Traduction et Langues Volume 25 Numéro 01/2026
Journal of Translation Languages **مجلة الترجمة واللغات**
ISSN (Print): 1112-3974 **EISSN (Online): 2600-6235**

Exploring the Role of AI in Diplomacy-related Translation Tasks: An Empirical Study on ChatGPT's Potential Use in Diplomatic Texts' Translation

Yaşar Akgün
Izmir University of Economics - Turkey
yasar.akgun@ieu.edu.tr
Nihal Yetkin-Karakoç
Izmir University of Economics – Turkey
nihal.yetkin@ieu.edu.tr
Mirza İşi
Izmir University of Economics – Turkey
mirza.isi@std.ieu.edu.tr

To cite this paper:
Akgün, Y., Yetkin-Karakoç, N.,& İşi, M. (2026). Exploring the role of AI in diplomacy-related translation tasks: An empirical study on ChatGPT's potential use in diplomatic texts' translation. *Traduction et Langues, 25(01*), 135-156.

Received: 13/04/2025; **Accepted**: 15/01/2026, **Published:** 30/06/2026

Corresponding author: Yaşar Akgün

Keywords

Diplomatic Translation; ChatGPT; Large Language Models (LLMs); AI-Assisted Translation; Diplomatic Discourse; Translation Quality Assessment; English–Turkish Translation

Abstract

Diplomatic language, with its culturally sensitive and formally stratified character, presents distinctive challenges for translation. Although artificial intelligence (AI) and large language models (LLMs), such as ChatGPT, have shown considerable potential in technical and pragmatic translation tasks, their applicability to the specialised domain of diplomatic translation remains underexamined. This study addresses this gap by empirically assessing ChatGPT's performance in translating diplomatic texts in comparison with human translators. Three diplomatic texts—two translated from English into Turkish and one from Turkish into English—were assigned to five third- and fourth-year undergraduate students enrolled in the English Translation and Interpreting Department at Izmir University of Economics, all of whom had completed a course in political text translation. The same texts were translated by ChatGPT (GPT-4o) using the prompt: "Translate the following texts with a diplomatic tone." To ensure blind evaluation, all translations were handwritten and anonymised before submission. A jury consisting of two scholars in international relations and one retired ambassador evaluated the translations using a five-criterion rubric. ChatGPT obtained the highest overall score among the six participants and outperformed the five human translators across all three texts. The margin was narrowest in Text 3, translated from Turkish into English, yet ChatGPT still achieved the highest score. This result may be attributed to the formulaic and template-based nature of diplomatic correspondence, which aligns with the pattern-recognition capacities of LLMs trained on large, English-dominant datasets. The study suggests that ChatGPT has substantial potential for written diplomatic translation, particularly in texts conforming to established diplomatic conventions. These findings have implications for translation pedagogy, diplomatic communication, and the integration of AI tools into professional translation workflows. Future research should extend the inquiry to interpreting contexts and to more rhetorically complex diplomatic texts.

Available online at https://www.asjp.cerist.dz/en/Articles/155

Anahtar Kelimeler

Çeviri Diplomasisi; ChatGPT; Büyük Dil Modelleri (BDM'ler); Yapay zekâ destekli çeviri; Diplomatik söylem; Çeviri kalitesi değerlendirmesi; İngilizce–Türkçe çeviri

Öz

Diplomatik dil, kültürel açıdan hassas ve resmî olarak katmanlı yapısıyla çeviri açısından belirgin zorluklar ortaya koyar. Yapay zekâ ve ChatGPT gibi büyük dil modelleri teknik ve pragmatik metin çevirilerinde önemli bir potansiyel göstermiş olsa da diplomatik çeviri gibi uzmanlık gerektiren alanlardaki uygulamaları yeterince araştırılmamıştır. Bu çalışma, ChatGPT'nin diplomatik metin çeviri performansını insan çevirmenlerle karşılaştırmalı olarak inceleyerek bu boşluğu ele almaktadır. İzmir Ekonomi Üniversitesi İngilizce Mütercim-Tercümanlık Bölümü'nde öğrenim gören ve siyasi metin çevirisi dersi almış beş üçüncü ve dördüncü sınıf öğrenciye ikisi İngilizceden Türkçeye, biri Türkçeden İngilizceye olmak üzere üç diplomatik metin verilmiştir. Aynı metinler ChatGPT (GPT-4o) tarafından "Aşağıdaki metinleri diplomatik bir üslupla çevirin" yönergesiyle çevrilmiştir. Tüm çeviriler anonimleştirilmiş ve kör değerlendirme için hazırlanmıştır. Değerlendirmeler, uluslararası ilişkiler alanında iki akademisyen ve bir emekli büyükelçiden oluşan jüri tarafından beş ölçütlü bir rubrikle yapılmıştır. Sonuçlar, ChatGPT'nin tüm katılımcılar arasında en yüksek toplam puanı aldığını ve üç metinde de insan çevirmenleri geride bıraktığını göstermiştir. Bu durum, diplomatik metinlerin büyük ölçüde kalıplaşmış ve şablon temelli yapısına bağlanmaktadır. Bulgular, ChatGPT'nin özellikle yerleşik diplomatik kurallara uygun yazılı çeviri görevlerinde güçlü bir potansiyele sahip olduğunu göstermektedir ve çeviri eğitimi ile profesyonel uygulamalar açısından önemli çıkarımlar sunmaktadır.

1. Introduction

Diplomacy aims at resolving the prospects of disputes between the nations and ideally cements the cordial relations so that communication is not derailed. It is driven by reciprocity, implying that each nation necessitates to be engaged taking its addressee's set of values, culture and norms into consideration. Since each nation possesses unique ways of perceiving and tackling the issues, states engaged in diplomacy aspire to be well-informed and cognizant with regards to each other's Achilles' Heel in order to conduct a precise communication. Pokhrel (2020) states that: "… it seems to be evident that efficient use of diplomatic language demands that its practitioners be interculturally competent enough to understand the core values of the culture(s) they work in or with" (p. 181). The ability to maintain a fruitful and constructive dialogue with a state rooted in vastly different geographic and cultural ideas, depends on the linguistic acumen of the conductors – that is – diplomats. Indeed, the interlocutors' use of language in diplomacy appears to be sui generis, drawing scholars' attention and leading to many studies delving into its defining characteristics (Al Salem et al., 2026; Martirosyan, 2022; Bobeica, 2022; Yetkin-Karakoç, 2022; Pokhrel, 2020; Acquisto, 2017).

