Law and Humanities
ISSN Print: 3079-5117
ISSN Online: 3079-5125
https://doi.org/10.63313/LH.9023

Application and Limitations of Large Language Models in the Translation of the Contract Book of the Civil Code

Xiaoyu Li. Guifang Wang
Zhejiang Yuexiu University, Shaoxing, China

How to cite this paper: Li, X. Y., & Wang, G. F. (2025). Application and Limitations of Large Language Models in the Translation of the Contract Book of the Civil Code. Law and Humanities, 1(3), 31-40. ISSN Print: 3079-5117; ISSN Online: 3079-5125.
https://doi.org/10.63313/LH.9023
Published: 2025-07-07

Abstract

In recent years, the rapid advancement of Large Language Models (LLMs) has created new possibilities for legal translation. This study employs a combination of quantitative analysis and case studies to systematically compare the textual features of the authoritative human-translated version and the GPT-4o-translated version of the Contract Book of the Civil Code across three dimensions: terminological accuracy, syntactic complexity, and the handling of semantically ambiguous expressions. The findings are as follows: (1) In trans-lating the Contract Book of the Civil Code, the human version demonstrates su-perior rigor and precision in legal terminology, in contrast, the GPT-generated version displays deviations in term usage that may introduce ambiguities in le-gal interpretation. (2) Although the GPT version enhances readability by reduc-ing syntactic complexity, this simplification compromises the precision and formality required in legal discourse. (3) In addressing semantically ambiguous expressions, human translators employ terminological transformation, lexical supplementation, and tense modulation, while the GPT version tends to rely on literal translation, increasing the risk of omission or distortion in legal provi-sions. This study highlights the current limitations of LLMs in legal translation and underscores the importance of human-machine collaboration, offering in-sights and guidance for producing high-quality legal translations with the assis-tance of LLMs.

Keywords

Large Language Models; legal translation; terminological accuracy; human-machine collaboration

1. Introduction

The Large Language Model ChatGPT, released by OpenAI in 2022, quickly emerged as a prominent force in the field of AI-driven translation due to its advanced architecture, remarkable performance, and high-quality output. Its rise has sparked extensive academic discourse concerning the future development of machine translation (Li and Niu, 2024, p. 25). Based on an empirical study of legal text translation, Zhang (2020) identified persistent deficiencies in current AI systems, particularly in

the accurate rendering of legal terminology, the reproduction of legal stylistic features, and the logical coherence of legal provisions. These findings effectively challenge the "replacement theory" of machine translation. Against the backdrop of continuous advancements in digital translation technologies, a critical question arises: how can legal texts and LLMs be effectively integrated to achieve high-quality translation while enhancing efficiency? To explore this question, this study constructs a specialized corpus based on the *Contract Book of the Civil Code* and conducts a comparative analysis between the authoritative human version and the GPT version. Through both quantitative and qualitative methods, the study examines textual feature differences across three key dimensions: terminological accuracy, syntactic complexity, and the handling of semantically ambiguous expressions. This research aims to elucidate the current limitations of LLMs in legal translation and propose a viable framework for human-machine collaboration, thereby offering practical insights and guidance for producing high-quality legal translations assisted by LLMs.

2. Literature Review

Both domestic and international has explored the application of LLMs to legal text translation. For instance, Zhang (2020) through empirical analysis of legal text machine translation cases, pointed out the core challenges faced by artificial intelligence technology in legal translation, and confirmed the irreplaceability of human translators. Liang (2022) focused on contract text, compared the handling of long sentences by three types of machine translation systems, and found significant limitations in the translation of complex syntactic structures, synonymous and near-synonymous expressions, and specialized legal terminology. Yang (2023) utilized GPT to translate legal texts and compared it with other machine and human translation, and found persistent deficiencies in the handling of specialized legal terminology and grammatical interpretation. In international scholarship, Eva Wiesmann (2019) evaluated the applicability of the computer-assisted translation tool MateCat to Italian legal texts. Her comparative experiments assessed the extent to which machine translation systems can achieve cross-linguistic legal translation in specific document types with minimal post-editing. Ding (2024) conducted a comparative analysis of GPT and four online neural machine translation systems, examining the quality of English-to-Chinese and Chinese-to-English translations of legal texts. The findings underscore the current limitations of LLMs in maintaining accuracy and legal fidelity across languages.

