

Translating legal texts: A four-criteria evaluation of ChatGPT, DeepL, and Google Translate

Merabet Mohamed Hamza, PhD student

University of Malaga (Spain)

merabet.mhamza@uma.es

ORCID: <https://orcid.org/0009-0007-7433-319X>

Roser Nebot Nicolas, Full Professor

University of Malaga (Spain)

nroser@uma.es

ORCID: <https://orcid.org/0000-0002-4732-4366>

Abstract: Translating legal texts – especially between linguistically and culturally distinct languages such as Arabic and English – poses significant challenges within Translation Studies. This study evaluates the performance of three machine translation tools: ChatGPT (OpenAI GPT-3.5, March 2024), Google Translate (web version, April 2024), and DeepL (free web version, April 2024), focusing on a corpus of Algerian Supreme Court judgments. Drawing on key theoretical frameworks including functionalism, equivalence and legal semiotics, we assess translation quality across four criteria: accuracy, clarity, fluency, and cultural appropriateness. Detailed documentation of tool versions, prompting strategies, and input constraints

ensures transparency and supports reproducibility. Our findings indicate that while all tools exhibit limitations inherent to the complexity of legal language, ChatGPT shows particular strength in handling cultural nuances and producing more contextually appropriate translations. Nonetheless, human expertise remains vital for ensuring reliability in professional legal translation. Though the small dataset limits the generalizability of results, this study offers important insights for Translation Studies scholarship and lays groundwork for further research into the evolving role of AI in legal translation.

Keywords: legal translation, translation evaluation, ChatGPT, Neural Machine Translation, evaluation criteria

1. Introduction

For machine translation systems, the translation of specialised content is a major challenge, as it must accurately convey the meaning of the text and at the same time require a thorough understanding of the vocabulary typical of the subject area. For this reason, there is an increasing need for machine translation systems that can reliably process specialist texts. Artificial intelligence (AI) has therefore influenced the lives, ideas and experiences of our time. It is of crucial importance for solving current problems, including the translation industry. Artificial intelligence (AI) has significantly influenced the translation industry, giving rise to a wide range of cutting-edge technologies that have proven indispensable in both general translation and other specialised fields where translation is needed. In the literature, these tools have attracted a great deal of scholarly interest (Wu et al. 2016), (Dong et al., 2022), (Khoshafah, 2023), (Moneus & Sahari, 2024).

The main difficulty with machine translation is not just translating words and phrases correctly, but also successfully communicating legal and cultural nuances and concepts in a natural and concise way. This clarity has the potential to greatly affect how the reader understands the material in general. As such, a number of academics (Šarčević, 1997; Cao, 2007; Biel, 2014; Garzone, 2000)

stress the significance of accuracy in conveying the text's intended meaning. They contend that, particularly when cultural factors are present in the legal context, fluency and clarity are crucial to ensuring that readers can easily understand the content. As a result, machine translation ought to be able to live up to the reader's expectations by offering a precise and understandable translation while maintaining the cultural components included in the source text. Therefore, if machine translation systems, such as the newly introduced ChatGPT and other neural systems can go beyond literal translations to include the accuracy of the meaning, clarity of the message, fluency of the language as well as cultural appropriateness, they could greatly enhance communication in different fields of specialization.

This article assesses the extent any of ChatGPT, Google Translate and DeepL provides an accurate and fluent translation of legal texts. Such a qualitative analysis evaluates which of the tools is suitable for in-depth legal translations according to the parameters of accuracy, clarity, fluency, and culture. Nevertheless, the technology evolution widely known as AI within translation is most likely to contribute to the revolution of the translation industry. This paper explains what machine translation can provide and where its limitations are in order to evaluate whether it is, or can be made, as accurate and subtle as human translation. It also discusses the potential use of ChatGPT, an LLM, within a professional context relative to more sophisticated NMT tools. As AI-driven translation becomes more integrated into professional workflows, evaluating tools not only through empirical comparison but also through theoretical engagement becomes essential for the advancement of Translation Studies as a field. Our research focuses on a careful assessment of each individual AI tool towards productivity enhancement and the adverse impact of each of them as they become part of routine human processes. As far as the researchers know, there has not been any other trial of these types of machine translation systems performed on Algerian Arabic-English legal documents using these four evaluation criteria before. Thus, this research work is unique within such describing literature.

2. Research questions

1. What distinguishes ChatGPT as a LLM from Google Translate and DeepL?
2. In what ways do these technologies meet or fall short of expectations for professional practice concerning accuracy, clarity, fluency, and cultural appropriateness?
3. In contexts involving professional translation, are these translation tools trustworthy?

3. Literature review

Previous research on the translation of legal texts has emphasised the importance of striking a balance between accuracy, clarity, fluency and cultural appropriateness (Šarčević, 1997; Cao, 2007; Biel, 2014; Garzone, 2000). In addition, numerous evaluations of machine translation (MT) systems over the years have shown significant advances in their performance, and our research extends the theoretical framework of machine translation evaluation (Wu et al. 2016, Cotelli Kureth et al., 2023; Hidalgo-Ternero, 2020) and large language model systems such as ChatGPT (Dong et al., 2022, Gao et al., 2023). These frameworks form the theoretical basis for the evaluation of translations using NMT and AI. The review focuses on the evaluation of machine translation and artificial intelligence, how they are used, what disadvantages they entail and which of them is more reliable in the field of work.

3.1 Evaluating Legal Translation

When evaluating the quality of translations, four crucial elements are usually taken into account: Accuracy, clarity, fluency and cultural appropriateness. A translation is considered accurate if it authentically conveys the entire original meaning, so that no details are omitted or

altered. Indeed, Šarčević (1997) emphasises the importance of accuracy in the translation of legal documents. Her work serves as a fundamental example of how even small errors in translation can have a major impact on the law, and emphasises the need for accuracy in order to preserve the legal meaning of the original text. Because translation work is so complex, the translator's task is not an easy one. Concepts and practises specific to each legal language and culture are closely interwoven with the language of law. It is almost impossible to translate them between languages and systems without making mistakes, misinterpreting them or even conveying meaningless information (Sacco, 1999, p. 169). Colson therefore emphasises that a translator working with a technical document must not embellish it at the cost of content accuracy (Colson, 1994, p. 69).