Even though diplomats are highly-educated professionals equipped with abilities to talk to their foreign interlocutors in a common language, they prefer to adhere to their

mother tongue to highlight independence and prioritize their national identity[1]. Thus, it can be argued that throughout the history, both translators who have conducted the correspondence between the diplomats and interpreters who have actively taken part in the diplomatic settings have proved to be sine que non in the international relations. This fact reinforces the argument that translation is not merely about converting words from one language to another, but about fostering understanding, cooperation, and peace among nations. Yet, given the rise and development of AI in the translation studies, the field has been revolutionized be integrated in a highly competitive market. Indeed, the growing usage of AI in the translation market has occurred in tandem with the growing demand of translation tasks (Çetiner, 2019, pp. 110-11; 2018, p. 153).

In a globalized world characterized by increased volume of trade and business related documents and technical texts, translators have needed come-in-handy means to meet and tackle competitive business demands. As Vukalović states (2021), translators are bound to cooperate with technology to survive in a competitive market. However, it needs to be accentuated that the type of documents translated in the market is predominantly business and technical. In 2002, Kingscott stated that 90% of translation output on a global level was technical documents (p. 247). Likewise, B (2020) pointed out that finance-related documents have risen to dominate the translation market especially for the last several years due to globalization (p. 552). In this respect, AI-driven applications have proved to be useful especially for technical and business-related documents where the repetitions and terminological phrases are the norm. Kingscott (2002) goes on to state that:

> Essentially, of course, the idea is to save the translator having to re-translate any segment of text, large or small, for which a translation already exists. Many technical documents contain repetitions, and there are immense savings to be had if the previous translation can be generated (p. 253).

Even though AI-driven applications can be utilized by professional translators to mitigate the burden of fast-circulating translation tasks, it is argued that they are frequently applied to the technical and business texts (Mercan et al., 2024), which are summarized as "pragmatic texts" in Odacıoğlu (2021). It is stated that technical texts are accuracy-based, impersonal and unemotional (Bowker, 2023). The effectiveness of the use of AI-driven applications and machine translation may vary in accordance with the type of the texts being translated. Although texts that are rife with repetitions and terminological

[1] For more details on the extra-linguistic use of interpreters see: Akgün, Y., & Yetkin Karakoç, N. (2024). The Different Aspects of Diplomatic Mediation: The Extralinguistic Motives of Having an Interpreter in the Diplomatic Settings. *İletişim Ve Diplomasi* (12), 75-89. https://doi.org/10.54722/iletisimvediplomasi.1488687

phrases are more suitable for AI and machine translation, delicately articulated texts – diplomatic language being one of the examples- may not be properly translated through mentioned means. Therefore, the investigation of the prospects of the use of AI for delicate language translation within the scope of diplomatic language has been assessed to be potentially lucrative for the observed literature gap in the field of translation studies. Within this scope, the present study, which -to the best of the authors' knowledge, is the first of its kind, aims at filling an assessed literature gap in the field. However, before proceeding with the characteristics of diplomatic language and method in this study, the working principle of Large Language Models through which ChatGPT is coded needs to be clarified.

2. Large Language Models

Large language models (LLMs) have changed natural language processing (NLP) by enabling machines to engage with human language in sophisticated and complex ways. Indeed, the concept of artificial intelligence (AI) dates back to Alan Turing's question, "Can machines think?" in his 1950 paper (Turing, 1950, pp. 433–435). However, the recent boom in artificial intelligence is driven by the development of large language models such as GPT-3 and GPT-4. These models are trained on massive datasets and are capable of generating human-like text in various fields (OpenAI, 2023, pp. 4–6; Brown et al., 2020, pp. 1–3; Chang et al., 2024). In earlier stages, humans had to communicate with machines using programming languages because computers could not understand human language as it is. However, today's advancements in artificial intelligence enable users to interact with machines using everyday language (OpenAI, 2023, p. 4; Chang et al., 2024).

The key thing to comprehend the principle at work is that the conceptual foundation of LLMs is not based on symbolic manipulation or deterministic grammatical parsing. It rather uses probabilistic modelling, pattern recognition, and distributed representation learned from massive corpora. Simply put, an LLM is a very complex mathematical function that predicts what word comes next for any piece of text, based on the patterns it has learned during training. As Hutchins and Somers (1992, pp. 5–8) argue, these models do not understand language like us humans do. Instead, they operate by associating linguistic patterns based on statistical frequencies. This gap between prediction and understanding has been a central philosophical concern. Searle's Chinese Room argument (1980) illustrates the addressed issue. Imagine a human in a room; his task is to translate Chinese texts to English, but he does not speak any Chinese. However, he has a very complex dictionary that hypothetically includes almost all words and phrases from Chinese and their counterparts in English. This human can translate the texts using the dictionary, but does it mean that he speaks and understands Chinese? This thought experiment tells us that even if a system manipulates symbols well enough to appear fluent in a language, it does not imply that the system understands the language in any meaningful sense (Searle, 1980, pp. 417–418).

2.1 How do LLMs Work?

A fundamental step in the process is named tokenisation. Before any meaningful computation can happen, language input must be separated into smaller, more manageable, and meaningful units. For example, BERT uses WordPiece tokenisation to break down text into subword units, each of which is then transformed into what is known as an embedding (Devlin et al., 2019, p. 3). After that, these embeddings are placed into high-dimensional vector spaces where their relationships are shaped through training on vast amounts of text (Devlin et al., 2019, p. 3). Importantly, these representations are highly context-dependent (Devlin et al., 2019, pp. 4–5). The meaning of a word like "bank" will shift depending on whether it appears alongside the word "river" or "loan". It is also powered by the self-attention mechanism used in transformer models.