While existing domestic and international studies have explored the application of LLMs to legal texts, relatively few have focused specifically on legislative texts such as the Civil Code. Moreover, most domestic research has concentrated on the limitations of traditional machine translation systems, such as DeepL and Google Translate, with limited attention given to the application of newer LLMs like GPT4o in the context of

legal translation. This study addresses this research gap by focusing on the *Contract Book of the Civil Code*, conducting a systematic comparison between its human version and the GPT version. The analysis centers on three core dimensions: terminological accuracy, syntactic complexity, and the handling of semantically ambiguous expressions. By integrating a mixed research methodology combining quantitative statistics and case analysis, the limitations of traditional automatic Assessment metrics in capturing the subtle requirements of legal translation are overcome, aiming to reveal the existing deficiencies of GPT in legal translation, construct a human-machine collaboration framework, and provide practical guidance and theoretical reference for generating high-quality legal translations with the assistance of LLMs.

3. Research Design

- This study, based on a self-constructed corpus of the *Contract Book of the Civil Code*, incorporating translations generated by GPT, and focuses on examining textual feature differences between the GPT version and the authoritative FaBao human version across three key dimensions: terminological accuracy, syntactic complexity, and the handling of semantically ambiguous expressions. Specifically, the study addresses the following three core questions:
- (1) In translating the *Contract Book of the Civil Code,* how does GPT perform in terms of terminological accuracy compared to the human version?
- (2) Does the GPT version simplify syntactic structures relative to the human version, and if so, how does this simplification impact the legal precision and readability of the provisions?
- (3) What are the differences in processing strategies between the GPT and FaBao versions when handling semantically ambiguous expressions?
- This corpus based on the *Contract Book of the Civil Code*, drawing from both the Chinese and English versions published on the official website of Pkulaw.com (hereinafter referred to as the FaBao Version), comprising three components: the original Chinese text, the authoritative human version, and the GPT version. The human version is an officially sanctioned translation produced by a team of legal experts and senior translators. It has undergone multiple rounds of revision and proofreading. GPT-4o, as the upgraded model released in the previous year, it has significantly improved in natural language understanding and generation capabilities, and the selected corpus serves as a representative sample of high-quality machine and human translations. By conducting a comparative analysis of these texts, this study aims to uncover fundamental differences in translation performance, particularly in terms of terminological accuracy, syntactic complexity, and the handling of semantically ambiguous expressions.
- In terms of research methodology, first, the AntConc 4.3.1 tool is employed, using the stop word list such us *if*, *in* and *is* import function to filter out function words,

and conducting bidirectional retrieval of high-frequency legal terms in both translations (alternately setting target and reference corpora), thereby laying the foundation for comparison of terminological accuracy; second, GitHub open-source tools are utilized to conduct micro-level comparison of syntactic complexity based on key linguistic parameters such as lexical diversity, average sentence length, and ratio of subordinate clauses; finally, Yang' s (2023) classification framework of legal vague expressions is adopted, and typical vague term frequency data are retrieved through AntConc, combined with the CUC_ParaConc tool, to comparatively analyze the handling of semantic vagueness in the Chinese and English texts as well as in both translations.

4. Results and Discussion

4.1. Terminological accuracy

Legal terminology serves as a highly condensed expression of legal concepts and phenomena. (Shi, 2019). To compare the human and GPT versions in terms of terminological accuracy, high-frequency terms retrieved from both English translations using AntConc 4.3.1 are presented in Table 1 and Table 2.

Table 1 GPT4o Translation's High-Frequency Terms (Top 10)

Word	Frequency		Keyness
	GPT Version	FaBao Version	
guarantee	60	1	89.316
custodian	48	0	78.676
goods	71	7	77.430
obligor	47	0	77.036
obligee	40	0	65.556
unclear	38	0	62.277
transport	42	1	60.498
guarantor	39	1	55.725
principal	129	52	54.843
agent	89	26	53.182
debt	38	3	44.296

Table 2 FaBaoTranslation's High-Frequency Terms (Top 10)