Secondly, in order to make the language easy to understand for the target audience, clarity is assessed by whether the translation is comprehensible and free of ambiguity. In her book "Translating Law", Cao discusses the complexity of legal language and the importance of clarity in legal translations. Her work focuses on the importance of making legal texts comprehensible to both lawyers and laypeople, which is essential for effective legal communication (Cao, 2007). A group of scholars have pointed out that word polysemy can lead to stylistic errors. This can be seen in the disruption of sentence structure when the wrong word is chosen in translation. Instead of a smooth, coherent reading experience, the reader encounters something that does not match their expectations, resulting in poor style and a feeling of detachment from the text (Cornu, 2005, pp.102-103). However, several other authors confirm that the enormous amount of material translated daily in this field – which will only increase with globalisation – refutes the above view. In reality, it is perhaps more appropriate to speak of the success of translation rather than its impossibility. This is particularly important as translation scholars are working hard to develop methods that make translation feasible. As Christine Durieux has noted, translation is above all understanding with the aim of conveying understanding. It is therefore ultimately about making concepts understandable and obvious to the recipient (Durieux, 1992, p. 102).

Thirdly, to ensure that the translation reads naturally and adheres to the grammatical norms of the target language, fluency focuses on how naturally the language is used. This was discussed in an article by (Biel, 2014) which examined how fluency and readability affect legal translations, particularly in multilingual legal systems such as the European Union. Cao picked up on this point, stating that the translator's task when working on a legal document is to identify the legal meaning and distinguish it from the general meaning before translating it into the target language. This means that the translator must rely on the writing style and fluency of the legal text in the target language to effectively convey the legal information and its implications from the source text (Cao, 2007, p. 67).

Finally, to ensure that culturally conditioned references are appropriately rendered or modified to make sense to the audience, cultural appropriateness evaluates how effectively the translation reflects and adapts to the cultural context of the target language. In this respect, translation theorists almost unanimously agree that no text can be free of flaws when translated between cultures. It is even claimed that while linguistic transfer between languages is possible, direct transfer of cultures is almost impossible without significant distortions occurring (Mohanty, 2005, p.9). In this context, Snell-Hornby pointed out that the translatability of a text depends on how deeply it is rooted in its original culture and how great the distance in time and space is between the cultural background of the source text and the target audience. The more the source text embodies its culture, the more difficult it will be for the translator to convey this through the translation. And the greater the temporal and spatial gap between the culture of the text and the culture of the target audience, the more difficult it is for the translator to bridge this gap (Snell-Hornby, 1988, p.41). In the legal field, Garzone (2000) argues that the translation of legal texts requires bridging cultural and legal system differences, which can prove particularly difficult when concepts have no direct equivalents in the target language, which is why Roser Nebot (2019) emphasizes that one of the most important aspects of intellectual production and translation is the cultural knowledge that must be consistently acquired throughout one's life (pp. 239-240). For this reason, the translation of specialized texts such as legal documents

and cultural content requires a deep empathy and understanding of the legal customs, traditions, symbols and social ideas in each language.

3.2 Neural Machine Translation as the Foremost Tool

3.2.1 Google Translate

Google Translate is an autonomous machine translation service provided by Google Inc. Boitet et al. (2009) state that initially it functioned as a Statistical Machine Translation (SMT) system that directly translated written source languages into other languages. On the other hand, Google used Neural Machine Translation (NMT) to introduce the Google Neural Machine Translation (GNMT) system in November 2016 (Turovsky, 2016). A recurrent neural network that applies the Encode-Decode paradigm is the basis of NMT. After encoding the source phrase into context vectors, it selects terms from the target vocabulary to construct the translation token-by-token (Liu et al., 2016).

Google Translate is one of the most popular websites globally, with about 500,000,000 users every day (Almahasees et al., 2021). This well-known online translator translates text between languages using machine learning techniques. It can translate full web pages, documents, and even voice in real time and supports more than 100 languages (James, 2020). As a result, Google Translate is now a necessary tool in our daily lives and is known to those who are computer knowledgeable. This tool is extensively used and acknowledged in many different languages.

Cheng (2019) points out that while NMT has advanced significantly recently, NMT systems still solely use parallel corpora for parameter estimation. Utilising monolingual corpora to improve NMT is attractive because parallel corpora are usually constrained in terms of number, quality, and coverage, especially for lower source languages. Wu et al. (2016) point out that the majority of NMT systems have trouble with uncommon terms, which hinders their use

in real-world applications and services where precision and speed are essential. Wu et al. (2016), however, show that Google's Neural Machine Translation, or GNMT, system decreases translation errors by about 60% on a number of popular language pairings when compared to the prior phrase-based output method. Extending this empirical investigation into the Arabic-English context, a study by Stiti et al. (2023) explores the challenges of translating specialized legal texts, particularly Algerian laws, which feature complex linguistic traits like culture-specific terms, honorifics, nominalizations, modals, doublets, lexical repetition, and syntactic intricacies. Their analysis of Google NMT outputs for selected Algerian law articles identifies persistent issues across lexical, semantic, syntactic, and pragmatic dimensions, concluding that for high-quality, specialized translation, post-editing remains an indispensable step. This finding confirms the earlier conclusion of Benounane et al. (2020) regarding the necessity of post-editing to refine the superior yet imperfect output of NMT systems for high-quality English-Arabic translation.

3.2.2 DeepL

DeepL made a big splash when it was introduced in 2017, quickly earning praise from all kinds of users—experts, researchers, and even casual users. It is a neural machine translation tool that really sets itself apart from older systems because of its impressive improvements in fluency and accuracy. A big reason for this is DeepL's massive and diverse training dataset, which helps it understand context better, and its use of advanced Transformer architecture. It is particularly good at handling tricky things like specialized jargon, complex sentences, and even everyday expressions (Kamaluddin et al., 2024).

In fact, some researchers argue that DeepL is more accurate than Microsoft Interpreter and Google Translate, making it one of the top options available today (Cotelli-Kureth et al., 2023). According to Kamaluddin et al. (2024), DeepL has found lots of practical uses in

business, academics, and teaching. People appreciate its accuracy, ease of use, and helpful features like the glossary function. It even helped improve international research collaborations, made communication smoother in global companies, and become a useful tool for language learners.