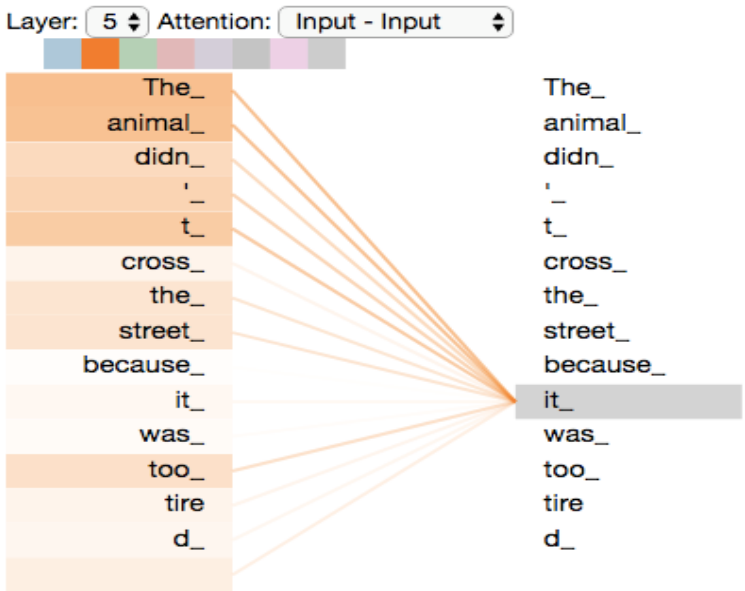

Figure 1. Self-attention at a high level. Reprinted from The Illustrated Transformer by J. Alammar (2018), retrieved from https://jalammar.github.io/illustrated-transformer/; licensed under CC BY-NC-SA 4.0

At the core of LLMs lies the math's. There is an interaction of linear algebra and probabilistic modelling. Each token is represented as a high-dimensional vector, and then these vectors are multiplied by learned weight matrices to extract and transform semantic information. As described by Vaswani et al. (2017, p. 4), transformer architectures use query, key, and value matrices to compute attention scores to determine how much focus a token should place on others in the sequence. This can be accomplished through scaled dot-product attention, followed by matrix multiplication and softmax functions that

normalise these attention weights. All of these mathematical operations allow the model to simulate linguistic dependencies without predefined grammar rules. The resulting representations pass through multiple layers; each layer refining the context. In modern LLMs, by distributing computation across multiple heads and layers, attention can capture both local and global patterns in parallel, enabling remarkable fluency and coherence even over long texts.

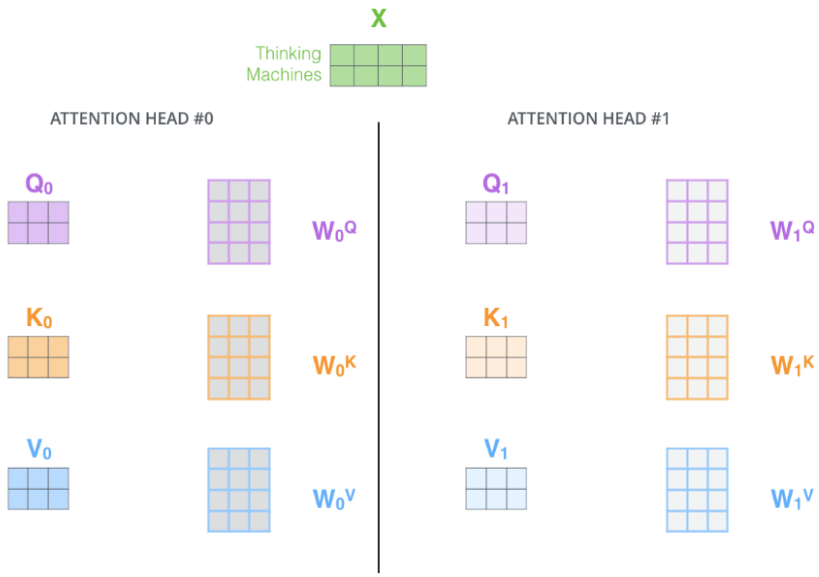

Figure 2. The beast with many heads (multi-headed attention). Reprinted from The Illustrated Transformer by J. Alammar (2018) https://jalammar.github.io/illustrated-transformer/; licensed under CC BY-NC-SA 4.0

A breakthrough was when the transformer architecture marked a significant turning point in AI history. Vaswani et al. (2017, p. 2) introduced a model that processes all input tokens simultaneously using a mechanism called self-attention. This self-attention mechanism enables the model to weigh the relevance of each word in the sequence in relation to every other word, allowing for the computation of global dependencies regardless of token position. The process can be thought of as arrows on text going from each word to every single other word, and on top of them, their relationship is defined. That is, every word in the text influences the rest of the words. The result is a flexible and context-sensitive representation of meaning that does not depend on strict word order. This idea appears to be crucial because it was a major departure from earlier architectures like recurrent neural networks (RNNs) and long short-term memory networks (LSTMs) (Vaswani et al., 2017, p. 2).

2.2 LLMs and Translation

The introduction of the attention mechanisms by Bahdanau et al. (2015, pp. 2–5) is another important event. Their model enabled the decoder to selectively focus on relevant parts of the source sentence during translation, which addressed the bottleneck caused by compressing an entire sentence into a single vector. This flexibility, introduced by attention, enabled more accurate alignment between source and target languages. As mentioned, transformer models then integrated multi-headed attention, increasing the model's ability to detect different linguistic features in parallel (Vaswani et al., 2017, p. 5).

Johnson et al. (2017, p. 1) showed that a single multilingual neural machine translation (NMT) model can handle multiple language pairs using a shared encoder–decoder architecture. Instead of training a separate model for each language direction, they came up with the idea of introducing a system in which the target language is indicated by a token at the beginning of the source sentence (Johnson et al., 2017, p. 1). This simple yet effective strategy enables the models to generalise beyond language pairs they had seen before (Johnson et al., 2017, p. 2). Remarkably, the very same system was able to perform translations between language pairs it had never encountered during training, which the authors referred to as "zero-shot translation" (Johnson et al., 2017, pp. 4–5). It can be said that Hutchins and Somers (1992, pp. 74–75) had discussed the idea of an intermediate representation as a theoretical construct in machine translation, and Johnson et al. (2017) offer empirical evidence of such a mechanism stemming from data-driven training.