Word	Frequency		Keyness
	GPT Version	FaBao Version)	
suretyship	72	0	83.749
depositary	54	0	62.828
mandatary	49	0	57.007
management	99	13	56.078
mandator	47	0	54.679
power	52	1	52.214
limit	72	6	51.297
obligation	182	51	50.468
mandate	42	0	48.859

Word	Frequency		Keyness
	GPT Version	FaBao Version)	
surety	38	0	44.204
suretyship	38	3	48.859

In the GPT translations, *guarantee* and *custodian* appear with significantly higher frequency and prominence. In contrast, the human translations demonstrate a greater prevalence of more legally precise terms such as *suretyship*, *depositary*, and *mandatary*. For instance, with respect to the Chinese term *wei tuo ren,* the GPT version uses custodian 48 times, whereas the human version employs *depositary* 47 times. This comparison highlights the divergence in terminological choices between the two versions. The following is an example.

Original Text: 第八百八十八条 保管合同是保管人.....

Human Translation: Article 888 A deposit contract is a contract whereby the depositary keeps the deposited property delivered...

GPT Translation: Article 888 A custody contract is an agreement wherein the custodian keeps the entrusted property delivered...

The legal term *bao guan ren* is translated in the human version as *depositary,* aligning with established legal usage. The *Yuan Zhao Dictionary of Anglo-American Law* defines *depositary* as "*custodian; bailee,*" highlighting the duty to safeguard the deposited item and return it upon request. This reflects the role of a bailee in a gratuitous bailment under civil law, where safekeeping is voluntary and without remuneration. *Black's Law Dictionary* similarly defines a *depositary* as one who receives property for safekeeping and is bound to return it per the original trust, consistent with the clause requiring the custodian to protect the deposited item. In contrast, the GPT version uses *custodian*, which, though superficially similar, carries a different legal meaning. According to the *Yuan Zhao Dictionary, custodian* refers to a supervisory or managerial role, often in professional contexts such as asset management during insolvency or statutory duties. *Black's Law Dictionary* adds that a custodian is a representative responsible for safeguarding securities and executing financial transactions, implying an institutional or fiduciary obligation. This diverges from the *Civil Code* context, which concerns ordinary, non-professional deposit relationships, such as temporary safekeeping of personal belongings. Thus, custodian fails to capture the voluntary, non-commercial nature of the legal relationship, creating conceptual inconsistency with the legislative intent.

The human version demonstrates greater terminological precision, while the GPT version's deviations risk legal ambiguities, conceptual inconsistency, and potential reader misunderstanding

4.2. Syntactic complexity

In the evaluation of translation quality, the degree of syntactic complexity is of critical importance, as it directly pertains to the readability of the text and the depth of

expression. This metric is employed to assess the diversity and intricacy of syntactic structures in linguistic expression. (Ortega, 2003). The results of data analysis conducted by the author using an open-source GitHub tool on the respective syntactic complexity of the two English translations are presented in Table 3.

Table 3 Comparison of FaBao Version Translation and GPT4o Translation's syntactic complexity

Indicator Name	GPT Version	Fabao Version
Lexical diversity	0.584	0.95
Average T-unit Length	11.2	6.9
dependent clause ratio	2.68	1.07
complex noun phrase Quantity	3.36	1.72
Verb Phrase Quantity/T-unit	2.44	1.53
demonstrative pronoun ratio	24.6%	40.5%

In metrics such as average clause length, complex noun phrases, and dependent clause ratio, the GPT version displays lower syntactic complexity than the human version. The human translation features longer sentences and more complex structures, with higher average clause and T-unit lengths. Its greater use of dependent clauses and complex noun phrases reflects a preference for embedded, intricate constructions. By contrast, the GPT version favors shorter sentences and simpler cohesion. However, in lexical diversity and the use of demonstrative pronouns, the GPT version surpasses the human translation, indicating greater variety and potentially enhanced readability. An example follows.

Original Text: 第五百零九条 当事人.....

Human Translation: Article 509 Each party shall fully perform its own obligations as agreed. The parties shall abide by the principle of good faith, and perform obligations of notification, assistance, confidentiality, etc., in accordance with the nature and purpose of the contract and the usage of trade. In performing a contract, the parties shall avoid wasting resources, polluting the environment, and compromising ecology.