Research by Hidalgo-Ternero (2020) points out that DeepL's advanced neural networks and well-prepped data might explain why it is reliable at understanding context and delivering more natural translations than its competitors. In professional and academic settings, DeepL shines by offering translations that are specifically tailored to the user's needs, especially when it comes to technical terms and particular contexts (Burchardt et al., 2021). That said, it can still face challenges when it comes to translating texts that are heavily context-driven or tied to cultural nuances (Kamaluddin et al., 2024).

3.3 LLMs as a Translation Tool

3.3.1 ChatGPT

ChatGPT, developed by OpenAI, is an intelligent chatbot that is starting to change the landscape of translation, showing promising results (Zhao et al., 2023). It can help users with many tasks, including translation, thanks to its advanced Natural Language Processing (NLP) technology. This allows it to understand questions quickly and accurately, interpret them, and deliver context-aware translations (Haleem et al., 2022).

Large language models (LLMs) like ChatGPT, with hundreds of millions to billions of parameters, are trained on massive amounts of text data. They're self-supervised and nonparametric, which makes them highly flexible. OpenAI's GPT models are a perfect example of this (Xie et al., 2023). According to Xie et al. (2023), even though translation is a challenging task, ChatGPT-3.5 performs impressively well. They note that while there are still some hurdles to overcome, LLMs like ChatGPT have shown real promise as effective translators,

though more research is needed before they are ready for widespread, general-purpose translation use.

Dong (2022) points out that ChatGPT has an edge over traditional machine translation systems because it can incorporate additional information, like user interaction, via prompts (Dong et al., 2022). Thanks to the advancements in NLP, ChatGPT is now seen as a powerful translation tool, often more versatile and accurate than older models. As a result, it is becoming a valuable tool for tasks that require precise language translation and real-time communication.

Gao et al. (2023) conducted several tests to explore ChatGPT's potential in machine translation, discovering that it often matched or even outperformed established professional systems, especially for high-resource languages. By using a range of translation prompts, the study found that ChatGPT consistently delivered accurate translations, sometimes better than commercial systems. One key takeaway was ChatGPT's ability to adapt its translations to specific contexts by understanding domain-specific terms, significantly improving accuracy. This highlights ChatGPT's potential as a tool for specialized and professional translators.

On the other hand, Khoshafah (2023) looked at ChatGPT-3.5's performance and found that while it handles simple tasks well, it struggles with more complex content that requires deeper human understanding. ChatGPT's limitations are particularly noticeable in translating legal, medical, scientific, and literary texts, where human intervention is crucial to accurately capture nuances like idioms and metaphors. Additionally, Bang et al. (2023) observed that ChatGPT can occasionally produce incorrect translations, and Moneus & Sahari (2023) compared AI systems with human translators in Arabic-English legal translation. They concluded that AI systems were less effective at preserving the legal intent and context, although they noted that AI could improve as the technology advances.

In conclusion, the quality of neural machine translation (NMT) tools, including ChatGPT, should be assessed based on four main criteria: accuracy, clarity, fluency, and cultural appropriateness. These factors have been discussed in the studies mentioned earlier. While ChatGPT, DeepL, and Google Translate have come a long way, more research is necessary to fully evaluate their performance across

different languages and types of texts, especially as the technology continues to evolve.

4. Methodology

4.1 Tool Selection and Settings

To ensure a fair and consistent comparison among the selected translation systems, three widely used machine translation tools were employed in this study: ChatGPT, Google Translate, and DeepL. Each tool was accessed through its respective web interface under comparable conditions, with default settings maintained to preserve authenticity and replicability. The following section outlines the specific configurations and prompts used for each system during the translation tasks:

- i. **ChatGPT:** OpenAI, GPT-3.5, web interface, March 2025.
Prompt:
“Translate the following legal text from Arabic to English with attention to accuracy, clarity, fluency, and cultural appropriateness. Do not paraphrase or summarize”.
- ii. **Google Translate:** Web interface, accessed April 2025; default settings.
- iii. **DeepL:** Free web version, accessed April 2025; default settings.

4.2 Corpus

Seven legal passages from Algerian Supreme Court rulings (1984–2006), rich in technical terminology and embedded cultural meaning, were selected. While this sample limits quantitative generalization, it

allows for qualitative depth and detailed error typologies, in keeping with Translation Studies' emphasis on interpretive nuance.

4.3 Evaluation Criteria

Scoring System:

Each translation for each passage was scored (scale: 1=Poor, 5=Excellent) across:

- i. **Accuracy,**
- ii. **Clarity,**
- iii. **Fluency,**
- iv. **Cultural Appropriateness.**

The outputs of the three tools will be evaluated based on the aforementioned criteria as follows:

1. Accuracy: This criterion assesses how successfully the translations capture the intended meaning of the source material.
2. Clarity: Measures how well the translation offers the original meaning so that the reader can understand it with ease.
3. Fluency: This evaluates the translations' grammatical and stylistic coherence in the target tongue.
4. Cultural Appropriateness: This criterion examines how well translations respect and convey cultural peculiarities that are significant in the context, particularly with regard to legal terminology.

Each criterion was scored on a scale from 1 to 5, where:

- 1:** Poor quality (the translation fails to convey the intended meaning or is very unclear).
- 2:** Below average (the translation contains significant errors or lacks clarity and fluency).
- 3:** Average (the translation conveys the general meaning but may have some inaccuracies, lack of clarity, or fluency issues).

4: Good (the translation is mostly accurate, clear, and fluent, with minor errors).

5: Excellent (the translation is accurate, clear, fluent, and culturally appropriate).

This method of scoring is based on a combination of previous studies that focus on key aspects of translation quality evaluation, including accuracy (House, 2015; SAE J2450 Standard, 2001), fluency (Papineni et al., 2002), clarity (Williams & Chesterman, 2009; Canavese, 2023), and cultural appropriateness (Baker, 1992; Venuti, 1995).

To guarantee appropriate interpretation, we will begin by ensuring that the reader understands the meaning of important legal terms and phrases by consulting a variety of trustworthy sources. After then, the translations produced by DeepL, Google Translate, and ChatGPT will be evaluated according to the four aforementioned criteria, so that every tool's output will be graded for each criterion on a scale from 1 to 5 as mentioned earlier.