GPT-3 and GPT-4 show the scale-driven evolution of LLMs. These models are trained with autoregression on massive datasets, which allows them to generate coherent and contextually appropriate responses from even minimal prompts (Brown et al., 2020, p. 1; OpenAI, 2023, pp. 1–2). The GPT-4 technical report showcases significant improvements in machine translation performance, particularly for high-resource languages such as English, German, and French (OpenAI, 2023, pp. 1–4). However, it also acknowledges that the model displays performance gaps in low-resource settings and struggles with specialised domains that require precise factual grounding or domain-specific terminology (OpenAI, 2023, p. 10).

Another important issue is hallucination. As the name suggests, it refers to the concept of an LLM hallucinating an idea or information. This could either mean that the LLM thinks something that does not exist actually exists, or something that is incorrect is actually correct (Huang et al., 2023, p. 6). The generated content is most probably entirely grammatically correct but may factually be incorrect. This can be especially problematic in translation as even small inaccuracies may distort the meaning. As noted by Johnson et al. (2017), zero-shot translation quality is lower than trained or bridged directions, particularly for distant language pairs. Moreover, omissions of words or phrases can turn out to be a significant issue as it is harder to detect than an incorrect piece of information, which can be easily detected by an expert.

Interpretability also presents certain challenges. While attention maps can suggest how information flows through a model, they do not reliably explain decision-making. As Searle (1980, p. 417) emphasises, statistical systems do not "understand" language in a semantic sense; rather, they rely on probabilistic associations and patterns derived from data (Wu et al., 2016, p. 2). This distinction between intelligent look-up and truly understanding language was one of the concerns raised by Turing in his seminal imitation game proposal. He argued that the ability to successfully mimic human responses could be a sufficient test for intelligence. However, the Turing Test remains a controversial issue in the field, particularly in light of systems that pass the test without any actual comprehension (Turing, 1950, pp. 433–435).

In recent years, the integration of LLMs and other LLM-based tools into professional workflows has been swift and life-changing. Tools like Google Translate, DeepL, and ChatGPT have become useful to translation workflows. However, while these systems objectively enhance convenience, they are also known to require caution. An important point also made by OpenAI CEO Sam Altman, who recently expressed surprise at how much people trust ChatGPT despite its tendency to hallucinate. He said, "People have a very high degree of trust in ChatGPT, which is interesting, because AI hallucinates. It should be the tech that you don't trust that much" (Altman, 2025; Huang et al., 2024). Likewise, Al Salem et al., (2026) point out that despite improvements in fluency and adequacy, these models may still miss specific nuances or generate misleading translations, especially in complex or specialised domains such as diplomacy. As Almahasees et al. (2025) state:

> In the modern era, characterized by the sudden emergence of artificial intelligence-driven translation programs in various media and diplomatic contexts, such programs must provide translations that are not only contextually correct but also unbiased, thereby upholding the integrity of political communication. (p. 12)

In short, large language models have transformed how computers perceive and handle language. They have not been set by truly comprehending the language but by detecting and replicating patterns in huge datasets. They produce fluent and convincing text, yet it can be said that their output is based on correlations, not comprehension. This matters, especially in areas like translation, where accuracy and context are key factors. Finally, as these models become part of everyday tools, it is important to recognise both their strengths and limits.

3. Characteristics of Diplomatic Language

Diplomats are bound to assume a delicate and prudent tone in their interstate communication not only to preserve the prestige of their own nations but also constitute an atmosphere where his/her interlocutor feels their nation's prestige is respected. As Al-

Azzawi points out (2019), "The choice of suitable words is very significant in diplomacy, because the language may contribute to communication problems if the diplomat does not give the subtle meaning to the words which he/she uses it in diplomatic language" (p. 15). Therefore, Abu Jaber appears to be justified in stating that ddiplomatic language is the child of the language of communication (2001).

The nations may readily find themselves at loggerheads due to wide range of conflicts pertaining to military, trade, economy and culture related issues. Since diplomacy prioritizes a peaceful solution, offensive language is avoided so that the process is not eclipsed. As Martirosyan states, hurdles are prone to being expressed not directly but rather through different naming processes in diplomacy (2022, p.130). Thus, naming the situation as is, is not one of the *modus-operandi* of diplomacy. The language sounds both fascinating and highly sophisticated, but in reality does not encapsulate the reality within itself (Bobeica, 2022). That is why, it might be argued that diplomatic language is a double-filtered version of language (Yetkin-Karakoç, 2022). Indeed, diplomatic language is structured to be stratified, with each layer essential for preventing any culturally or politically provocative expression. It follows that strategically vague use of language is among the required layers, with politeness observed in all stages of diplomacy for retaining a room for face-saving.

To abstract utterances, inclusion of figures of speech is not infrequent. With meticulously chosen words, lexical ambiguity containing concise phrases and politely manifested etiquette, stratification of diplomatic language is optimally completed, on the way to ensuring smooth relations between nations.

4. Method

4.1 Research Design and Procedure

This study is empirical in that it compares the diplomatic texts' translations of human translators with ChatGPT. In order to attain adequate data, the number of participants was designated as six, one of which was ChatGPT. Thus, five students from the Department of English Translation and Interpreting at Izmir University of Economics consented to take a part in the study. Students were in the third and fourth grade at the department, having successfully passed the course named Translation of Texts on Political Science. All the students passed the mentioned class with the grade of AA. Each student was handed out three same texts, which reflected a diplomatic tone, register and content.

During the study, it was made sure that students who were accompanied by an invigilator did not have any access to electronic devices but dictionary and glossary they prepared counting on the terminology stressed during the lessons. Their translations were collected to be compared with ChatGPT's outputs. The same texts were given to ChatGPT with the prompt of "Translate the following texts with a diplomatic tone". GPT-4o version was utilized while generating translations.

Having collected ChatGPT's translations, both students' and ChatGPT's translations were prepared for blind-review. Students translated the texts on the paper by their

hands. Thus, to anonymize ChatGPT, the translations done by it were handwritten on the paper so that the jury would not recognize ChatGPT. Translations were then submitted to a jury consisting of three members. The jury was not informed that the participant 3 was indeed ChatGPT. While two of the jury members were working in the field of international relations affiliated with different universities, the last jury member was a retired ambassador. The jury was also submitted a rubric through which they could evaluate the translations out of five points (see Appendix, Table 1, for rubric). They were also let to write their comments on the translations, where they deemed necessary. Having collected evaluations, their grades were compared with each other in order to craft an overall order.

