AI and the BoLC:
Streamlining legal translation

Patrizia GIAMPIERI, PhD

University of Camerino, Italy
University of Malta, Malta
patrizia.giampieri@unicam.it

ORCID: http://orcid.org/0000-0003-4443-1593

Abstract: Artificial Intelligence (AI) applied in the legal field has gained considerable ground in the recent years and is used in many fields, amongst which the legal one is of no exception. This paper wishes to explore the quality of the translation (from English into Italian) of an arbitration clause performed by the ChatGPT chatbot. To do so, the automatically generated target text is post-edited by consulting the BoLC (Bononia Legal Corpus) and the web as corpus in the gov.it Italian governmental domain. General and legal dictionaries are also used. The paper findings report some inaccuracies at word level which are easily tackled by corpus consultation. In view of the results obtained, however, complete reliance upon AI-driven solutions is not recommendable at the time being. Conversely, the use of ad hoc corpora and of targeted web searches are the most feasible and reliable solutions, although time-consuming.

Keywords: corpus-based translation; legal translation; legal corpora; AI; machine translation.

Intelligenza Artificiale ed il corpus BoLC: ottimizzare la traduzione giuridica

Sommario: L'Intelligenza Artificiale applicata in ambito giuridico ha guadagnato notevole terreno negli ultimi anni e viene utilizzata in molteplici ambiti, tra cui quello legale. Questo articolo intende esplorare la qualità della traduzione (dall'inglese all'italiano) di una clausola compromissoria eseguita dalla chatbot ChatGPT. A tale scopo, il testo tradotto automaticamente è post-editato consultando il BoLC (Bononia Legal Corpus) e il web come corpus nel dominio *gov.it*. Sono inoltre impiegati dizionari generali e giuridici. I risultati dell'analisi evidenziano alcune inesattezze a livello lessicale che sono facilmente corrette attraverso la consultazione del corpus. Alla luce dei risultati ottenuti, tuttavia, al momento non è auspicabile affidarsi completamente a soluzioni basate sull'intelligenza artificiale. Al contrario, l'uso di corpora specifici e di ricerche web mirate rappresentano le soluzioni più adeguate ed affidabili anche se richiedono tempo.

Parole chiave: traduzione basata sui corpora; traduzione legale; corpora giuridici; intelligenza artificiale; traduzione automatica.

1. Introduction

1.1 Legal language and legal translation

Translating legal texts is challenging due to the fact that legal translation implies rendering not only terms from a source language into a target language, but it also requires verifying correspondences in the source and target legal systems (Jacometti and Pozzo 2018: 12). Also, legal translation is demanding due to the preponderance of legalese elements (Coulthard and Johnson 2007; Bhatia 2010; Tiersma 1999; Tiersma and Solan 2012; Williams 2023). Legalese is the archaic and intricate form of legal language, characterised by long sentences, near-synonyms (i.e., words with the same or very similar meanings, such as "last will and testament", Bhatia 2010: 28), polysemy (words that can have multiple meanings, such as "instrument", Bhatia 2010: 28), passive constructs (whereby the "agent" of an action is omitted), complex adverbial or prepositional

constructs (such as "for the purposes of and in connection with", or "notwithstanding", Bhatia 2010: 26), cataphora and anaphora (namely, prepositional or adverbial elements which refer back or forward to other elements in a given text, such as "hereto", "thereof", Abate 1998: 14–16), Latinism (as in "ab initio", Bhatia 2010: 367, or "ipso facto", Berezowski 2021: 38), vagueness (as in the expression "or whatsoever", or "or otherwise") and ambiguity (such as the one in the modal verb "shall", when it is used to impose obligations rather than to describe future actions; Williams 2023: 32–33, 156–160).

1.2 AI and legal translation

Artificial Intelligence was discovered and defined early in the 1950s. McCarthy et al. (1955), for example, postulated the possibility for machines to generate intelligent human behaviours.

The application of AI in the legal field firstly appeared in the 21st century, mainly thanks to advancements in machine learning (ML) and natural language processing (NLP). The former was investigated because it helped reduce human intervention in the performance of repetitive tasks (as in "labour-intensive processes", Brooks et al. 2020: 139). The latter has gradually changed existing practices in law firms (Brooks et al. 2020). In particular, programming, computational linguistics and NLP have largely contributed to the development of AI (Moneus and Sahari 2024).

Given the complexity of the language of the law, it is evident that legal translators should only use *ad hoc* language tools. At the same time, however, in view of the limited number of reliable and system-specific legal language resources available, translators may opt for non-targeted sources and/or AI-driven solutions, such as machine translation.

The use of chatbots such as ChatGPT, for example, has dramatically increased over the last few years (Thomson Reuters Institute 2023a). ChatGPT is a chat-based interface developed by OpenAI. It is applied by a wide array of users, amongst which are law firms and legal professionals (Thomson Reuters Institute 2023a).

AI-driven interfaces, in fact, present many advantages, such as the possibility to automate labour-intensive processes, perform

contract drafting and review, as well as expedite daily operations and work-flows (Brooks et al. 2020). Nonetheless, there is also scepticism over an uncritical approach to AI. Some scholars, for example, describe AI as generating "hallucinations" in the form of nonsense responses (Guerreiro et al. 2023; Thomson Reuters Institute 2023a). Others argue that AI "hallucinations" are actually outright fabrications and, hence, they should never be quoted or used in professional and academic settings (Emsley 2023).

The reason why this paper explores AI-driven tools such as ChatGPT lie in the fact that, as mentioned, it is a widespread tool amongst lawyers (Gigi et al. 2023). As shown in Thomson Reuters Institute's report (2023b: 3), for example, it took 5 days for ChatGPT to reach 1Mln users in 2022, compared to 2.5 months for Instagram in 2010 and 10 months for Facebook in 2004.