Each passage will receive a total score out of 20, based on the combination of the scores, which will represent the translation's overall quality. This grading scheme will draw attention to the distinct advantages and disadvantages of every instrument. After that, a thorough examination will be carried out to determine whether these translation technologies are fit for usage in such specialised situations and how reliable they are when used with technical writings, such legal papers.

5. Results

Original Passage 1

من المتفق عليه فقها أن الخلوة الصحيحة توجب جميع الصداق المسمى

Key term definitions

The term ***khalwah*** (خلوة) is defined in legal terminology as the husband being alone with his wife after a valid marriage contract has been concluded (Ibn Qudāmah, 1984, p. 191). Thus, it can be said that ***khalwah*** (خلوة) is the husband's seclusion with his wife in a place where no other person is present, with the possibility of intimate relations occurring between the spouses, which is permitted in Islamic law following the valid contract and the absence of any legal, natural, or physical impediments, as previously mentioned. Therefore, the translation must address these two aspects (seclusion and the potential for intimate relations).

Translations

1. Google Translate:

It is agreed upon in Islamic jurisprudence that a valid seclusion requires the entire specified dowry.

2. DeepL:

It is agreed upon in jurisprudence that a valid khuluwwah is entitled to all the named dowry.

3. ChatGPT:

It is unanimously agreed upon in jurisprudence that a valid seclusion necessitates the payment of the entire specified dowry.

Evaluation

Criterion	Google Translate	DeepL	ChatGPT
Accuracy	4: Correctly translates the main idea but fails to convey the concept of <i>ijmā'</i> (إجماع) / <i>muttafaq'alay-hi</i> (متفق عليه) representing consensus; phrase omitted	3: Misses the emphasis on consensus terms and leaves Arabic word <i>khuluwwah</i> untranslated – which is misleading	5: Accurately conveys both legal meaning and cultural context; appropriately uses “unanimously”
Clarity	4: Generally clear but lacks the nuance	3: Use of <i>khuluwwah</i> –	5: Clear and straightforward;

Criterion	Google Translate	DeepL	ChatGPT
	of consensus, which may confuse readers	incorrect transcription of <i>khalwah</i> – is not familiar to English readers; leads to confusion	clarifies consensus and concept
Fluency	5: Grammatically smooth and fluent	4: Mostly fluent; awkwardness due to unfamiliar term <i>khuluwwah</i>	5: Natural phrasing with no awkwardness
Cultural Appropriateness	4: Understandable but lacks full depth of Islamic legal terminology	2: Use of untranslated Arabic term without explanation hurts accessibility	5: Successfully captures cultural and Islamic legal context
Total scores	17	12	20

Table 1. Passage 1 Evaluation

Evaluation Discussion

- i. Google Translate effectively captures general meaning but omits the critical notion of **consensus** (*ijmāʿ* / *muttafaqʿalay-hi*), reducing cultural and theological richness.
- ii. DeepL’s use of *khuluwwah* is both a misspelling and an inappropriate transliteration, confusing readers unfamiliar with Arabic legal terms, which reflects poor cultural adaptation.
- iii. ChatGPT accurately translates both meaning and terms, with “unanimously” reflecting legal consensus, fulfilling expectations from Translation Studies on functional equivalence and cultural transposition.

Original Passage 2

إن الحبس يعتبر من أعمال التبرع

Key termdefinitions

Ḥabs (حبس) or **waqf** (وقف) denotes “restraining the property and dedicating its benefit” (Al-Zarkashi, 1993, p. 268). **Ḥabs** derives from “to restrain” (*taḥbīs*), referring to preventing transfer of ownership while dedicating profits to specified goals (Al-Bahūtī, 1983). The “property” (*al-asl*) is the endowed asset, while “dedicating its benefit” (*tasbīl al-manfaʿa*) refers to directing returns to beneficiaries or causes (Al-Bahūtī, 1983).

Translations

1. **Google Translate:**
Imprisonment is considered an act of donation.
2. **DeepL:**
Confinement is an act of kindness.
3. **ChatGPT:**
Endowment (or waqf) is considered an act of donation.

Evaluation

Criterion	Google Translate	DeepL	ChatGPT
Accuracy	3: Captures donation concept but translates ḥabs as “imprisonment” inaccurately	2: Translates ḥabs as “confinement” and inaccurately equates donation with “kindness”	5: Correctly translates ḥabs as “endowment” and includes synonym <i>waqf</i>
Clarity	3: Somewhat clear but misleading due to “imprisonment”	2: Confusing and misleading due to inaccurate terms	5: Clear, logical, accurate
Fluency	5: Fluent and grammatically sound	4: Mostly fluent but confuses meaning	5: Smooth, natural
Cultural Appropriateness	3: Misleading use of “imprisonment” affects cultural understanding	2: Fails to convey legal and cultural meaning	5: Effectively captures cultural context
Total scores	14	10	20

Table 2. Passage 2 Evaluation

Evaluation Discussion

- i. Google Translate partially captures “donation” but the key term **ḥabs** mistranslated as “imprisonment” causes confusion.
- ii. DeepL’s rendering of **ḥabs** as “confinement” and donation as “kindness” misrepresents key Islamic legal concepts.
- iii. ChatGPT excels with culturally informed accurate terms and synonyms, improving understanding and mirroring **Nikulin’s** call for linguistic-cultural transparency in translation of legal texts.

Original Passage 3

أن الطاعة كانت قدمت شهوداً أثبتوا انعقاد الزواج واعتماد قضاة الموضوع في رفض الدعوى على عدم توفر ولي الزوجة يعد خطأ قانونياً وشرعياً إذ أنه يحق للمرأة زواج نفسها ما دامت ثيباً.

Key term definitions

ثيب (Thayyib) refers to a woman previously married or no longer a virgin (Ibn Manẓūr, 1999). This includes women divorced, widowed, or otherwise separated after consummation.

Translations

1. Google Translate:

The appellant had presented witnesses who proved the conclusion of the marriage, and the judges’ reliance in rejecting the case on the lack of the wife’s guardian is a legal and religious error, as a woman has the right to marry herself as long as she is not married.

2. DeepL:

The appellant had presented witnesses who proved that the marriage had taken place, and the trial judges’ reliance on the absence of the wife’s guardian is a legal and legal error, as a woman has the right to marry herself as long as she is a virgin.