3.2 Selection of the Texts

During the process of preparing the texts that would be translated, the book named *Diplomatik Yazışma ve Konuşma Metinlerinden Örnekler* (Examples from Diplomatic Correspondence and Speech Texts) (Tulun, 1996) was applied in order to extract out diplomatic phrases authentically used in different occasions such as condolences, congratulations and invitations. Text 1 and Text 2 were created by applying some of the most common patterns in diplomatic correspondence, as mentioned in *Examples from Diplomatic Correspondence and Speech Texts* (Tulun, 1996). Text 3, which is the only Turkish text, was taken from *Immunities and Privileges Booklet* prepared by Aylin Taşhan on behalf of the Turkish Ministry of Foreign Affairs[2].

Whereas Text 1 and Text 2 (correspondence between diplomats) had to be translated from English to Turkish, Text 3 (a diplomatic notice) was from Turkish to English. Thus, two of the source texts were in English while the only Turkish source text was Text 3. The texts were structurally simple and devoid of euphemism and ambiguity (See Appendix, Texts).

3.3 *Limitations*

- The number of participants of the study was designated to be six, five of which were human translators and the Participant 3 was ChatGPT. Increasing the number of participants would enable varied results.
- Texts were structurally plain and devoid of ambiguous patterns, potentially leading ChatGPT to operate better given that there appeared to be no implicit meanings that could only be deciphered through *read-between-the-lines.*
- ChatGPT's translation competence ought not to be limited to written translation. Indeed, in sensitive occurrences entailing human interaction, interpreting competence needs to be accentuated. Therefore, the success of AI-driven applications used in interpreting services may be compared to human interpreters

[2] Available at: https://www.mfa.gov.tr/disisleri-bakanligi-mensuplari-ve-emekli-buyukelcilerin-eserleri.tr.mfa

5. Results and Discussion

When five students translated the texts handed out to them, ChatGPT was requested to translate the same texts with a diplomatic tone, thus, being the sixth participant of the study. Collected translations were prepared for a blind-review and submitted to the jury, consisting of two scholars working in the field of international relations and one retired ambassador. When the jury's grading process finalized, overall translation scores of the participants were obtained. It was observed that Participant 3, which was ChatGPT, ranked as the outperforming translator of the diplomatic texts among other participants. Participant 4 had the second highest score with a slightly lower score compared to ChatGPT.

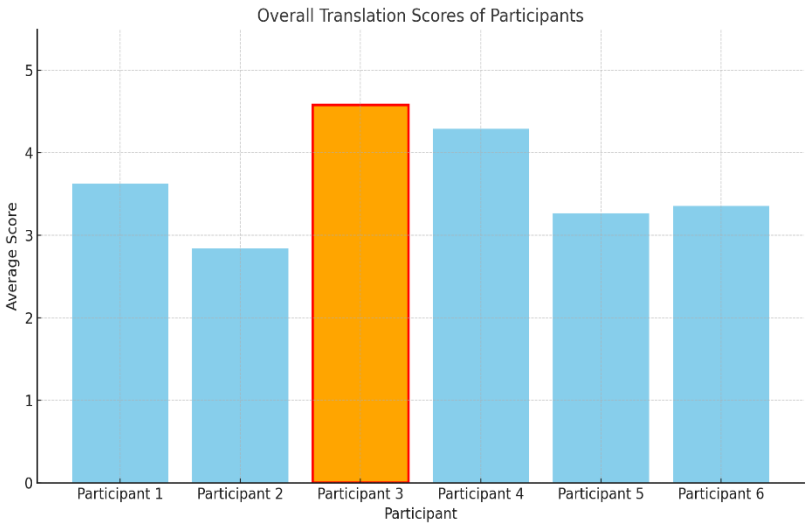

Figure 1. Overall Translation Scores of Participants

Although it appears to be clearly-put that ChatGPT performed better compared to five human participants on an overall scores analysis, it might be crucial to take text-based analysis into account. As it is mentioned above, Text 1 and Text 2 needed to be translated from English to Turkish, while Text 3's direction was from Turkish to English. As can be observed below, ChatGPT performed better in all three texts. Whereas ChatGPT outscored human translators in Text 1 and 2, Text 3 seems to be the only task, where two poles drew nearest, yet, with ChatGPT eventually prevailing.

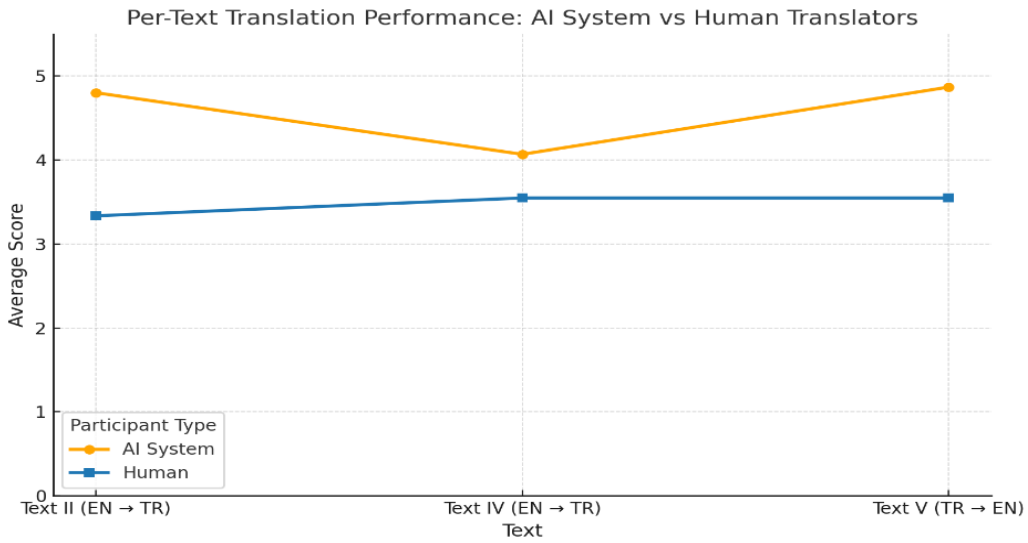

Figure 2. Per-Text Translation Performance: AI System vs Human Translators

Although diplomatic language is full of cultural nuances and patterns, its forms and rules of writing are well-structured; that is to say, diplomatic correspondence is conducted through on an established diplomatic acquis of communication. Indeed, Berin Tulun, who had served as an officer within the Turkish Ministry of Foreign Affairs, set off to prepare a handbook for novice officers struggling to adapt to a diplomatic correspondence. In the Introduction of her handbook named Diplomatik Yazışma ve Konuşma Metinlerinden Örnekler (Examples from Diplomatic Correspondence and Speech Texts) (1996) she states that:

> ...writing a message and making it ready to be sent was a time-consuming and tedious task. Due to the heavy workload, my somewhat impatient nature, and also because I felt uncomfortable not being able to complete a task assigned to me within at most one day of its delegation, I began looking for practical methods that would save me the trouble of "reinventing the wheel" each time. In this context, I tried to create a reference file by collecting examples of diplomatic correspondence and speech texts that I had either written myself or come across, after removing the parts containing confidential information or identifying the addressee. (p.8)

The reason why Tulu decided to prepare a handbook of diplomatic correspondence might be attributed to the fact that diplomatic writing tends to be tackled with an accustomed set of rules, mostly giving way to correspondence templates to be used by officers. Likewise, Erdil. K. Akay, who was the Director of the Training Center in the Turkish Foreign Ministry, wrote a foreword for the same handbook and advised readers

with regards to the optimal use of the handbook:

> When using a sample message as a reference, find and review a message that has been sent in recent years to a foreign figure on a similar occasion. In doing so, you help prevent the sending of a message that is almost identical to one previously sent to the predecessor of the same individual (p.7).

Thus, it can be argued that the repetitive use of similar correspondance templates enables interlocutors to apply for diplomatic acquis. In a similar vein, given that artificial intelligence performs better in technical and pragmatic texts (Odacıoğlu, 2021), ChatGPT's prevailing performance compared to human outputs might be explained by its well-improved reasoning skill drawing on its large accumulated dataset. In a research conducted by Almahasees et al. (2025), DeepSeek and ChatGPT's translation competence in English-Arabic translation of political texts was compared, which concluded that both of the programs performed adequately, with ChatGPT appearing to be better. Given that diplomatic and political texts are perceived on a common reception ground, it might be feasible to claim that ChatGPT and similar AI programs are equipped with fundamental skills to translate diplomatic and political texts due to their accustomed patterns of diplomatic and political acquis of writing. Al Salem et al. (2026) state that AI can be utilized in order to translate political documents with the aid of human translators, so that the process might be expedited. In the same vein, Vraila (2024) claims that AI can "effectively" translate political texts. Likewise, Jia and Jia (2024) find in their empirical research that:

> While human raters achieve a perfect 100% accuracy in literary texts, their accuracy drops to 70% when identifying political texts, which is significantly lower than the machine's 94% accuracy. Consequently, the human raters' accuracy ends up being approximately 10% below GPT-4's accuracy on the overall average accuracy. Even when calculating the mean accuracies of GPT-3.5 and GPT-4, machines still achieve higher identification accuracies than humans in the case of political texts. (p.14)

Zhou et al. claim that since the data ChatGPT draws on happens to be English-dominated, its ability to perform in English-related tasks might be better functioning (p. 115, 2025). Likewise, study conducted by Özcan (2024) demonstrates that eTranslation, which is AI-supported machine translation utilized in EU, performs better when English is taken as the interlanguage. That is to say, output tends to be limping when input is merely Turkish. That English happens to be a language of international relations as lingua franca might be attributed to the fact that ChatGPT generates well-established output in diplomatic texts. Furthermore, it is underscored in the same work that, "The model can

distinguish multiple registers (formal, casual), adjust to varied tones (e.g., technical, literary), and suggest idiomatic words based on huge, diverse text corpora. This assistance improves efficiency and helps translators navigate intricate linguistic structures" (p.115).

With that being stated, it ought to be addressed that diplomatic relations conducted via written means and spoken means differ from each other. That is to say, diplomatic verbal communication necessitates interlocutors to possess specific diplomatic etiquette, through which awareness towards the target culture is resolved and conducted with optimized prudence. Therefore, translating written diplomatic documents and interpreting diplomatic speech entails a different set of standards. Therefore, it would be of vital importance to claim that the results obtained in this study is not likely to be in tandem with the alternative study delving into the performance of AI-powered interpreters and human interpreters serving in diplomacy.

6. Conclusion and Recommendations

Executed with five human translators and one AI translator, the present pioneering study aims at scrutinizing the abilities of AI-powered ChatGPT in translating diplomatic texts. The results yielded by the jury consisting of two scholars working in the field of international relations and one retired ambassador indicate that ChatGPT prevails over the five human translators based on an evaluation crafted with five criteria given in Appendix Table 1. Even though the results may turn out to be unexpected, given the fact that an ideal diplomatic text is handled with cultural nuances, thus, enabling translators as human agencies to perform better, the set of standards for correspondence assumed in diplomatic writing might be a catalyst in the results favoring ChatGPT. It should also be taken into account that the texts were chosen to exclude all the rhetorical ambiguities which may result in a more complicated diplomatic text.

The form and pattern of diplomatic texts constitute a modus operandi with which such texts are prepared, hinting that a wide dataset in English wielded by ChatGPT may have ratcheted its performance up during the process of translation. Yet, it must be taken into account that ChatGPT's observed performance in written translation may not apply to the process of interpreting, which is laden with social interaction necessitating human agencies. In the final analysis, it seems to be feasible to assert that world knowledge and cultural understanding will eventually need to be internalized by AI systems to operate both in written and spoken interactions, just like needed by humans (Song and Kao, 2025). Thus, further research is needed in order to constitute a solid literature projecting the future of AI in diplomacy-related texts and occasions, whether spoken or written.

References

Akgün, Y., & Yetkin Karakoç, N. (2024). The different aspects of diplomatic mediation: The extralinguistic motives of having an interpreter in diplomatic settings. *İletişim ve Diplomasi*, *12*, 75–89. https://doi.org/10.54722/iletisimvediplomasi.1488687

Al Salem, M. N., Alghazo, S., Abusalim, N., & Rabab'ah, G. (2026). How AI translation personas reframe diplomatic texts. *Studies in Media and Communication, 14*(2), 198–212. https://doi.org/10.11114/smc.v14i2.8391

Almahasees, Z., Al-Harahsheh, A., & Albudairi, Y. (2025). Assessing DeepSeek and ChatGPT in English-Arabic translation of political texts. *Research Journal in Advanced Humanities*, *6*(1). https://doi.org/10.58256/jrdwr620

Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. In *Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015)*. https://arxiv.org/abs/1409.0473

Bobeica, G. (2022). Diplomatic language as part of the language for specific purposes. *Études Interdisciplinaires en Sciences Humaines, 8*, 600–610.