1.3 Corpora and legal translation

Corpora are considered helpful in legal translation because they allow users to notice collocations (i.e., words that are frequently found together), formulae (or formulaic expressions) and retrieve ready-made phrases in given contexts (Vigier Moreno 2016; Vigier Moreno and Sánchez Ramos 2017; Pontrandolfo 2019; Giampieri 2024). Corpora can also be applied in the post-editing of machine-translated output (Giampieri 2023a) because they offer an array of translation options and, hence, help cater for MT shortcomings (Giampieri 2023a).

Examples of legal corpora are the BoLC (Bononia Legal Corpus, Rossini Favretti et al. 2007), and the BLaRC (British Law Report Corpus, Rizzo and Pérez 2012). Some EU-based legal corpora are, for instance, the Europarl v7 (Koehn 2005; Tiedemann 2012), the JRC EU DGT Translation Memory Parsebank (Ljubešić and Erjavec 2018), the JRC-Acquis corpus (Steinberger et al. 2006), and the Multi-Eurlex (Chalkidis et al. 2021).

2. Research questions

This paper is aimed at investigating whether the BoLC can be successfully integrated in an AI-driven translation process and used to post-edit AI-generated texts.

To this aim, a contract clause is translated from English into Italian by using the ChatGPT chatbot. The automatically generated target text is then post-edited by consulting the Italian sub-section of the BoLC, the web as corpus (via advanced searches in the Italian governmental *gov.it* domain), and general and legal dictionaries.

The research questions that this paper wishes to address are the following ones: Can a legal corpus be consulted by lawyers, students in Translation Studies and legal translators to effectively address the shortcomings of AI-generated translations? To what extent can the BoLC help post-edit AI-produced target texts?

3. Methodology

As mentioned, this paper provides an in-depth analysis of the automated translation of an arbitration clause. In particular, the arbitration clause is firstly translated from English into Italian by using the ChatGPT chatbot, then it is post-edited by consulting the Italian sub-section of the BoLC, as well as the Hoepli bilingual dictionary and targeted web searches in the *.gov.it* domain.

As regards the automated translation, the following prompt is written in the ChatGPT field: "Translate the following text from English into Italian: '[*text to translate*]'". As can be seen, the language of the prompt is plain and straightforward. Therefore, it would be quite difficult for the AI solution to misinterpret it. Nonetheless, a test can be carried out to verify the effectiveness of the prompt. By changing the wording of the prompt, the results do not vary. For example, by entering the following test prompt: "Write in Italian the following text: '[*text*]'", the target text obtained is the same as the one of the first prompt (i.e., "Translate the following text from English into Italian: '[*text to translate*]'").

The source text is shown in Table 1 below.

Table 1. Text to translate

> All questions of differences whatsoever which may at any time hereafter arise between the parties hereto or their respective representatives touching these presents or the subject matter thereof or arising out of or in relation thereto respectively and whether as to construction or otherwise, shall be referred to arbitration in England, in accordance with the provisions of the Arbitration Act 1950 or any re-enactment or statutory modification thereof for the time being in force.

As mentioned, the clause above relates to arbitration. Arbitration is a dispute resolution method alternative to court trials. The parties to a contract may refer to arbitration in case of litigation, misunderstandings or differences in the interpretation or performance of a contract. By resorting to arbitration, the parties must abide by the final decision (referred to as an "award") issued by an arbitrator, who is an expert in the specific subject-matter of the contract.

As can be seen from Table 1, the clause contains some *legalese* elements, such as repetitions (such as the phrases "arising out of or in relation thereto", and "any re-enactment or statutory modification thereof"); vagueness (such as "whatsoever" and "or otherwise"); deixis ("hereafter", "thereof", "thereto"); polysemy (as in the word "construction", meaning "interpretation" in legal contexts); formulaic expressions (such as "or any re-enactment or statutory modification thereof"); ambiguity in the use of modal verbs ("shall"), and archaic language (such as "touching these presents"). In addition, the word "statutory" may generate false cognates in Italian, given its similarity with the word *statutario* (meaning "deriving from a by-law") (Giampieri 2024: 78).

It makes sense to explore the automated translation of AI solutions such as ChatGPT for various reasons. Firstly, as mentioned above, such a tool is widely used in the legal profession. Secondly, machine translation (MT) platforms are AI-driven interfaces which are resorted to by many law firms (Nunes Vieira et al. 2021; Giampieri 2023b: 22–25). Consequently, it can be useful to analyse the automated translation performed by a popular AI tool and verify whether it can be relied upon. In addition, for reasons of clarity and for comparisons, Appendix 1 reports the automated translations performed by ChatGPT and by DeepL (an MT platform). As can be

seen, apart from the word "construction" (which is misinterpreted by DeepL), the target texts are very similar.

It is now interesting to explore whether and to what extent ChatGPT can successfully render the source text reported in Table 1. As mentioned, post-editing is performed by consulting the Italian sub-section of the BoLC, together with evidence produced by targeted web searches in the *.gov.it* Italian governmental domain and general or legal dictionaries.

4. Analysis

Table 2 below reports the source text, the ChatGPT-generated target text and its (human-made) back-translation.

Table 2. Arbitration clause: source text, AI-generated target text and its back-translation

Source Text	ChatGPT translation into English	Human-made back-translation
All questions of differences whatsoever which may at any time hereafter arise between the parties hereto or their respective representatives touching these presents or the subject matter thereof or arising out of or in relation thereto respectively and whether as to construction or otherwise, shall be referred to arbitration in England, in accordance with the provisions of the Arbitration Act 1950	*Tutte le questioni di qualsiasi natura di differenze che possano insorgere in qualsiasi momento in futuro tra le parti qui presenti o i loro rispettivi rappresentanti in merito a queste disposizioni o all'oggetto della stessa o che sorgano in relazione ad esse rispettivamente e sia in termini di interpretazione o altro, saranno deferite all'arbitrato in Inghilterra, in*	All questions of difference of any nature that may arise at any time in the future between the parties here present or their respective representatives with regard to these provisions or the subject of the same or that arise in relation to them respectively and whether in terms of interpretation or other, will be deferred to arbitration in England, in conformity with the provisions of

or any re-enactment or statutory modification thereof for the time being in force.	*conformità alle disposizioni dell'Arbitration Act 1950 o di qualsiasi sua rinnovazione o modifica statutaria in vigore al momento.*	the Arbitration Act 1950 or of any of its renewal or by-law modification in force at the moment.