3. ChatGPT:

The appellant had presented witnesses who confirmed that the marriage had taken place. The reliance of the trial judges on the absence of a guardian for the wife to reject the claim is

considered a legal and religious error, as a woman has the right to marry herself if she is previously married (a widow or divorcee).

Evaluation

Criterion	Google Translate	DeepL	ChatGPT
Accuracy	2: Incorrectly renders ثيب as “not married” and translates <i>shar‘iyan</i> as “religious” instead of “Islamic law”	2: Misinterprets ثيب as “virgin”, redundantly uses “legal and legal error”, mis-translates <i>shar‘iyan</i>	4: Correctly translates ثيب with cultural nuance; incorrectly uses “religious” instead of “According to Islamic law”
Clarity	3: Generally clear but misleading about marital status	2: Confusing due to redundancy and incorrect term	5: Clear and precise
Fluency	5: Fluent and well-structured	3: Awkward redundancy affects flow	5: Fluent and coherent
Cultural Appropriateness	3: Misses vital cultural distinction	2: Omits cultural significance, confusing marital context	4: Effectively conveys cultural nuance
Total scores	13	9	18

Table 3. Passage 3 Evaluation

Evaluation Discussion

- i. Google and DeepL make serious errors with **ثيب**, thus misrepresenting women’s status in Islamic legal terms.
- ii. ChatGPT’s translation is most accurate and culturally sensitive, enhancing legal and cultural comprehension despite a minor error translating *shar‘iyan*.

Original Passage 4

يثبت الزواج بعد الدخول بصدّاق المثل

Key termdefinitions

ṣadāq (صدّاق) is an Islamic dowry paid by the husband to the wife to consummate marriage rights, differing from Western dowry (Al-Faḍīlāt, n.d, p.148). From that definition, it can be said that the Islamic dowry (*ṣadāq*/صدّاق) is what is paid to a woman in money or other forms to express a man's intention to take her as his wife.

ṣadāq bi-l-mithl (صدّاق بالمثل), or the dowry of parity, is Algerian law's provision comparing bride to similar women for dowry equivalency.

Translations

1. Google Translate:

Marriage is proven after consummation with a dowry of equal value.

2. DeepL:

Marriage is fixed after entry with the same dowry.

3. ChatGPT:

The marriage is confirmed after consummation with a dowry equivalent to that of peers.

Evaluation

Criterion	Google Translate	DeepL	ChatGPT
Accuracy	4: Generally accurate but “proven” and “equal value” not fully contextually precise	2: Misleading phrasing; “fixed” & “entry” unclear	5: Clear and accurately reflects cultural and legal concepts
Clarity	3: Ambiguous wording	2: Lacks clarity	5: Clear, informative
Fluency	4: Slightly awkward	3: Awkward, non-idiomatic	5: Smooth and fluent
Cultural Appropriateness	3: Partially captures dowry significance	2: Fails to communicate cultural nuance	5: Successfully explains cultural context

Criterion	Google Translate	DeepL	ChatGPT
Total scores	14	9	20

Table 4. Passage 4 Evaluation

Evaluation Discussion

- Google Translate’s “proven” may mislead readers unfamiliar with confirmation through consummation concept.
- DeepL’s wording obscures meaning, lacking necessary cultural insight.
- ChatGPT provides the clearest explanation suited to readers with no prior knowledge.

Original Passage 5

حيث أنه بالفعل فقد جاء في القرار المنتقد بأن الطاعن لم يببر تأخره في الدخول بالمطعون ضدها وأن العدول عن الخطبة لم يكن من جانبها حتى ترد الهدايا والجهاز مع أن الطاعن قد دفع باتفاق العائلتين على أن الدخول يكون بعد تأديته للخدمة الوطنية.

Key termdefinitions

Al-dukhūl (الدخول) refers to consummation of marriage – the lawful intimate relations after marriage, including periods without physical obstacles such as illness or menstruation (Al-Muṭarrizī, n.d., p.161).

Al-jihāz (الجهاز) means the bride’s trousseau or furnishings needed for her new household (Al-Bustānī, 1956; Ibn Fāris, 1985).

Translations

1. Google Translate:

The criticized decision states that the appellant did not justify his delay in consummating the marriage and that the withdrawal from the engagement was not on her part, so she should return the gifts and the bridal trousseau, even though the appellant claimed that both families had agreed that the consummation would occur after he completed his national service.

2. DeepL:

Indeed, the criticised decision stated that the appellant did not justify his delay in entering the respondent and that the engagement was not cancelled by her until the gifts and the device were returned, although the appellant argued that the two families had agreed that the entry would take place after he performed his national service.

3. *ChatGPT*:

Indeed, the contested decision stated that the appellant did not justify his delay in consummating the marriage with the respondent and that the engagement was not broken off by her, which would warrant the return of gifts and trousseau. However, the appellant argued that both families had agreed that the marriage consummation would take place after he completed his national service.

Evaluation

Criterion	Google Translate	DeepL	ChatGPT
Accuracy	4: Mostly accurate but incorrectly translates “criticized” instead of “contested”; does not specify “respondent”	2: Uses “entering” (literal), “device” instead of “trousseau”; awkward	5: Correct terminology and legal clarity
Clarity	4: Generally clear but some awkwardness in phrasing	2: Phrasing confusing	5: Clear and precise
Fluency	4: Fluent with minor awkwardness	3: Awkward phrases disrupt flow	5: Smooth and natural
Cultural Appropriateness	4: Captures culture, lacks deep contextualization	2: Fails to convey cultural/legal importance	5: Fully culturally sensitive
Total scores	16	9	20

Table 5. Passage 5 Evaluation

Evaluation Discussion

- i. **Google Translate:**
Conveys the main ideas clearly, especially the dispute over gifts and the trousseau. However, it uses less precise terms, such as “criticized decision” instead of “contested decision”, and some expressions lack clarity.
- ii. **DeepL:**
Struggles with legal and cultural terms, translating *الدخول* as “entering” and *الجهاز* as “device”. It also misinterprets *حتى*, which distorts the meaning and creates confusion.
- iii. **ChatGPT:**
Offers the most accurate and culturally sensitive translation. It correctly renders *الدخول* as « consummating » and *الجهاز* as « trousseau », ensuring both clarity and legal-cultural relevance.

Original Passage 6

أن قاضي الموضوع لم يعلل حكمه محل الطعن فيما يتعلق بعرض المبلغ المالي مقابل الخلع.

