translation in Arabic means that ' *as this (turning our back to refugees) is against our values and principles*'. This shows that, in regards to meaning, the LLM translation is almost identical to the Google-provided MT.

EX 10:	SL	In an increasingly volatile region, welcoming refugees has become an indelible part of Jordan's national identity. We cannot turn our backs on refugees, because that is who we are.
	TL-LLM	في منطقة متزايدة الاضطراب، أصبح استقبال اللاجئين جزءًا لا يمحى من الهوية الوطنية الأردنية. لا يمكننا أن ندير ظهورنا للأشخاص النازحين، لأن هذا هو من نحن
	TL-MT	وفي منطقة مضطربة بشكل متزايد، أصبح الترحيب باللاجئين جزءاً لا يمحى من الهوية الوطنية الأردنية لا يمكننا أن ندير ظهورنا للاجئين، لأن هذا هو ما نحن عليه
	TL-HT	ان منح الملاذ الآمن للاجئين جزء لا يتجزأ من المبادئ الوطنية الأردنية، خصوصا في هذه المنطقة المضطربة، فلا يمكننا أن ندير ظهورنا لهم لأن ذلك يتنافى مع صميم هويتنا.

SL: Source language; **TL-MT**: Target language-Machine Translation; **TL-HT**: Target language-Human translation

Example 11 also shows that the LLM translation has totally distorted the meaning of the source text. '*But Jordanians have been increasingly feeling that the world is turning its backs on them, as refugee hosts*' was translated as 'لكن الأردنيين يشعرون بشكل متزايد بأن العالم يدير ظهره لهم، باعتبارهم كمضيفين للاجئين'. If the machine translation were translated back to English, it would mean, '*But Jordanians have been increasingly feeling that the world is turning its backs on them as hosts of refugees.*' As in Google Translate, Jordanians shall not host refugees to avoid being left behind. This is apparent distortion. The ideal translation of King Abdullah II's speech excerpt shall be ' لكن الأردنيين يشعرون بشكل متزايد بأن العالم يدير ظهره لهم، ويتجاهل جهودهم كمستضيفين للاجئين' which if translated

back to English would means that the world ignores the efforts of Jordan to host refugees and does not provide enough eid.

EX 11:	SL	However, Jordanians have been increasingly feeling that the world is turning its backs on them, as refugee hosts.
	TL-LLM	لكن الأردنيين يشعرون بشكل متزايد بأن العالم يدير ظهوره لهم، كمضيفين للاجئين
	TL-MT	لكن الأردنيين يشعرون بشكل متزايد بأن العالم يدير ظهره لهم، باعتبارهم مضيفين للاجئين
	TL-HT	لكن الأردنيين يشعرون بشكل متزايد بأن العالم يدير ظهره لهم، ويتجاهل جهودهم كمستضيفين للاجئين

SL: Source language; **TL-MT**: Target language-Machine Translation; **TL-HT**: Target language-Human translation

The functional and textual inadequacy in machine translation is also evident in ChatGPT translation. In example 6, '*as serious crises compete for international attention*' was translated as 'مع تنافس الأزمات الجسيمة على الانتباه الدولي'. The use of the word '*تنافس*' to translate '*compete*' is literal and does not provide a functional meaning. It gives the impression that '*crises strive to achieve international attention.*' The intended meaning is that there are several international crises. As such, they '*crowd*' to achieve attention. In Arabic, the translation shall be '*تتزاحم*' as it gives a functional meaning rather than a literal meaning for the word use. In addition, in example 12, '*the plight of refugees and their host countries has taken a backseat*' was translated by ChatGPT as 'انحصرت معاناة اللاجئين وبلدان الاستضافة لهم على المقاعد الخلفية' If translated to English, this sentence means '*the plight of refugees and their host countries have taken the backseats.*' Even though this translation does not fully provide an accurate picture of the situation, it is better than Google Translate's translation to describe the situation. The translation here is not distorted, as in the translation given by Google Translate, but it requires minor edits.