For the purpose of the analysis, the text of the clause of Table 1 is split into separate segments. The sections which follow present detailed corpus-based analyses, thus highlighting the words and phrases searched for in the BoLC and the results obtained.

The first source phrase to address is "All questions of differences whatsoever which may at any time hereafter arise between the parties hereto or their respective representatives". The chatbot renders it literally (i.e., *tutte le questioni di qualsiasi natura di differenze che possano insorgere in qualsiasi momento in futuro tra le parti qui presenti o i loro rispettivi rappresentanti*). In particular, "differences" is rendered as *differenze*, and "hereto" is translated as *qui presenti* ("here present"). However the target words actually mean "controversies" (as regards "differences") and "here concerned" or better "of this agreement" (as regards "hereto").

To start with the post-editing process of the source phrase, equivalents of the word "differences" and of the verb phrase "arise between the parties hereto" are explored. By consulting the Hoepli bilingual dictionary, the words *disaccordo*, *controversia*, *divergenza*, *divario*, and *discrepanza* (back-translations: "disagreement", "controversy", "divergence", "divide" and "discrepancy") are suggested as possible translation options of "difference". As can be seen, amongst others the dictionary proposes minimal pairs, i.e., pairs of words that differ in only one phonological element.

By writing the query *(("qualunque"|"qualsiasi") ("divergenza"|"disaccordo"|"divario"|"discrepanza")* in the Italian sub-section of the BoLC, it is possible to verify whether the words *qualunque* and/or *qualsiasi* (back-translation of both: "any") collocate with the various translation options proposed by the Hoepli dictionary. The straight line ("|") triggers the OR Boolean operator. In light of the results obtained, it is possible to notice the prevalence of the expression *qualsiasi divergenza* (3 hits, against 1 of *qualsisasi*

discrepanza). This is confirmed by targeted web searches. If the string *"qualunque|qualsiasi divergenza|disaccordo|divario|discrepanza" site:.gov.it* is queried in the Google search field, it is possible to notice many results with phrases such as *qualsiasi divergenza*, whereas *qualsiasi discrepanza* collocates with numbers, data or figures, not people or situations.

As regards the verb phrase "arise between the parties hereto", the Hoepli dictionary proposes *sorgere*, *insorgere*, and *presentarsi* (back-translations: "rise", "arise", "present itself") as translation candidates of "arise". As can be noticed, the dictionary proposes other minimal pairs, which corpus consultation can help further investigate.

By writing the following query in the BoLC (*"insorgere"|"sorgere"|"presentarsi") []{0,6} "parti"*, the various verbs proposed by the dictionary are queried together with the noun *parti* (back-translation: "parties") within a span of 0-6 words. In particular, the verb phrase *insorgere tra le parti* comes to the fore and there are no adverbial elements corresponding to "hereto" which follow. Sample phrases are the following: *ogni controversia che dovesse insorgere tra le parti* (back-translation: "any controversy which should arise between the parties"), and *l'insorgere di una controveria tra le parti* (back-translation: "the arise of a controversy between the parties"). The verb *insorgere* is a correct equivalent, as it collocates with "controversy", which is the subject-matter addressed by the clause. This is confirmed by targeted web searches. If the string *"insorgere|sorgere|presentarsi * parti" site:.gov.it* is googled, it is possible to read verbs such as *insorgere* and *sorgere*, both collocating with *controversia* and *divergenza*. Also in this case, there are no apparent equivalents of "hereto". Hence, *qui presenti* (as suggested by ChatGPT) is considered redundant and can be omitted in the target text. Therefore, the chatbot-generated verb *insorgere* can be considered as partly correct.

On the basis of the results obtained, the phrase "All questions of differences whatsoever which may at any time hereafter arise between the parties hereto or their respective representatives" is post-edited and translated as follows: *Qualsiasi divergenza [controversia] di qualunque tipo che possa insorgere [sorgere] in qualsiasi momento tra le parti o i loro rappresentanti.*

The next challenging phrases are the following ones: 1) "touching these presents", 2) "subject matter thereof" and the near-synonyms 3) "arising out of or in relation thereto". ChatGPT renders

them as follows: 1) *in merito a queste disposizioni,* 2) *all'oggetto della stessa,* and 3) *che sorgano in relazione ad esse*. As can be seen, the chatbot translates "these presents" as *queste disposizioni* (back-translation: "these provisions"). However, "these presents" actually means "this document". The Black's Law Dictionary, in fact, explains "these presents" as "the instrument under consideration". The Hoepli dictionary proposes *riguardare*, *toccare*, and *concernere* (back-translations: "regard", "touch", "concern") as translation candidates of "touch", whereas it generates *documento*, *dichiarazione*, *scritto*, and *affermazione* (back-translations: "document", "declaration", "writing", "affirmation / statement") as equivalents of "presents". By writing the following search string in the BoLC: *("riguardante"|"concernente"|"inerente")* *[]{0,3}* *("atto"|"document")*, it is possible to notice *inerente all'atto* and *inerente ad un atto* (back-translation: "concerning the/a deed"), which are the most recurrent results (14 hits). Moreover, as the whole sentence deals with disputes related to a contractual relationship, the terms *controversia* and *divergenza* (back-translation: "controversy" and "divergence") can be queried together with *documento*, *atto*, or *contratto* (i.e., "document", "deed" or "contract", respectively). Unfortunately, there are no relevant hits resulting from the search string *("controversia"|"divergenza")* *[]{0,3}* *("atto"|"document"|"contratto")*. For example, the following unrelated phrase is read: *controversia in atto* (back-translation "current controversy"). The same is obtained by carrying out targeted web searches. If the search string *"controversia|divergenza * atto|documento|contratto" site:.gov.it* is googled, the results obtained mainly revolve around unrelated expressions such as *controversia in atto* (back-translation: "ongoing controversy"). Nonetheless, some other hits show interesting formulae, such as *controversia derivante dal presente contratto* (back-translation: "controversy arising out of this contract"). Therefore, "touching these presents" can be post-edited and rendered as *derivante dal presente contratto*, or *inerente al presente atto*.