Key term definitions

Khul' (خلع) is a form of divorce initiated by the wife, involving compensation to the husband to effect divorce, distinct from Talaq (divorce initiated by husband) (Belhadj, 2005).

Translations

1. **Google Translate:**
The judge did not justify his ruling under appeal with regard to the offer of the financial amount in exchange for divorce.
2. **DeepL:**
The trial judge did not justify his judgement regarding the offer of money in exchange for khula.
3. **ChatGPT:**
The trial judge did not provide reasoning for the contested ruling regarding the offer of the financial sum in exchange for khul' (divorce initiated by the wife).

Evaluation

Criterion	Google Translate	DeepL	ChatGPT
Accuracy	4: Omits the term khul' , less precise	3: Retains “khula”, but lacks ruling context clarity	5: Precise, includes explanation
Clarity	4: Clear but vague	3: Possibly confusing for unfamiliar readers	5: Clear and informative
Fluency	4: Fluent but awkward phrasing	3: Awkward phrasing	5: Smooth and well-structured
Cultural Appropriateness	3: Lacks adequate cultural framing for khul'	3: Uses term bare	5: Explains cultural/legal significance
Total scores	15	12	20

Table 6. Passage 6 Evaluation**Evaluation Discussion**i. **Google Translate:**

Conveys the main idea about the judge's lack of justification but uses less precise language like “financial amount” instead of “financial sum”. It also omits the key term *khul'* (خلع), reducing cultural accuracy.

ii. **DeepL:**

Keeps the term *khula* but fails to explain it, which may confuse readers unfamiliar with Islamic law. The phrasing is awkward and lacks clarity, especially regarding the legal appeal context.

iii. **ChatGPT:**

Delivers an accurate and fluent translation, clearly explaining *khul'* (خلع) to maintain both cultural relevance and reader understanding. It strikes a strong balance between accuracy and accessibility.

Original Passage 7

أن قضاة الإستئناف ارتكزوا بإلحاق نسب الطفل (م-هـ) للطاعن على كون هذا الأخير لم يأت بالدليل على نكرانه نسب الطفل وراحوا يصرحون بأن المطعون ضدها تكون قد حملت بالطفل (هـ) خلال العلاقة الزوجية بالتالي فهو للفراش.

Key termdefinitions

Firāsh (فراش) indicates a valid marital relationship in Islamic law, conferring legitimacy to children born within it (Al-Jurjānī, 1986; Al-Kāsānī, 1986).

Translations

1. Google Translate:

The appeal judges based their attribution of the child's (M-H) lineage to the appellant on the fact that the latter did not provide evidence of his denial of the child's lineage, and they went on to state that the respondent had conceived the child (H) during the marital relationship, and therefore he was the child of the bed.

2. DeepL:

The appellate judges relied on the fact that the appellant did not provide evidence of his denial of the child's paternity and went on to state that the respondent had conceived the child (M-H) during the marital relationship and therefore the child belongs to the bed.

3. ChatGPT:

The appellate judges based the child's (M-H) paternity on the fact that the appellant did not provide evidence to refute the child's lineage. They stated that the respondent must have conceived the child (H) during the marriage, thus establishing that the child belongs to the marital union.

Evaluation

Criterion	Google Translate	DeepL	ChatGPT
Accuracy	4: Generally accurate but unclear on cultural term	3: Clear legally but vague culturally	5: Accurate and culturally clear
Clarity	3: Phrase "child of the bed" confusing	3: Similar vagueness	5: Clear, culturally

Criterion	Google Translate	DeepL	ChatGPT
			accessible
Fluency	4: Smooth but some awkwardness	4: Fluent but awkward phrase	5: Smooth and professional
Cultural Appropriateness	3: Fails to explain cultural/legal meaning well	3: Lacks explanation	5: Fully contextualized culturally
Total scores	14	13	20

Table 7. Passage 7 Evaluation

Evaluation Discussion

- i. **Google Translate:**
Translates the text fairly accurately but renders *الفراش* (*firāsh*) as “child of the bed”, which is unclear to English readers and weakens the legal-cultural meaning tied to marital paternity.
- ii. **DeepL:**
Conveys the idea of paternity but fails to explain the judges’ reasoning. Phrasing like “child belongs to the bed” lacks clarity and does not reflect the cultural or legal depth of the term *firāsh*.
- iii. **ChatGPT:**
Provides a clear and accurate translation, explaining *الفراش* (*firāsh*) as “the marital union”, which effectively conveys the child’s legal legitimacy. It captures both the legal and cultural significance for readers unfamiliar with Islamic law.

6. General Discussion

The evaluation of ChatGPT, Google Translate, and DeepL in translating Arabic legal texts highlights several patterns that corroborate the complexities outlined by established legal translation scholarship. As underscored in our analysis, legal translation requires not only linguistic accuracy but also a careful negotiation of cultural and institutional particularities. The challenge of translating

specialized legal terminology – such as *khalwah*, *ħabs*, and *thayyib* – demonstrates how these tools engage variably with cultural nuance, an issue repeatedly emphasized in works that examine the cross-cultural transfer of legal concepts.

Consistent with Šarcevic (1997), the results suggest that legal translation is inherently an interpretative process, not a mere linguistic substitution. ChatGPT's improved performance in rendering culturally and legally significant concepts points to its comparative ability to approximate this interpretative dimension, albeit imperfectly. Meanwhile, Google Translate and DeepL's more literal or awkward renderings highlight the risks of relying on purely automated systems that do not integrate contextual and institutional knowledge integral to legal meaning.

We also shed light on the critical necessity of the functioning translator's agency, echoing insights on the indispensability of human expertise and ethical responsibility in legal translation workflows. Given the complexity of legal language and its performative power, the study reaffirms the viewpoint that machine translation technologies, while promising, remain assistive tools rather than substitutes for professional translators.

Methodological transparency, especially regarding tool versions, prompt strategies, and input constraints, responds directly to calls for reproducibility and ethical research practices within Translation Studies. However, as acknowledged, the limited number of analyzed excerpts imposes constraints on the generalizability of findings, necessitating cautious interpretation while providing a valuable pilot perspective.

Overall, this study substantiates theoretical and practical claims regarding the intricate interplay between legal language, culture, and machine translation technology, offering empirical evidence aligned with existing scholarship and pointing to directions for future research.