EX 12:	SL	As serious crises compete for international attention, the plight of refugees and their host countries has taken a backseat. But this is a lapse that the international community can ill afford.
	TL-LLM	مع تنافس الأزمات الجسيمة على الانتباه الدولي، انحصرت معاناة اللاجئين وبلدان الاستضافة لهم على المقاعد الخلفية. ولكن هذا هو انقطاع يمكن للمجتمع الدولي أن يحمل ثمنه بشكل صعب
	TL-MT	وبينما تتنافس الأزمات الخطيرة على جذب الاهتمام الدولي، تراجعت محنة اللاجئين والدول المضيفة لهم ولكن هذه هفوة لا يستطيع المجتمع الدولي أن يتحملها
	TL-HT	وبينما تتزاحم الأزمات الخطيرة لتستحوذ على الاهتمام الدولي، يتراجع التركيز على محنة اللاجئين والبلدان المستضيفة لهم، إلا أن المجتمع الدولي لا يملك ترف تجاهل هذه القضية.

SL: Source language; **TL-MT**: Target language-Machine Translation; **TL-HT**: Target language-Human translation

Regarding target language mechanics (i.e., spelling, grammar, and punctuation), as in Google Translate, the target text is almost free of spelling and grammar errors. However, regarding punctuation, the target text follows the punctuation marks used in the source text. Therefore,

placing punctuation marks in the target text in place of punctuation marks in the source text disturbs the smooth flow of ideas in the target text. The use of punctuation marks in machine translation shall follow the language mechanics of the target language.

Table 4: Style, register, and idiomaticity in LLM English-Arabic translation

Source Language	LLM Translation	Machine Translation	Human Translation
And now, as we speak	والآن، ونحن نتحدث	والآن، بينما نتحدث	والآن، بينما نجتمع
With all eyes on Gaza	ينما تتجه الأنظار نحو غزة	ومع كل الأنظار نحو غزة	ومع توجه كل الأنظار نحو غزة
global crises demand long-term responsibility-sharing	أن الأزمات العالمية تتطلب مشاركة على المدى المسؤولية الطويل	وأن الأزمات العالمية تتطلب تقاسم المسؤولية على المدى الطويل	وأن الأزمات العالمية تستوجب التشارك في تحمل المسؤولية على المدى الطويل
Jordan is pushing for a more coordinated humanitarian response in Gaza.	يسعى الأردن إلى تعزيز استجابة إنسانية متنسقة في غزة	ويدفع الأردن من أجل استجابة إنسانية أكثر تنسيقا في غزة	يضغط الأردن باتجاه استجابة إنسانية أكثر تنسيقا في غزة
we recognize that this is a long-term commitment that we are undertaking on behalf of the international community	لكننا ندرك أن هذا التزامًا طويل الأمد نتخذه نيابة عن المجتمع الدولي.	ولكننا ندرك أن هذا التزام طويل الأمد نتعهد به بالنيابة عن المجتمع الدولي.	لكننا نعي أن هذا الالتزام طويل المدى نتحمله بالنيابة عن المجتمع الدولي
Let's make this forum count.	لنجعل هذا المنتدى يحقق نتائج.	دعونا نجعل هذا المنتدى مهمًا	فلنعمل معا لنجاح هذا المنتدى

Concerning style, register, and idiomaticity, as Table 4 shows, there are some occurrences of amateurish style and lower register. The context of the speech is the Global Refugee Forum in Geneva. The attendees are heads of state and international delegates. The style of the original speech is formal and professional. In the machine translation, '*as we speak*' was literally translated as 'ونحن نتحدث,' which gives the impression of an informal context. The ideal translation shall be 'بينما نجتمع' to maintain the style and the formality (register) of the speech. In addition, we also have idiomaticity inaccuracy, as we see in the translation of 'responsibility-sharing,' which was translated literally to 'مشاركة على المدى المسؤولية.' The ideal and idiomatic translation shall be 'التشارك في تحمل المسؤولية.' 'مشاركة على المدى المسؤولية' is a totally is idiomatic in Arabic, whereas 'في تحمل المسؤولية' inaccurate translation that is not used in the Arabic language. However, unlike Google Translate, LLM translation provided a better translation for '*let us make this forum count.*' It translated this sentence as 'لنجعل هذا المنتدى يحقق نتائج,' which is acceptable as it provides an accurate translation. This translation means, '*Let us make this forum achieve results.*'

Overall, the English-Arabic LLM translation provided by ChatGPT 4 is rated between 'Deficicint' and 'Acceptable.' This rating is higher than Google Translate's as LLM translation does have meaning, functional and textual inadequacy, mechanics, style, and register mistakes. However, these mistakes are less than those generated by Google Translate.