As far as the deictic phrase "subject matter thereof" is concerned, the chatbot suggests *oggetto della stessa* (back-translation: "subject of the same"). The Hoepli dictionary proposes *oggetto* and *argomento* as translations of "subject matter" or "subject-matter". It is, hence, helpful to search for *riguardante*, *concernente* and *inerente* (back-translations: "regarding", "concerning", and "inherent")

together with *argomento* and *oggetto* (back-translations: "topic" and "subject") in the BoLC. The following search query is written: *("riguardante"|"concernente"|"inerente")* []{0,8} *("argomento"|"oggetto")*. In this way, some phrases are obtained, such as *concernente (...) e il medesimo oggetto* (back-translation: "concerning (…) and the same subject"), and *riguardante il medesimo oggetto* (back-translation: "regarding the same subject"). The word *medesimo* (back-translation: "the same") can be considered as an acceptable equivalent of "thereof". The Proz translators' forum also suggests *materia oggetto della medesima* as a translation of "subject matter thereof"[1]. As *stesso* is a synonym of *medesimo*, the string *"materia" "oggetto" ("dello"|"della"|"del") ("medesima"|"medesima"|"stesso"|"stessa")* is queried in the BoLC; as a result, the phrase *materia oggetto dello stesso* is found. Also, by investigating the *gov.it* domain, the formula *materia oggetto del regolamento stesso* comes to the fore (search string: *"materia oggetto dello|della|del * medesima|medesima|stesso|stessa" site..gov.it*). Therefore, to some extent, the chatbot-produced target phrase *oggetto della stessa* can be considered acceptable.

The latter deictic phrase, i.e., "arising out of or in relation thereto", is rendered as *che sorgano in relazione ad esse* (back-translation: "that rises in relation to them") by ChatGPT. It can be intuitively guessed that "arising out of" and "in relation thereto" mean *derivante* and/or *riguardante* (back-translation: "deriving" and "regarding"). Hence, by querying *"derivante" []{0,10} "riguardante"* in the BoLC, it is possible to obtain the following phrases: *derivante dall' (...) riguardante il medesimo* and *derviante dalla (...) riguardante la stessa*. In light of these results, the source phrase "touching these presents or the subject matter thereof or arising out of or in relation thereto respectively" can be post-edited and translated as *inerente al presente atto [derivante dal presente contratto] o alla materia oggetto del medesimo [dello stesso], o derivante da o riguardante lo stesso [il medesimo]*. As can be seen, the corpus-based post-editing is more formal than the AI-generated target text.

The next phrase to tackle is "and whether as to construction or otherwise, shall be referred to arbitration in England". In this expression, the polysemous word "construction", the vagueness

---

[1] See https://www.proz.com/kudoz/english-to-italian/law-general/5172400-subject-matter-thereof.html. Accessed May 14, 2024.

element "or otherwise," and the verb "referred to" relating to "arbitration" may pose challenges due to their ambiguous nature. However, ChatGPT renders the whole phrase quite satisfactorily as follows: *e sia in termini di interpretazione o altro, saranno deferite all'arbitrato in Inghilterra* (back-translation: "and both in terms of interpretation or other, will be deferred to arbitration in England"). The Hoepli dictionary, in fact, suggests *interpretazione* as a translation of "construction" in legal contexts. As regards "or otherwise,", by searching for *"o" "altro" ","* in the BoLC (comma included) (back-translation: "or other,"), 8 hits are obtained. The phrase *"o" "altrimenti" ","* (back-translation: "or otherwise,") generates 12 hits. As can be seen, such results are almost equivalent in terms of occurrences and they both express vagueness. However, the latter (i.e., *o altrimenti,*) may be more adherent to the source phrase, as *altro* means "other", whereas *altrimenti* strictly translates "otherwise". Therefore, *o altro*, as suggested by ChatGPT should be replaced by *o altrimenti*.

The phrase "referred to arbitration" could be challenging due to the verb "referred" which must collocate with "arbitration". As mentioned, the chatbot renders it as *deferite ad arbitrato* (back-translation: "deferred to arbitration"). In this case, by searching for *"ad" "arbitrato"* in the BoLC, it is possible to notice the verb *sottoporre* (back-translation: "to subject / submit") before *ad abitrato*, together with other verbs, albeit less frequent, such as *deferire*, *devolvere*, *sottomettere* and *rimettere* (back-translations: "defer", "devolve", "submit" and "put", respectively). By googling *"sarà * ad arbitrato" site:.gov.it* (back-translation of the phrase: "will be * to arbitration"), the following verbs are obtained: *sottoposta*, *demandata*, and *devoluta* (back-translation: "submitted", "deferred", "devolved"). In light of the corpus-based analyses and of targeted web searches, the phrase "and whether as to construction or otherwise shall be referred to arbitration in England" can be post-edited and rendered as follows: *ed avente ad oggetto la sua interpretazione o altrimenti, sarà sottoposta [devoluta; deferita; sottomessa; rimessa] ad arbitrato in Inghilterra*. As can be noticed, corpus consultation provides an array of possible translation solutions, which are probably more adherent to legal language in use.