Bowker, L. (2019). Fit-for-purpose translation. In M. O'Hagan (Ed.), *The Routledge handbook of translation and technology* (pp. 453–468). Routledge.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901. https://arxiv.org/abs/2005.14165

Çetiner, C. (2018). Analyzing the attitudes of translation students towards CAT (computer-aided translation) tools. *Journal of Language and Linguistic Studies*, *14*(1), 153–161.

Çetiner, C., & İşisağ, K. U. (2019). Undergraduate level translation students' attitudes towards machine translation post-editing training. *International Journal of Languages' Education, 7*(1), 110–120. https://doi.org/10.18298/ijlet.3242

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y., Ye, W., Zhang, Y., Chang, Y., Yu, P. S., Yang, Q., & Xie, X. (2024). A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology*, *15*(3), 1–45. https://doi.org/10.1145/3641289

D'Acquisto, G. (2017). *A linguistic analysis of diplomatic discourse*. Cambridge Scholars Publishing.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 4171–4186). Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423

Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., Qin, B., & Liu, T. (2024). A survey on hallucination in large language models:

Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems, 43*(2), Article 1. https://doi.org/10.1145/3703155
Hutchins, W. J., & Somers, H. L. (1992). *An introduction to machine translation*. Academic Press.
Jaber, A. (2001). Language and diplomacy. In J. Kurbalija & H. Slavik (Eds.), *Language and diplomacy* (pp. 49–54). DiploProjects.
Jiang, L., Jiang, Y., & Han, L. (2024). The potential of ChatGPT in translation evaluation: A case study of Chinese-Portuguese machine translation. *Cadernos de Tradução, 44*(1). https://doi.org/10.5007/2175-7968.2024.e98613
Johnson, M., Schuster, M., Le, Q. V., Krikun, M., Wu, Y., Chen, Z., Thorat, N., Viégas, F., Wattenberg, M., Corrado, G., Hughes, M., & Dean, J. (2017). Google's multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics, 5*, 339–351. https://doi.org/10.1162/tacl_a_00065
Kingscott, G. (2002). Technical translation and related disciplines. *Perspectives: Studies in Translatology, 10*(4), 247–255.
Martirosyan, A. (2022). Modality's features of diplomatic discourse. *Bulletin of Eurasia International University*, 129–144. https://doi.org/10.53614/18294952-2022.1-129
Mercan, H., Akgün, Y., & Odacıoğlu, M. C. (2024). The evolution of machine translation: A review study. *Lotus: International Journal of Language and Translation Studies, 3*(1), 53–72. https://doi.org/10.55003/lotus.1453321
Odacıoğlu, M. C. (2021). Çeviri edincinden çevirmen edincine geçişte çevirmen adaylarının sorumlulukları ve çeviri eğitiminin bu sürece katkısı. *RumeliDE Dil ve Edebiyat Araştırmaları Dergisi, (20)*, 897–912. https://doi.org/10.29000/rumelide.897299
OpenAI. (2023). *GPT-4 technical report*. https://arxiv.org/abs/2303.08774
Özcan, O. (2024). Avrupa Komisyonu'nun makine çeviri sistemi eTranslation'da metin çevirisi. *Journal of Academic Studies in World Languages, Literatures and Translation (WOLLT), 5*(2), 191–216.
Pokharel, S. (2020). Diplomatic language: An analysis of salutations from speeches used in international diplomacy. *Diplomatic Language: An Analysis of Salutations From Speeches Used in International Diplomacy, 3*(1), 180–193. https://doi.org/10.3126/joia.v3i1.29094
Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3*(3), 417–424. https://doi.org/10.1017/S0140525X00005756
Song, C., & Kao, Q. (2025). Comparing generative AI and advanced learners in translation tasks using the House model: Similarities and pedagogical implications. *International Journal of Chinese Language Teaching*, 1–28. https://doi.org/10.46451/ijclt.250503
Turing, A. M. (1950). Computing machinery and intelligence. *Mind, 59*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30*, 5998–6008. https://arxiv.org/abs/1706.03762

Vraila, S. (2024). AI in political translation: Revolutionising communication or risking manipulation? In *New Trends in Translation and Technology Conference 2024* (pp. 226–229). https://doi.org/10.26615/issn.2815-4711.2024_019

Yetkin-Karakoç, N. (2022). *Çeviri ve diplomasi* (2nd ed.). Çeviribilim Yayınları.

Zhao, W. X., Zhou, K., Li, J., Tang, T., Dong, Z., Hou, Y., Zhang, B., Min, Y., Zhang, J., Liu, P., et al. (2026). A survey of large language models. *Frontiers of Computer Science, 20*, Article 2012627. https://doi.org/10.1007/s11704-026-60308-3

Zhou, L., Muhammad Alif, M. S. N., & Peng, J. (2025). The impact of ChatGPT on part-time translators working with the English language: A threat or a complementary tool? *World Journal of English Language, 15*(7), 103. https://doi.org/10.5430/wjel.v15n7p103

Appendices

Appendix (1): Table 1

Instructions for Jury Members: For each translation (Participant 1–5 and ChatGPT), please rate from 1 (Very Poor) to 5 (Excellent) for each criterion.

Category	Description	Participant 1	Participant 2	Participant 3	Participant 4	Participant 5	Participant 6
Diplomatic Tone and Politeness Strategies (Weight: High)	Does the translation maintain a neutral, formal, and tactful tone suitable for international diplomatic communication?						
Register and Formality Level	Is the level of formality appropriate for diplomatic correspondence (not too casual, not overly rigid)?						
Cultural and Contextual Sensitivity	Does the translation reflect cross-cultural awareness and avoid language that could cause offense or misunderstanding						

Category	Description	Participant 1	Participant 2	Participant 3	Participant 4	Participant 5	Participant 6
	in diplomatic settings?						
Accuracy and Content Faithfulness	Does the translation correctly convey the factual content, nuance, and implied meanings of the source text?						
Fluency and Readability	Is the translation grammatically correct, coherent, and smooth in the target language?						