Discussion

This study examined the utilization of Large Language Models (LLMs), specifically focusing on ChatGPT as computational linguistics machine translation tools, contrasting them with traditional machine translation tools, particularly Google Translate. The study investigated both Arabic and English as source and target languages. Translations generated by ChatGPT and Google Translate were compared against the official translations of speeches of King Abdullah II of Jordan. The Arabic speech was delivered on November 11, 2023, at the joint Arab-Islamic Extraordinary Summit on Gaza in Riyadh, while the English speech was delivered on December 13, 2023, at the Global Refugee Forum in Geneva.

The evaluation was conducted using a rubric provided by [55]. The rubric evaluates translations based on five categories: meaning, functional and textual adequacy, target language mechanics, style, register, idiomaticity, and an overall rating. The evaluation incorporated textual and contextual analysis to ensure translations were evaluated within context.

Analysis of the Arabic-English Google Translate translation revealed misunderstandings, textual and contextual inaccuracies, and biases. While not deemed 'unacceptable,' it was rated as 'deficient' and required major revisions before publication. Google Translate provided a relatively satisfactory translation in terms of meaning but exhibited deviations, contextual misunderstandings, and occasional changes in meaning.

The English-Arabic Google Translate translation also exhibited literal translation practices and was rated as 'deficient.' Although it seemed 'acceptable' at first glance, a detailed examination revealed the need for significant edits, mainly in meaning, functional and textual adequacy, target language mechanics, style, register, and idiomaticity.

In contrast, the overall rating of the Arabic-English ChatGPT translation was 'Acceptable,' indicating the need for minor edits before publishing. While also exhibiting deficiencies in meaning, it offered more natural-sounding translations than Google Translate, making it a better choice.

Similarly, the English-Arabic LLM translation provided by ChatGPT 4 was rated between 'Deficient' and 'Acceptable,' with fewer mistakes than Google Translate. Despite exhibiting meaning, functional, and textual inadequacies, it was deemed better than Google Translate and only required minor edits for clarity. However, it still showed some mechanics, style, and register issues.

Overall, Google Translate translation is mainly deficient and can be nearly acceptable with significant edits. LLM-ChatGPT translation is better than Google Translate and ranges, in terms of evaluation, between low deficient and mainly acceptable with minor edits. As this study shows, human translation cannot and shall not be replaced to ensure that the translation is accurate and transfers source texts' textual and contextual aspects. Neither Google Translate nor ChatGPT can replace the human touch in translations to provide an accurate and comprehensive picture of source texts.

Implications and Possible Actions Based on the Study

The study demonstrates that ChatGPT provides more natural and contextually appropriate translations than Google Translate, but both tools still exhibit significant deficiencies, particularly in handling complex, context-specific content. Developers of translation tools should focus on enhancing LLMs' contextual understanding and natural language processing capabilities. They shall continue refining these models, particularly in understanding and preserving the nuances of different languages, which is crucial.

The findings underline the irreplaceability of human translators for ensuring the accuracy and contextual integrity of translations, especially for complex texts like speeches. Neither ChatGPT nor Google Translate can fully replicate the human ability to grasp and convey subtle nuances and cultural context. There shall be an emphasis on integrating human oversight in machine translation workflows, particularly for critical or sensitive content. Developers shall explore hybrid approaches that leverage the strengths of both human translators and machine translation tools.

Using Ustaszewski's rubric to evaluate translation quality highlights the importance of comprehensive evaluation criteria that cover meaning, functionality, textual adequacy, and more. This rigorous evaluation framework helps identify specific areas where machine translations fall short. Researchers and lecturers shall adopt similar comprehensive rubrics for ongoing assessment and improvement of translation tools. They shall also update these rubrics to reflect evolving standards and expectations in translation quality.