The next expression to tackle is the prepositional phrase "in accordance with the provisions of the Arbitration Act 1950". This phrase is interesting from several perspectives. Firstly, it is useful to

78

verify which Italian prepositional phrase can be considered equivalent of "in accordance with"; then, it is necessary to understand which words best render "provisions" collocating with "act" (meaning "law" or "statute"). ChatGPT addresses the source phrase as follows: *in conformità alle disposizioni dell'Arbitration Act 1950* (back-translation: "in conformity with the provisions of the Arbitration Act 1950"). As can be noticed, the chatbot "intelligently" transferred "Arbitration Act 1950" without translating it. As regards post-editing, the Hoepli dictionary suggests *in conformità con*, *conformemente a*, and *secondo*, as translations of "in accordance with". By searching for the following string *("in" "conformità" "a")|("conformemente" "a")|("secondo") []{0,4} ("legge"|"leggi")* in the BoLC, several results are obtained. Unfortunately, there is some "noise" due to many unrelated concordances, such as *secondo comma della legge* (back-translation: "second paragraph of the law"), or *secondo motivo, violazione di legge* (back-translation: "second reason, breach of law"). By taking out the noise-generating word *secondo* and re-writing the search string as follows: *("conformità"|"conformemente") []{0,4} ("legge"|"leggi")*, it is possible to retrieve more targeted and consistent results, such as *in conformità a quanto previsto dalla legge* (back-translation: "in accordance with what envisaged / foreseen by law"); *in conformità a quanto disposto dall'articolo (...) della legge* (back-translation: "in accordance with what provided for by art... of the law"), and *in conformità a quanto statuito dal legislatore* (back-translation: "in accordance with what established by the law drafter"). Similar results are noticed by carrying out advanced web searches in the *gov.it* domain. The search string *"conformità|conformemente * legge|leggi" site:.gov.it*, in fact, generates the following results: *in conformità alla legge* and *conformemente alla legge*. Therefore, the chatbot-driven target phrase *in conformità alle disposizioni* can be considered as satisfactory.

As regards the word "provision", the Hoepli dictionary proposes the following translation candidates: *clausola*, *disposizione*, *disposto*, and *dettato* (back-translation: "clause", "disposition / provision", "provision", "principle"). By querying *("clausole"|"disposizioni"|"disposto"|"dettato") []{0,6} "legge"* in the BoLC, the phrase *disposizioni di legge* (back-translation: "law provisions") comes to the fore and is preceded by the following adverbial or prepositional phrases: *conformemente alle*, *in ottemperanza a*, and *nel rispetto delle* (back-translation: "conforming

to", "by complying with", "by respecting the"). Also, by searching for *("conformità"|"conformemente")[]{0,3}("disposizioni")*, these phrases are noticed: *in conformità delle disposizioni legislative* (back-translation: "in compliance with legal provisions"); *in conformità con le specifiche disposizioni* (back-translation: "in compliance with specific provisions"), and *in conformità con tutte le disposizioni e procedure di legge* (back-translation: "in compliance with all law provisions and procedures"). By searching for *"conformità|conformemente * disposizioni" site:.gov.it*, results similar to the ones obtained from the BoLC are found, such as *conformemente alle disposizioni dell'articolo 12, paragrafo 6 del regolamento* (back-translation: "in accordance with the provisions of article 12, comma 6 of the regulations"), and *in conformità alle disposizioni di leggi e regolamenti* (back-translation: "according to the provisions of laws and regulations").

On the basis of the analysis carried out, the phrase "in accordance with the provisions of the Arbitration Act 1950" can be post-edited and rendered as *in conformità con le [conformemente alle] disposizioni della Arbitration Act 1950*. As can be noticed, the "Arbitration Act 1950" is transferred (i.e., kept) in the target text, given that it is the name of an English statutory document. The chatbot target phrase is, hence, correct.

The expression "or any re-enactment or statutory modification thereof for the time being in force" ends the clause and can be difficult to translate as it is a fixed formulaic expression whose equivalent must be found in the target language and legal system. This expression states that the original Arbitration Act 1950 is applicable, as well as any of its later modifications or integrations. ChatGPT renders it as follows: *o di qualsiasi sua rinnovazione o modifica statutaria in vigore al momento* (back-translation: "or of any of its renewal or by-law modification in force at the moment"). In this translation, the word *rinnovazione* ("renewal") and *statutaria* ("by-law(ish)") are erroneous, as laws are not "renewed" and no "by-law" modification is entailed. The Hoepli dictionary proposes the following translation candidates of "re-enactment": *ripristino*, *reintroduzione*, and *rimessa in vigore* (back-translation: "reinstatement", "reintroduction", "reactivation in force"). By writing the following query: *("ripristino"|"reintroduzione"|"rimessa") []{0,6} "legge"* and/or *"legge" []{0,6} ("ripristino"|"reintroduzione"|"rimessa")*, no relevant concordance is obtained. However, by querying *legge* (back-

translation: "law") together with *modifica* (back-translation: "change / modification" (search string: *"legge" []{0,10} ("modifica"|"modifiche")*), it is possible to read *legge n. (...) e successive modifiche* (back-translation: "law no.... and further modifications"), and *legge n. (...) e successive modifiche ed integrazioni*, (back-translation: "law no.... and further modifications and integrations"), which can be considered acceptable equivalents. The same results are obtained by querying *"legge e|o * modifica|modifiche" site:.gov.it* (e.g., *legge e successive modifiche e integrazioni*). Therefore, the chatbot-driven translation presents some shortcomings which must be addressed. The source phrase "or any re-enactment or statutory modification thereof for the time being in force" can be post-edited and translated as follows: *o successive modifiche e integrazioni*. Evidently, the phrase "for the time being in force" is not necessarily mentioned in Italian legal texts.