6.1 AI in the Institutional Context

Deploying large language models (LLMs) in legal workflows introduces complex challenges concerning traceability, accountability, and legal certainty. These technologies often operate as opaque systems, making it difficult to verify the reasoning behind generated outputs or to attribute responsibility in cases of error or bias. Such concerns echo growing calls within the field of legal informatics and AI ethics for an “ethics of transparency”, emphasizing the need for explainable models, auditable processes, and clear accountability frameworks to ensure that technological efficiency does not undermine the fundamental principles of justice and legal reliability.

7. Limitations and Directions for Future Research

7.1 Sample constraint

This study was limited to seven legal text excerpts, which constrains the generalizability of the findings. While these samples provided meaningful insights into the comparative performance of machine translation systems, they serve primarily as groundwork for broader, statistically robust investigations. Future research should include a larger and more diverse dataset encompassing different legal genres and text complexities to enable quantitative analysis and enhance the validity of conclusions.

7.2 Evolving tool parameters

The rapid evolution of neural machine translation (NMT) and large language models (LLMs) poses challenges for reproducibility and longitudinal comparison. Since model architectures and parameter

updates can occur frequently and without detailed public disclosure, results obtained at one point in time may not remain stable or replicable in future studies. Therefore, meticulous reporting of tool versions, access dates, prompts, and settings is essential to ensure transparency and allow future researchers to contextualize findings within a specific technological state.

7.3 Cross-jurisdictional and cross-linguistic considerations

Given that legal language is inherently shaped by cultural, linguistic, and jurisdictional contexts, future research should broaden its scope to include a wider range of languages and legal systems. Comparative studies that involve civil law, common law, and hybrid systems could illuminate how translation tools handle varying terminologies, syntactic structures, and conceptual frameworks. Incorporating corpus-based methodologies would further strengthen analysis by enabling the identification of recurring translation patterns, terminological inconsistencies, and culturally specific renderings across diverse legal domains.

8. Ethical and Professional Implications

Professional translators, legal practitioners, and technology developers must remain critically attentive to several emerging concerns in the integration of machine translation (MT) and large language models (LLMs) within legal contexts. First, confidentiality and data privacy are paramount when handling sensitive legal materials through online MT platforms, as such systems may expose privileged information to security risks or third-party access. Second, post-editing and rigorous quality assurance remain essential, since human oversight ensures that translations achieve the precision, clarity, and contextual appropriateness demanded by legal communication. Even as

automated systems grow increasingly sophisticated, certified human translators continue to play an indispensable role in verifying linguistic and legal accuracy. Finally, the rise of AI technologies is transforming translator identity and professional training. LLMs challenge traditional pedagogical models in Translation Studies, shifting the translator's role from a mere linguistic executor to a critical mediator and quality controller. Consequently, training programs must incorporate technological literacy, ethical competence, and interdisciplinary awareness to prepare future professionals for this evolving landscape.

9. Conclusion

This research fortified the understanding that translating Arabic legal texts demands more than surface-level linguistic transfer – it requires cultural sensitivity, legal interpretative competence, and adherence to professional ethical standards. Evaluated against criteria informed by our referenced literature, ChatGPT consistently delivered more accurate, clear, fluent, and culturally appropriate translations compared to Google Translate and DeepL. However, the persistent shortcomings observed across all three systems – particularly for Arabic – suggest a deeper, systemic limitation. These shortcomings may be attributed to the insufficient integration of Arabic data into large-scale digital training corpora compared to higher-resource languages. This aligns with the perspective that the quality of machine translation output is intrinsically linked to the volume, variety, and domain-specificity of its underlying linguistic data. Consequently, while AI tools like ChatGPT show promising advancements, this study reinforces that human oversight remains indispensable in legal translation. Machine translation serves best as a powerful augmentation tool for legal translators, not a replacement – especially for languages with less extensive digital representation. Recognizing the preliminary nature of this study due to its limited dataset, these findings offer a meaningful contribution that bridges theoretical frameworks with practical evaluation. This work opens clear avenues

for future inquiry, not only into AI's evolving role but also into the critical need to develop richer, legally-oriented Arabic datasets to better balance the demands of law, language, and culture in the digital age.

Disclosure Statement:

The authors declare that they have no relevant financial or non-financial competing interests to disclose.

References

- Al-Bahūtī, M. B. Y. (1983). *Kashshāf al-qinā' 'an matn al-iqnā'* (Vol. 2). Dār Al-'Ilm li-l-Malayīn.
- Al-Faḍīlāt, J. M. (n.d.). *Binā' al-usrah al-muslimah 'alā ḍaw' al-fiqh wa-l-qānūn*. Dār Al-Shihāb.
- Al-Jurjānī, A. B. M. A.-S. (1986). *Al-ta'rīfāt. Dār Al-Shu'ūn Al-Taḳāfiyyah Al-'Āmmah*.
- Al-Kāsānī, A. A. B. M. (1986). *Badā'i' al-ṣanā'i' fī tartīb al-sharā'i'* (Vol. 2). Dār Al-Kutub Al-'Ilmiyyah.
- Al-Zarkashī, B. A. M. (1993). *Sharḥ Mukhtaṣar al-Kharqī* (Vol. 4). Maktabat al-'Ubīkayn.
- Baker, M. (1992). *In other words: A coursebook on translation*. Routledge.
- Bang, Y., Cahyawijaya, S., Lee, N., Dai, W., Su, D., Wilie, B., Lovenia, H., Ji, Z., Yu, T., Chung, W., Do, Q. V., Xu, Y., & Fung, P. (2023). *A multitask, multilingual, multimodal evaluation of ChatGPT on reasoning, hallucination, and interactivity*. arXiv. <https://doi.org/10.48550/arXiv.2302.04023>
- Belhadj, L. (2005). *Al-wajīz fī sharḥ qānūn al-usrah al-jazā'irī* (Part 1, 4th ed.). Dīwān Al-Maṭbū'āt Al-Jāmi'iyyah.
- Benounane, H. M., & Naceur, D. (2020). On the current state of machine translation: Investigating the statistical approach limitations and the neural model implications. *Academic Review of Social and Human Studies*, 12(2), 38–47.