The study identifies deficiencies in Google Translate, such as literal translations and contextual inaccuracies. This suggests areas where targeted improvements can significantly enhance performance. Developers shall focus on reducing literal translations and improving Google Translate's handling of context and meaning. They shall also consider incorporating feedback from detailed human reviews to fine-tune the algorithms.

Despite its relative superiority, ChatGPT still requires significant improvements to reach higher levels of accuracy and naturalness in translations. This points to the need for ongoing development and training of LLMs. Developers shall invest in expanding the training datasets for LLMs to include more diverse and context-rich examples. They shall also enhance the models' ability to handle different registers, styles, and idiomatic expressions. They also need to consider future research directions that explore new architectures and techniques to improve translation quality.

The practical comparison using speeches by King Abdullah II underscores the relevance of these findings for real-world applications. The study indicates where current tools stand in handling important and contextually rich texts. Developers shall encourage developing and testing translation tools using real-world texts across various domains. These insights guide improvements and tailor tools for specific applications, such as diplomatic communications, technical manuals, and literary translations.

The findings suggest a potential shift in the translation industry towards a more collaborative model where machine translation aids human translators, enhancing efficiency while maintaining quality. We shall promote collaborative tools and platforms integrating machine translation with human expertise. We shall also develop educational resources and training programs to help translators effectively use these tools and maximize their potential benefits.

By addressing these implications and actions, the study's insights can drive meaningful advancements in developing and applying machine translation tools, ensuring they complement rather than replace human translators.

Conclusion

This study provides valuable insights into the effectiveness of Large Language Models (LLMs) in machine translation, mainly focusing on ChatGPT and contrasting it with traditional tools like Google Translate. By examining both Arabic and English translations of speeches by King Abdullah II of Jordan, the study evaluated the accuracy and quality of translations generated by these tools. The translations were carefully analyzed within their contextual framework using a comprehensive rubric provided by [55], which considers various aspects such as meaning, functional adequacy, style, and overall rating. The analysis revealed significant shortcomings in the translations produced by Google Translate, including misunderstandings, inaccuracies, and biases. While Google Translate provided relatively satisfactory translations in terms of meaning, it often exhibited deviations and occasional changes that affected the overall quality. Similarly, the literal translation practices observed in the English-Arabic translation underscored the need for substantial edits to ensure clarity and coherence. On the other hand, ChatGPT's translations, while also showing deficiencies, were generally rated as acceptable, with minor edits required. Despite some issues in mechanics, style, and register, ChatGPT offered more natural-sounding translations than Google Translate, positioning it as a preferable option for machine translation. While both Google Translate and ChatGPT have their limitations, ChatGPT emerges as a promising alternative, providing translations closer to acceptable with fewer mistakes. However, this study underscores the irreplaceable role of human translation in accurately conveying the textual and contextual nuances of source texts. While machine translation tools continue to advance, they cannot fully replace the human touch required to ensure an accurate and comprehensive understanding of the original content. Therefore, integrating machine translation with human oversight remains crucial for achieving high-quality translations. The study answered the following research questions: Can LLMs be used as computational linguistics tools? Yes, they can, and the translations are acceptable with minimal edits in accuracy. Did MT tools develop to overcome the weaknesses listed in previous research? Yes, they did. However, they still have

several weaknesses regarding misunderstanding, textual inaccuracy, contextual inaccuracy, and biases. Do LLMs provide a better alternative than the traditional MT tools? Yes, they are. Based on the findings of this study, they, especially ChatGPT, provide a better machine translation option.

Conflict of Interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

[1] Lazer, D, Duncan, J. Watts, S, Athey, S, Contractor, N, Freelon, D, Gonzalez-Bailon, S, King, G, Margetts, H. (2020). Computational social science: Obstacles and opportunities. *Science*, *369*(6507):1060–1062. https:// doi.org/10.1126/science.aaz8170,

[2] Lazer, D, Pentland, A, Adamic, L, Aral, S, Barab´asi, A, Brewer, D, Christakis, N, Contractor, N, Fowler, J, Gutmann, M. (2009). Computational social science. *Science*, *323*(5915):721–723. https://doi.org/10.1126/science.