4.1 Proposed translation of the arbitration clause

On the basis of the analysis carried out above, the clause shown in Table 1 and 2 can be post-edited and translated as reported in Table 3 below.

The first column contains the source text, the second column the chatbot-driven translation and the third column the corpus-based post-editing and translation (alternative target terms are reported in squared brackets). Chatbot-generated shortcomings are underlined.

Table 3. Translation proposals of the arbitration clause

Source Text	ChatGPT translation into English	Corpus-based post-editing and translation
All questions of differences whatsoever which may at any time hereafter arise between the parties hereto or their respective representatives touching	*Tutte le questioni di qualsiasi natura di differenze che possano insorgere in qualsiasi momento in futuro tra le parti qui presenti o i loro rispettivi rappresentanti*	*Qualsiasi divergenza [controversia] di qualunque tipo che possa insorgere [sorgere] in qualsiasi momento tra le parti o i loro rappresentanti*

these presents or the subject matter thereof or arising out of or in relation thereto respectively and whether as to construction or otherwise, shall be referred to arbitration in England, in accordance with the provisions of the Arbitration Act 1950 or any re-enactment or statutory modification thereof for the time being in force.	*in merito a* <u>*queste*</u> <u>*disposizioni*</u> *o all'oggetto della stessa o che sorgano in relazione ad esse rispettivamente e sia in termini di interpretazione o altro, saranno deferite all'arbitrato in Inghilterra, in conformità alle disposizioni dell'Arbitration Act 1950 o di qualsiasi sua* <u>*rinnovazione*</u> *o* <u>*modifica*</u> <u>*statutaria*</u> *in vigore al momento.*	*inerente al presente atto [derivante dal presente contratto] o alla materia oggetto del medesimo [dello stesso], o derivante da o riguardante lo stesso ed avente ad oggetto la sua interpretazione o altrimenti, sarà sottoposta [devoluta; deferita; sottomessa; rimessa] ad arbitrato in Inghilterra, in conformità con le [conformemente alle] disposizioni della Arbitration Act 1950 o successive modifiche e integrazioni.*

As can be seen, the Italian translation contains vague language, such as *qualsiasi*, *qualunque*, (back-translation of both: "any"); *in qualsiasi momento* (back-translation: "at any time"), and *o altrimenti* (back-translation: "or otherwise"). There is also deixis, as in the phrases *presente atto* (back-translation: "present act"); *del medesimo* (back-translation: "of the same"), and *lo stesso* (back-translation: "the same").

The formulaic expression *o successive modifiche e integrazioni* can be considered the full equivalent of "or any re-enactment or statutory modification thereof". Adverbial elements such as "hereafter", "thereof", and "thereto" are rendered with deictic expressions, such as *del medesimo* ("of the same"), or possessive adjectives (as *sua* in the phrase *sua interpretazione*, back-translation: "its interpretation").

The target text proposed by ChatGPT is overall satisfactory, although it generates some shortcomings (underlined in Table 3 above), such as *queste disposizioni* (back-translation: "these provisions") to render "these presents" (actually meaning "this

document"), *rinnovazione* (back-translation: "renovation") to translate "re-enactment" (referring to the creation of a law) and *modifica statutaria* (back-translation: "by-law modification") which addresses "statutory modification" (actually meaning "modification of law").

As a final analysis, the target text proposed by ChatGPT can be compared with the automated translation proposed by DeepL (see Appendix 1). As can be noticed, some minor differences come to the fore. As mentioned, DeepL misinterprets the word "construction" and proposes a literal translation, i.e., *di natura costruttiva* (back-translation: "of a building/constructive nature"). In addition, the MT platform proposes *in futuro* (back-translation: "in the future") as a rendering of "at any time". Such a translation solution is debatable, as "at any time" is slightly different from "in the future". Conversely, the phrase "touching these presents" is correctly rendered as *che riguardino i presenti documenti* (back-translation: "regarding the present documents").

Finally, the formula "re-enactment or statutory modification" is translated literally as *riedizione o modifica statutaria* (back-translation: "re-edition or statute modification"). As already discussed, a law cannot be "edited". Also, the English "statutory" clearly refers to laws, whereas the Italian *statutaria* may be interpreted as relating to by-laws or articles of associations. These mistranslations were already noticed in the target text generated by ChatGPT.

In light of the above, the analyses carried out in this paper bring to the fore the fact that the translation proposed by the chatbot is slightly better than the one suggested by the MT platform. This partly explains the reason why law firms and legal professionals often recur to ChatGPT.

5. Discussion

As could be seen in the sections above, the BoLC search syntax is not particularly user-friendly. Writing the search query correctly is important in order to obtain consistent findings. In addition, as the BoLC is not equipped with lemmatisation or POS tagging, the search strings must be carefully designed. Writing a word instead of another may yield different results and, hence, different conclusions as to

possible equivalences. Moreover, targeted web searches require accuracy and knowledge of the syntax to use. These aspects, however, can be considered as part of a thorough translation process. Despite difficulties, both the BoLC and targeted web searches generated very interesting results which helped post-edit the AI-generated target text successfully. Translation options and (full or near) equivalents were, in fact, always retrieved. Some expressions were also sourced quite effortlessly, such as "or otherwise" and "in accordance with". Others required more attention and care, such as the formula "or any re-enactment or statutory modification thereof".

In the case in point, the AI-driven solution did not give rise to any "hallucination" or nonsense target phrase, although some inconsistent language solutions were produced (see the underlined words in Table 3). For these reasons, complete reliance upon AI tools must be ruled out.

In light of the paper findings, it can be argued that a variety of different digital tools (such as legal corpora and the web as corpus) should be introduced in the translation process, especially if AI solutions are resorted to. Corpora, in fact, can cater for the needs of legal translators and may be used as both post-editing and translation tools. This, however, is possible as long as users are acquainted with the tools' specific search syntax.