Appendix (2): Texts

Text 1

II. Translate the following into Turkish:

……..(put an appropriate title in the address line)

I have very much appreciated Your kind letter of congratulations and good wishes on my assuming the post of Minister of culture of ……..I hasten to say that I fully reciprocate your sentiment as well as the need to promote bilateral relations.

I am moved by Your recollection of our past contacts and cooperation in the interests of our respective countries. Indeed, our nations are liked by indissoluble ties of shared history, cultural affinity and above of love and affection that happily exist between our two peoples. I avail myself of this opportunity to extend to your Excellency an invitation for an official visit to Türkiye at Your earliest convenience.

Please accept, Excellency, my best wishes for Your success in Office as well as Your personal health and happiness.

(5 points each for address line and sentences)

Text 2

IV. Translating a diplomatic Message

Excellency,

It was a great pleasure for me to meet with Your Excellency and to participate in the 5th Peace and Progress Summit which took place in the beautiful city of Georgia.

Please allow me to take this opportunity to thank Your Excellency in person and all the authorities that contributed to the organization of this excellent conference which I believe was a significant milestone in the achievement of enhanced cooperation and solidarity among the member countries.

Text 3

V. Translate the following text into English.

i. Emniyet yetkililerinden alınan bilgiye göre, hava meydanları, deniz limanları ve sınır kapıları gibi stratejik noktalardaki güvenlik önlemleri, uluslararası sözleşmelere uygun olarak tesis edilmektedir. Büyükelçilikler milli gün, sergi, heyet ziyareti gibi vesilelerle rezidansta veya otel gibi başka mekanlarda düzenleyecekleri etkinlikler için koruma talep etmelidir. Bu talep genel olarak Dışişleri Bakanlığı aracılığıyla ilgili emniyet makamlarına iletilebildiği gibi, kabul eden Devlet tarafından uygun görülen herhangi bir makama da doğrudan iletilebilir…Ülkemizde bulunan Büyükelçiler ile ülkemizi ziyaret eden üst düzeyli yabancı Devlet ve Hükümet erkanı, mütekabiliyet ilkesi çerçevesinde, ücretli veya ücretsiz olarak VIP salonlarından yararlanabilmektedirler. Eş ve çocuklar da, hak sahibi kişi ile birlikte seyahat ediyorlarsa, bu ayrıcalığa sahip olur.

Acknowledgements

The authors wish to thank the jury members – two scholars in international relations and one retired ambassador – for their time and expertise in evaluating the translations. The authors also extend their gratitude to the five student participants from the Department of English Translation and Interpreting at İzmir University of Economics for their willing participation in the study.

Authors' Biodata

Yaşar Akgün graduated from the Department of Translation and Interpreting at Kutahya Dumlupınar University as High Honour student. Having finished his master degree at Ege University in the field of Translation Studies with the thesis named *The analysis of note taking language in consecutive interpreting within the scope of English and Turkish language pair: A case study with professionals and semiprofessionals*, he is currently pursuing his PhD at Dokuz Eylül University in the field of Translation Studies. His main interests are diplomatic interpreting, consecutive interpreting and postcolonial translation studies. He has penned several articles published in both national and international journals in specified fiels and translated several literary books published in Türkiye. Affiliated with the Department of Translation and Interpreting, he is currently working as a research assistant at İzmir University of Economics.

Nihal Yetkin Karakoç graduated from the Department of English Translation and Interpreting at Hacettepe University. Between 1996 and 1999, she performed consecutive interpreting at the Ministry of National Defense for high-level officials, including the minister and the undersecretary, during bilateral meetings. From 1999 to 2002, she worked as a translator and interpreter at TAI, an international aviation company with Turkish-American partnership. In 2005, she completed her master's thesis in English Linguistics at Hacettepe University, and in 2009, she completed her doctoral dissertation. She received the title of associate professor in 2017 and was promoted to professor in 2023.

Since 2007, she has been working at İzmir University of Economics. Her main fields of interest include translation of diplomatic texts, interpreting, linguistics, subtitling and women's studies. In these fields, she has published numerous articles and book chapters in both national and international journals and books.

Mirza İşi is currently pursuing a double major in English Translation and Interpreting and Computer Engineering at İzmir University of Economics, with a primary focus on artificial intelligence. He also works at Outlier, a subsidiary of Scale AI, as a Quality Control Auditor, where he contributes to the training of various large language models (LLMs) for major clients, including through Reinforcement Learning from Human Feedback (RLHF), one of the most widely used methods in the field. Through this work, he has developed a solid understanding of how LLMs function and are trained. His academic background in translation studies, combined with his technical expertise in computer engineering, enables him to approach AI and language-related research from an interdisciplinary perspective, with particular interest in the intersection of AI and translation.

Authors' Contributions

Nihal Yetkin Karakoç prepared the texts, organized jury design, performed all the experiments and analyzed and discussed the findings in collaboration with her colleagues. Yaşar Akgün executed the process of statistically analyzing the translations done by students and their comparisons with ChatGPT. Mirza İşi predominantly covered the sections on large language models and technical principles underlying such systems. All authors contributed equally in the process of writing the article and reviewed and approved the final version of the manuscript.

Disclaimer status on the use of Generative AI

ChatGPT (GPT-4o, OpenAI) was used in this study strictly as the subject of empirical investigation, serving as one of the six translation participants whose outputs were evaluated by a jury alongside those of human translators. Generative AI tools were not used in the drafting, editing, or preparation of the manuscript itself. All textual content, analysis, and conclusions presented in this article are the sole intellectual product of the authors.

Declaration of conflicting interest

The authors declared no conflicts of interest to the article's research, authorship, and/or publication.

Ethical Approval

Ethical Approval was ensured and informed consent was obtained from all human participants prior to their involvement in the study. Participants were informed of the purpose of the research, the voluntary nature of their participation, and their right to withdraw at any time without consequence. No personally identifying information was collected, and all data were handled confidentially and used solely for academic research purposes.

Available online at https://www.asjp.cerist.dz/en/Articles/155