[3] Golder, S. & Macy, M. (2014).Digital footprints: Opportunities and challenges for online social research. *Annual Review of Sociology*, 40:129–152. https://doi.org/10.1146/annurev-soc-071913-04314

[4] Stone, P. J., Dunphy, D, & Smith, M. (1966). *The general inquirer: A computer approach to content analysis*. Chicago, IL: University of Chicago Press.

[5] Bang, Y, Cahyawijaya, S, Lee, N, Su, W, Wilie, B, Lovenia, H, Ji, J, Yu, T, Chung, Y. (2023). A multitask, multilingual, multimodal evaluation of ChatGPT on reasoning, hallucination, and interactivity. ArXiv preprint, abs/2302.04023.

[6] Goyal, T, Li, J, & Durrett, G. (2022). News summarization and evaluation in the era of GPT-3. ArXiv preprint, abs/2209.12356.

[7] Zhuo, T, Huang, Y, Chen, C, & Xing, Z. (2023). Exploring AI ethics of ChatGPT: A diagnostic analysis. ArXiv preprint, abs/2301.12867.

[8] Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., ... & Zhang, Y. (2023). Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint arXiv:2303.12712*.

[9] Chung, H. W., Hou, L., Longpre, S., Zoph, B.; Tay, Y.; Fedus, W., Li, E., Wang, X., Dehghani, M., Brahma, S. (2022). Scaling instruction-finetuned language models. arXivpreprintarXiv:2210.11416.

[10] AlAfnan, M. A., Dishari, S., & Siti Fatimah MohdZuki. (2024). Developing Soft Skills in the Artificial Intelligence Era: Communication, Business Writing, and Composition Skills. *Journal of Artificial Intelligence and Technology*. https://doi.org/10.37965/jait.2024.0496.

[11] Min, B, Ross H, Sulem E, et al. (2021). Recent advances in natural language processing via large pre-trained language models. arXiv 2111.01243 [preprint]. https://arxiv.org/abs/2111.01243.

[12] Vaswani A, Shazeer N, Parmar N, et al. (2027). Attention is all you need. Advances in Neural Information Processing Systems 2017;30. https://papers.nips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

[13] Shen, Y., Heacock, L., Elias, J., Hentel, K. D., Reig, B., Shih, G., & Moy, L. (2023). ChatGPT and other large language models are double-edged swords. Radiology, 307(2), e230163.

[14] Rohrbach A, Hendricks LA, Burns K, Darrell T, Saenko K. (2018). Object hallucination in image captioning. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing 2018; 4035–4045.

[15] Xiao Y, Wang WY. (2021) On hallucination and predictive uncertainty in conditional language generation. In: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume 2021.

[16] Nguyen, D, Oz, A, Rose, E, & Jong, F. (2016). Computational sociolinguistics: A survey. *Computational Linguistics*, *42*(3):537–593. https://doi.org/10.1162/COLI_a_00258

[17] Bucholtz, M., & Hall, K. (2005). Identity and interaction: A sociocultural linguistic approach. *Discourse studies*, *7*(4-5), 585-614.

[18] Del, M & Fernandez, R. (2017). Semantic variation in online communities of practice. In IWCS 2017 - 12th International Conference on Computational Semantics - Long papers. https://aclanthology.org/W17-6804.pdf

[19] Purschke, P, and Hove D . (2019). Lorres, moppes, and the swiss.(re) discovering regional patterns in anonymous social media data. *Journal of Linguistic Geography*, *7*(2):113–134. https://doi.org/10.1017/jlg.2019.10

[20] Preoţiuc-Pietro, D., Lampos, V., & Aletras, N. (2015, July). An analysis of the user occupational class through Twitter content. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* (pp. 1754-1764).