6. Conclusion

This paper was aimed at verifying to what extent an AI-generated target text can be trustworthy and whether additional language tools must be accounted for. The findings highlight the fact that, at the time being, chatbots such as ChatGPT are not to be used as a single translation tool, as erroneous translations may occur. The AI tool produced a target text which made sense and was not affected by "hallucinations". However, translators are always recommended using multiple language tools. In the case in point, corpora and the web as corpus proved to be satisfactory post-editing resources. It is evident that using the BoLC for contract-related translations can be foreseen provided that users are acquainted with its search syntax and, occasionally, can consult other *ad hoc* resources, such as dictionaries

and targeted web domains. The BoLC, in fact, is a useful language tool in the legal field, but it requires acquaintance with its advanced search commands. Also, its content may not be sufficient for all kinds of legal translations (e.g., contract-related ones). Therefore, a combination of digital tools is apparently the best approach to meet the needs of legal translators.

The first research question asked whether a legal corpus could be consulted by lawyers, students in Translation Studies and legal translators to effectively address the shortcomings of AI-generated translations. The answer to this question is affirmative provided that users are acquainted with corpus consultation techniques. The second question asked to what extent the BoLC helped post-edit AI-produced target texts. The sections above proved that without corpus consultation, the automated target text presented some minor and major shortcomings which would have not made it accurate and reliable. Therefore, the (manual) consultation of *ad hoc* resources is still mandatory.

It may be argued that the limits of this paper lie in the software tools chosen (i.e., the BoLC and the web as corpus). Quite to the contrary, it was the purpose of this paper to explore whether and to what extent the BoLC could be useful and resourceful, especially if consulted in conjunction with other targeted legal language resources. This paper showed that this platform can be a reliable option, especially if supplemented with other language tools.

Further research could explore other online publicly available resources that can be used to translate AI-generated legal texts or to post-edit automatically translated target texts.

Conflict of interest statement:

The author has no financial or proprietary interests in any material discussed in this article. There is no conflict of interest.

AI Use statement:

AI was used strictly for the purposes of translating the clause extract from English into Italian, as indicated in the paper.

Bibliography

Abate, Salvatore C. 1998. *Il documento legale anglosassone*. Milan: Hoepli.

Berezowski, Łukasz J. 2021. Translating Latin legalese: between dynamic equivalence and formal correspondence. A comparative study case of Latin legal terms and maxims in English, Italian and Polish language. In: *Traduttologia e Traduzioni Vol II*, ed. Joanna Ciesielka, 35–51. http://hdl.handle.net/11089/39976 (accessed May 14, 2024).

Bhatia, Vijay K. 2010. *Textbook on Legal Language and Legal Writing*. New Delhi: Universal Law Publishing Co.

Brooks, Chay, Cristian Gherhes, and Tim Vorley. 2020. Artificial intelligence in the legal sector: Pressures and challenges of transformation. *Cambridge Journal of Regions, Economy and Society*, 13(1): 135–152. DOI: https://doi.org/10.1093/cjres/rsz026.

Chalkidis, Ilias, Manos Fergadiotis, and Ion Androutsopoulos. 2021. MultiEURLEX – A Multi-lingual and Multi-label Legal Document Classification Dataset for Zero-shot Cross-lingual Transfer. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2021)*, on-line and Punta Cana, Dominical Republic. https://aclanthology.org/2021.emnlp-main.559 (accessed May 14, 2024).

Coulthard, Malcolm, and Alison Johnson. 2007. *An Introduction to Forensic Linguistics. Language in Evidence*. London/New York: Routledge.

Emsley, Robin. 2023. ChatGPT: these are not hallucinations – they're fabrications and falsifications. *Schizophrenia*, 9(52): 1–2. DOI: https://doi.org/10.1038/s41537-023-00379-4.

Giampieri, Patrizia. 2023a. Is machine translation reliable in the legal field? A corpus-based critical comparative analysis for teaching ESP at tertiary level. *ESP Today*, 11(1): 119–137. DOI: https://doi.org/10.18485/esptoday.2023.11.1.6.

Giampieri, Patrizia. 2023b. *Legal Machine Translation Explained: MT in legal contexts*. Newcastle Upon Tyne: Cambridge Scholars Publishing.

Giampieri, Patrizia. 2024. *The Use of Comparable Corpora on (General) Terms and Conditions as a Pedagogical Tool in Translation Training between English and Italian*. PhD dissertation. University of Malta. https://www.um.edu.mt/library/oar/handle/123456789/119427 (accessed May 14, 2024).

Gigi, Gisa M., Rajesh, Nikhitha, Sania, Ann, Mohan, Theresa, and Thekkanath Biju A. 2023. AI Contract: Future and Scope in Legal System. *International Journal of Advance and Applied Research*, 10(6): 237–240. DOI: https://doi.org/10.5281/zenodo.8344607.

Guerreiro, Nuno M., Elena Voita, and André F.T.Martins. 2023. Looking for a needle in a haystack: A comprehensive study of hallucinations in neural machine translation. *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, May 2023. DOI: https://doi.org/10.48550/arXiv.2208.05309 (accessed May 14, 2024).

Jacometti, Valentina, and Barbara Pozzo. 2018. *Traduttologia e linguaggio giuridico*. Milan: Wolters Kluwer.

Koehn, Philip. 2005. Europarl: A Parallel Corpus for Statistical Machine Translation. In: *The Tenth Machine Translation Summit Proceedings of Conference*, ed. John Hutchins, 79–86. International Association for Machine Translation. https://www.pure.ed.ac.uk/ws/portalfiles/portal/26315407/MTS_2005_Koehn.pdf (accessed May 14, 2024).

Ljubešić, Nikola, and Tomaž Erjavec. 2018. *JRC EU DGT Translation Memory Parsebank DGT-UD 1.0*. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1197 (accessed May 14, 2024).