- Biel, L. (2014). *Lost in the Eurofog: The textual fit of translated law*. In *Translation and the law* (pp. 57–76). Routledge.
- Boitet, C., Blanchon, H., Seligman, M., & Bellynck, V. (2009). Evolution of MT with the web. In *Proceedings of the Conference on Machine Translation* (Vol. 25, pp. 1–13).
- Canavese, P. (2023). Translation and clarity in Swiss multilingual legislation: A qualitative study on textuality. *Translation & Interpreting*, 15(2), 1–20.
<https://doi.org/10.12807/ti.115202.2023.a05>
- Cao, D. (2007). *Translating law*. Multilingual Matters.
- Cheng, Y., & Cheng, Y. (2019). Semi-supervised learning for neural machine translation. In *Joint training for neural machine translation: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers) (pp. 1965–1974). Association for Computational Linguistics.
- Colson, J. (1993). Les instruments de la docilité. In *Le traducteur et ses instruments*. Presses de la Sorbonne Nouvelle.
- Cornu, G. (2005). *Linguistique juridique*. Éditions Montchrestien.
- Cotelli Kureth, S., Delorme Benites, A., Haller, M., Noghrechi, H., & Steele, E. (2023). “I looked it up in DeepL”: Machine translation and digital tools in the language classroom. In *Human translation and natural language processing: Towards a new consensus?* (pp. 81–96). Fondazione Università Ca’ Foscari. <https://doi.org/10.30687/978-88-6969-762-3/006>
- Dong, Q., Li, L., Dai, D., Zheng, C., Ma, J., Li, R., Xia, H., Xu, J., Wu, Z., Liu, T., Chang, B., Sun, X., Li, L., & Sui, Z. (2023). A survey on in-context learning. arXiv. <https://doi.org/10.48550/arXiv.2301.00234>
- Durieux, C. (1992). La terminologie en traduction technique: apports et limites. *Terminologie et Traduction*, 2(3).
- Gao, Y., Wang, R., & Hou, F. (2023). *How to design translation prompts for ChatGPT: An empirical study*. arXiv. <https://doi.org/10.48550/arXiv.2304.02182>
- Garzone, G. (2000). Legal translation and functionalist approaches: A contradiction in terms? In *Translation studies: A reader* (pp. 127–136). St. Jerome Publishing.
- Haleem, A., Javaid, M., & Singh, R. P. (2022). An era of ChatGPT as a significant futuristic support tool: A study on features, abilities, and challenges. *BenchCouncil Transactions on*

- Benchmarks, Standards and Evaluations*, 2(4), 100089.
<https://doi.org/10.1016/j.tbench.2023.100089>
- Hidalgo-Tenero, C. (2020). Google Translate vs. DeepL: Analyzing neural machine translation performance under the challenge of phraseological variation. *MonTI. Monografías de Traducción e Interpretación, Special Issue* 6, 154–177.
<https://doi.org/10.6035/MonTI.2020.ne6.5>
- House, J. (2015). *Translation quality assessment: Past and present*. Routledge.
- Ibn Manzūr, M. (1999). *Lisān al-‘Arab* (Vol. 2, 3rd ed.). Dār Ihyā’ al-Turāth al-‘Arabī & Mu’assasat al-Ta’rīkh al-‘Arabī.
- Ibn Qudāmāh, A. A. (1984). *Al-Mughnī* (Vol. 7). Dār Al-Fikr.
- James, G. (2020). *Introduction to Google Translate*. Gilad James Mystery School.
- Kamaluddin, M. I., Rasyid, M. W. K., Abqoriyyah, F., & Saehu, A. (2024). Accuracy analysis of DeepL: Breakthroughs in machine translation technology. *Journal of English Education Forum* (JEEF), 4(2), 122–126.
- Khoshafah, F. (2023). ChatGPT for Arabic-English translation: Evaluating the accuracy. *Research Square*.
<https://doi.org/10.21203/rs.3.rs-2814154/v2>
- Liu, L., Utiyama, M., Finch, A., & Sumita, E. (2016). *Neural machine translation with supervised attention*. arXiv preprint arXiv:1609.04186.
- Mohanty, N. (2005). The limits of creativity and the translator’s responsibility. *Translation Today*, 2(2), 129–148.
- Moneus, A. M., & Sahari, Y. M. (2024). Artificial intelligence and human translation: A contrastive study based on legal texts. *Heliyon*, 10(6), e28106.
- Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). BLEU: A method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics* (pp. 311–318). Association for Computational Linguistics.
- Roser Nebot, N. (2019). El papel de la exégesis en la traducción del Corán. *REIDOCREA*, 8, 231–249.
- Sacco, R. (1999). Langue et droit. In R. Sacco & L. Castellani (Eds.), *Les multiples langues du droit européen uniforme* (pp. 126–143). L’Harmattan.
- Šarčević, S. (1997). *New approach to legal translation*. Kluwer Law International.

- Snell-Hornby, M. (1988). *Translation studies: An integrated approach*. John Benjamins. <https://doi.org/10.1075/z.38>
- Stiti, A., & Benaouda, A. (2023). Translating Algerian laws using neural machine translation. *Cahiers de Traduction*, 28(1), 421–445.
- Turovsky, B. (2016, November 15). *Found in translation: More accurate, fluent sentences in Google Translate*. Google Blog.
- Venuti, L. (1995). *The translator's invisibility: A history of translation*. Routledge.
- Williams, J., & Chesterman, A. (2002). *The map: A beginner's guide to doing research in translation studies*. St. Jerome Publishing.
- Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., & Dean, J. (2016). *Google's neural machine translation system: Bridging the gap between human and machine translation*. arXiv preprint arXiv:1609.08144.
- Xie, Y., Yu, C., Zhu, T., Bai, J., Gong, Z., & Soh, H. (2023). *Translating natural language to planning goals with large-language models*. arXiv. <https://doi.org/10.48550/arXiv.2302.05128>
- Zhao, Y., Zhang, M., Chen, X., Deng, Y., Geng, A., Liu, L., Liu, X., Li, W., Jiang, Y., Yang, H., Han, Y., Tao, S., Xie, N., Li, X., Ma, M., & Zhang, Z. (2023). Human evaluation for translation quality of ChatGPT: A preliminary study. In *Proceedings of the International Conference HiT-IT 2023* (pp. 282–287). https://doi.org/10.26615/issn.2683-0078.2023_023