[21] Del, M & Fernandez, R. (2017). Semantic variation in online communities of practice. In IWCS 2017 - 12th International Conference on Computational Semantics - Long papers. https://aclanthology.org/W17-6804.pdf

[22] Bamman, D, Eisenstein, J, & Schnoebelen, T. (2014). Gender identity and lexical variation in social media. *Journal of Sociolinguistics*, *18*(2):135–160. https://doi.org/10.1111/josl.12080.

[23] Piper, A, So, R, and Bamman, D. (2021). Narrative theory for computational narrative understanding. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pages 298–311. https://doi.org/10.18653/v1/2021.emnlp-main.26.

[24] Sap, M, Jafarpour, A, Choi, Y, Smith, N, Pennebaker, J, & Horvitz, E. (2022). Quantifying the narrative flow of imagined versus autobiographical stories. Proceedings

of the National Academy of Sciences, *119*(45):e2211715119. https://doi.org /10.1073/pnas.2211715119.

[25]   AlAfnan, M. A. (2022). Public discourse: Systemic functional analysis of Trump's and Biden's inaugural speeches. *Journal of Language and Linguistic Studies*, *18*(1), 1-14.

[26]   Jockers, M. L., & Mimno, D. (2013). Significant themes in 19th-century literature. Poetics, 41(6), 750-769.

[27]   Philipp, K. & Knowles R. (2017). Six challenges for neural machine translation. In Proceedings of the 1st Workshop on Neural Machine Translation. 28–39.

[28]   Somers, H. (2011). Machine translation: history, development, and limitations. In: Malmkjær K, Windle K (eds). *The Oxford handbook of translation studies*. Oxford University Press, Oxford, pp 427–440. https://doi.org/10.1093/oxfordhb/9780199239306

[29]   Hutchins, W. (1995). Machine translation: a brief history. In: Koerner EF, Asher RE (eds) *Concise history of the language sciences: from the Sumerians to the cognitivists*. Pergamon Press, Oxford, pp 431–445

[30]   AlAfnan, M. A. (2015). Language use in computer-mediated communication: An investigation into the genre of workplace emails. *International Journal of Education and Literacy Studies*, *3*(1), 1-11.

[31]   Almahasees. Z. & Mahmoud, S. (2022). Evaluation of Google Image Translate in rendering Arabic signage into English. *World J Engl Lang 12*(1):185–197. https://doi.org/10.5430/wjel.v12n1p185

[32]   Casacuberta-Nolla, F. & Peris-Abril, A. (2017) Traducción automática neuronal. *Rev Tradum Tecnol Traducció,* 15:66–74. https://doi.org/10.5565/rev/tradumatica.203

[33]   Crivellari A, & Beinat, E, (2020) Trace2trace—a feasibility study on neural machine translation applied to human motion trajectories. *Sensors, 20*(12): 3503. https://doi.org/10.3390/s20123503

[34]   Klimova, B, Pikhart, M, Benites, A, Lehr, C, & Sanchez-Stockhammer, C. (2022). Neural machine translation in foreign language teaching and learning: a systematic review. E*duc Inf Technol* 2022:1–20. https://doi.org/10.1007/s10639-022-11194-2.

[35]   AlAfnan, M. A. (2022). Uniting for peace: A speech act analysis of the United Nations General Assembly Resolution 377 A (V). *World*, *12*(6).

[36]   Yamada M (2019) Language learners and non-professional translators as users. In: Ohagan M (ed). *The Routledge handbook of translation and technology*. Routledge, New Yok, pp 183–199. https://doi.org/10.4324/9781315311258-11

[37]   Zhao, L, Gao, W, & Fang, J. (2021) High-performance English-Chinese machine translation based on GPU-enabled deep neural networks with domain corpus. *Appl Sci, 11*(22):10915. https://doi.org/10.3390/app112210915

[38]   Fuentes-Luque, A, & Santamaría-Urbieta, S. (2020). Machine translation systems and guidebooks: an approach to the importance of the role of the human translator. *Onomázein*. https://doi.org/10.7764/onomazein.ne7.04

[39]   Stewart, D. (2019). English for tourism in the non-native English classroom: machine translation and corpora. In: Ennis MJ, Petrie GM (eds) *Teaching English for tourism: bridging research and praxis.* Routledge, Amsterdam, pp 114–130. https://doi.org/10.4324/9780429032141

[40]   AlAfnan, M. A., & Dishari, S. (2024). ESD goals and soft skills competencies through constructivist approaches to teaching: an integrative review. Journal of Education and Learning (EduLearn), 18(3), 708-718.