McCarthy, John, Minsky, Martin, Shannon, Claude E., and Rochester, Nathanielì. 1955. "A proposal for the Dartmouth summer

research project on artificial intelligence". *AI Magazine*, 27(4): 12. DOI: https://doi.org/10.1609/aimag.v27i4.1904.

Moneus, Mohammed A. and Yousef Sahari. 2024. Artificial intelligence and human translation: A contrastive study based on legal texts. *Heliyon*, 10(6): 1–14. DOI: https://doi.org/10.1016/j.heliyon.2024.e28106.

Nunes Vieira, Lucas, O'Hagan, Minako, and O'Sullivan, Carol. 2021. Understanding the societal impacts of machine translation: a critical review of the literature on medical and legal use cases. *Information, Communication & Society*, 24(11): 1515–1532. DOI: https://doi.org/10.1080/1369118X.2020.1776370.

Pontrandolfo, Gianluca. 2019. Corpus methods in legal translation studies. In: *Research methods in legal translation and interpreting*, eds. Łucja Biel, Jan Engberg, Rosario Martín Ruano, and Vilelmini Sosoni, 13–28. London: Routledge.

Rizzo, Camino Rea, and Maria J.M. Pérez. 2012. Structure and design of the British Law Report Corpus (BLRC): a legal corpus of judicial decisions from the UK. *Journal of English Studies*, 10: 131–145. DOI: https://doi.org/10.18172/jes.184.

Rossini, Favretti, Rema, Fabio Tamburini, and Edoardo Martelli. 2007. Words from Bononia Legal Corpus. In: *Text Corpora and Multilingual Lexicography*, ed. Wolfgang Tuebert, 11–30. Amsterdam: John Benjamins.

Steinberger, Ralf, Bruno Pouliquen, Anna Widiger, Camelia Ignat, Tomaž Erjavec, Dan Tufis, and Dániel Varga. 2006. The JRC-Acquis: A multilingual aligned parallel corpus with 20+ languages. *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'2006)*. Genoa, Italy, 24–26 May 2006. http://www.lrec-conf.org/proceedings/lrec2006/pdf/340_pdf.pdf (accessed May 14, 2024).

Thomson Reuters Institute. 2023a. *ChatGPT and Generative AI within Law Firms*. https://www.thomsonreuters.com/en-us/posts/wp-content/uploads/sites/20/2023/04/2023-Chat-GPT-Generative-AI-in-Law-Firms.pdf (accessed May 14, 2024).

Thomson Reuters Institute. 2023b. *Stepping into the future: How generative AI will help lawyers improve legal service delivery*. https://legal.thomsonreuters.com/content/dam/ewp-m/documents/legal/en/pdf/reports/lawyers-harnessing-power-of-ai.pdf (accessed May 14, 2024).

Tiedemann, Jörg. 2012. Parallel Data, Tools and Interfaces in OPUS. *Proceedings of the Eight International Conference on Language Resources and Evaluation* (LREC'2012), European Language Resources Association, Istanbul, 2214–2218. http://www.lrec-conf.org/proceedings/lrec2012/pdf/463_Paper.pdf (accessed May 14, 2024).

Tiersma, Peter. 1999. *Legal language*. Chicago, IL: The University of Chicago Press.

Tiersma, Peter and Lawrence M. Solan. 2012. *The Oxford Handbook of Language and Law*. New York: Oxford University Press.

Vigier Moreno, Francisco. J. 2016. Teaching the Use of ad hoc Corpora. *Language and Law / Linguagem e Direito*, 3(1): 100–119. https://ojs.letras.up.pt/index.php/LLLD/article/view/1658 (accessed May 14, 2024).

Vigier Moreno, Francisco J., and María d. M. Sánchez Ramos. 2017. Using Parallel Corpora to Study the Translation of Legal System-Bound Terms: The Case of Names of English and Spanish Courts. In: *Computational and Corpus-Based Phraseology EUROPHRAS 2017, Lecture Notes in Computer Science*, ed. Ruslan Mitkov, 206–273. Cham: Springer.

Williams, Christopher. 2023. *The Impact of Plain Language on Legal English in the United Kingdom*. London/New York: Routledge.

Other Resources

ChatGPT: https://chat.openai.com (accessed May 14, 2024).

Hoepli dictionary: https://www.dizionari.repubblica.it (accessed May 14, 2024).

Garner, Bryan A. 2004. *Black's Law Dictionary*. Eighth edition. Eagan: Thomson West.

Appendix 1

Automated translation performed by ChatGPT and DeepL.

Source Text	Target Text (ChatGPT)	Target Text (DeepL)
All questions of differences whatsoever which may at any time hereafter arise between the parties hereto or their respective representatives touching these presents or the subject matter thereof or arising out of or in relation thereto respectively and whether as to construction or otherwise, shall be referred to arbitration in England, in accordance with the provisions of the Arbitration Act 1950 or any re-enactment or statutory modification thereof for the time being in force.	*Tutte le questioni di qualsiasi natura di differenze che possano insorgere in qualsiasi momento in futuro tra le parti qui presenti o i loro rispettivi rappresentanti in merito a queste disposizioni o all'oggetto della stessa o che sorgano in relazione ad esse rispettivamente e sia in termini di interpretazione o altro, saranno deferite all'arbitrato in Inghilterra, in conformità alle disposizioni dell'Arbitration Act 1950 o di qualsiasi sua rinnovazione o modifica statutaria in vigore al momento.*	*Tutte le questioni di divergenza di qualsiasi tipo che possano in futuro insorgere tra le parti o i loro rispettivi rappresentanti e che riguardino i presenti documenti o l'oggetto degli stessi o che derivino da o siano in relazione ad essi e che siano di natura costruttiva o di altro tipo, saranno deferite all'arbitrato in Inghilterra, in conformità alle disposizioni dell'Arbitration Act 1950 o di qualsiasi sua riedizione o modifica statutaria per il momento in vigore.*