[41]   Athanasiou, V, & Maragoudakis, M. (2016, September). Dealing with high dimensional sentiment data using gradient boosting machines. In: *IFIP international conference on artificial intelligence applications and innovations*, pp 481–489. Springer, Cham. https://doi.org/10.1007/978-3-319-44944-9_42

[42]   Kalita, I. (2016). Notes on the Life of Words "Hairy (shaggy)" Dumplings [О жизни слов.«Волосатый (лохматый)» кнедлик]. *XLinguae, 9*(3):2–24. https://doi.org/10.18355/XL.2016.09.03.2-24

[43]   Toral, A, Esplá-Gomis, M, Klubička, F, Ljubešić, N, Papavassiliou, V, Prokopidis, P, & Way, A. (2017). Crawl and crowd to bring machine translation to under-resourced languages. *Lang Resourc Eval*, *51*(4):1019–1051. https://doi.org/10.1007/s10579-016-9363-6

[44]   Sen, S, Hasanuzzaman, M, Ekbal, A, Bhattacharyya, P, & Way, A. (2021). Neural machine translation of low-resource languages using SMT phrase pair injection. *Nat Lang Eng, 27*(3):271–292. https://doi.org/10.1017/S1351324920000303

[45]   Zhao, L, Gao, W, & Fang, J. (2021) High-performance English-Chinese machine translation based on GPU-enabled deep neural networks with domain corpus. *Appl Sci, 11*(22):10915. https://doi.org/10.3390/app112210915

[46]   Ata, M, & Debreli, E. (2021). Machine translation in the language classroom: Turkish EFL learners' and instructors' perceptions and use. *IAFOR J Educ, 9*(4):103–122. https://doi.org/10.22492/ije.9.4.06

[47]   Deng, X, &Yu, Z, (2022), A systematic review of machine-translation-assisted language learning for sustainable education. *Sustainability, 14*(13):7598. https://doi.org/10.3390/su14137598

[48]   Vazquez-Calvo, B, & Cassany, D. (2017). Aprender lengua con el traductor automático en la escuela secundaria: un diálogo necesario. *Calidoscópio, 15*(1):180–189. https://doi.org/10.4013/cld.2017.151.14

[49]   AlAfnan, M. A. (2018). Language use as a resource: E-communication dimensions of register variation in a professional context. *Journal of Progressive Research in Social Sciences*, *7*(2), 502-526.

[50]   Vazquez-Calvo, B, & Cassany, D. (2017). Aprender lengua con el traductor automático en la escuela secundaria: un diálogo necesario. *Calidoscópio, 15*(1):180–189. https://doi.org/10.4013/cld.2017.151.14

[51]   AlAfnan, M. A., & MohdZuki, S. F. (2023). Do artificial intelligence chatbots have a writing style? An investigation into the stylistic features of ChatGPT-4. *Journal of Artificial intelligence and technology*, *3*(3), 85-94.

[52]   Ram B Misra, & Li, Y. (2024). A Constructive Use of ChatGPT in the Classroom: An Empirical Study. *Journal of Artificial Intelligence and Technology*. https://doi.org/10.37965/jait.2024.0485

[53]   Michael E. Bernal. (2024). Revolutionizing eLearning Assessments: The Role of GPT in Crafting Dynamic Content and Feedback. *Journal of Artificial Intelligence and Technology*. https://doi.org/10.37965/jait.2024.0513.

[54]   Schäferhoff, N. (2023). *Best machine translation software: 7 excellent solutions to try out*. Translation Press. Retrieved from: https://translatepress.com/machine-translation-software/

[55]   Ustaszewski, M. (2014). Towards a methodology for intercomprehension-based language instruction in translator training. *University of Innsbruck, Innsbruck.*