GPT Translation: Article 509 Parties shall fully perform their obligations as agreed. They shall observe good faith and fulfill auxiliary obligations such as notice, assistance, and confidentiality, based on the nature and purpose of the contract and trade customs. Parties shall avoid waste of resources, pollution, and ecological damage during contract performance.

The GPT version introduces *auxiliary obligations,* absent from the original text, while the human version more faithfully renders the phrase as *obligations of notification, assistance, confidentiality*, avoiding unnecessary terminological expansion. In common law, *auxiliary obligations* often refer to subordinate constructs like *guarantees*

or *suretyship,* risking conceptual confusion with the original civil law notion of ancillary duties grounded in good faith. Syntactically, the human version links two predicate *abide by and perform,* within a compound structure that preserves the original' s logical hierarchy. The GPT version, by contrast, fragments these into simple sentences, weakening causal and conditional connections. For example, its treatment of "based on the nature and purpose of the contract" as an isolated prepositional phrase undermines its restrictive function on the scope of obligations. In subject reference, the human version alternates between *each party* and *the parties,* distinguishing individual from collective duties. The GPT version consistently uses *the parties*, which blurs this distinction and weakens the emphasis on individual liability.

Thus, while the GPT translation may improve surface readability through reduced syntactic complexity, it does so at the expense of legal precision, highlighting current LLM limitations in processing texts requiring high logical density and strict definitional accuracy.

4.3. Handling of Semantically Ambiguous Terms

Vague terms refer to expressions that lack precise definitional boundaries or clear referents, and whose scope of application is inherently indeterminate (Du, 2001). As identified by Yang (2023) in the study, four primary categories of vague terms appear frequently in *the Contract Book of the Civil Code*: referential vague terms, degree vague terms, quantity vague terms, and abstract vague terms. To examine how semantically ambiguous expressions are handled in the GPT version versus the human version, this study adopts Yang 's classification framework as its analytical foundation. Representative vague terms from each category were extracted and searched using the corpus analysis tool AntConc 4.3.1, which generated frequency data for the different categories, as presented in Table 4. Subsequently, several high-frequency terms were selected from the search results and retrieved in CUC_ParaConc, where the corresponding Chinese source texts, along with their GPT translations and human equivalents, were aligned and analyzed.

Original Text: 第五百一十八条　债权人为二人以上.....

Human Translation: Article 518 A join... between not less than two creditors...

GPT Translation: Article 518 Where there are multiple obligees...

The Human version translates "……以上" as not less than..., while the GPT version renders it as multiple. In English, multiple generally denotes "more than one" without specifying a lower limit, and under common law, the term may be construed as "three or more". In contrast, the original Chinese expression explicitly includes the base number (meaning "two or more" when the base is two). When a clause applies to exactly two creditors, the GPT version risks misinterpretation as inapplicable, potentially creating legal loopholes or disputes. The Human version adopts the positive-negative translation strategy (Xu & Zhou, 2024), using not less than... to

accurately reflect this statutory meaning of inclusivity, faithfully reproducing the original text's linguistic ambiguity. The analysis above reveals that the human version addresses semantically vague expressions through strategies such as terminological transformation, lexical supplementation, and tense adjustment. In contrast, the GPT version predominantly relies on literal translation, which is more likely to result in omissions or inadequacies in conveying the intended legal meaning.

5. Implications of the Application of LLMs for Legal Translation

Although LLMs such as GPT still exhibit notable limitations and deficiencies in the translation of legal texts, their rapid advancement has nonetheless brought unprecedented convenience to legal translators (Wang , 2025). To enhance the effectiveness of LLMs in supporting high-quality legal translation, this study proposes feasible strategies for human–machine collaboration in three key areas: terminological accuracy, syntactic complexity, and the handling of semantically ambiguous expressions. These strategies aim to offer practical reference and theoretical guidance for improving the quality and reliability of legal translation facilitated by LLMs.

5.1 Layered Parsing and Post-Translation Reconstruction of the Human-Machine Collaborative Optimization Mechanism

In response to the syntactic complexity of legal text, a dedicated syntactic analysis module adapted to legal language has been developed by technical experts, for the precise identification of hierarchical relations such as principal-subordinate structures and nested modifiers in long sentences, and for the mapping of logical relations to generate structured semantic trees. When processing complex syntax, preliminary segmentation is first assisted by large models, providing translation suggestions segmented in accordance with syntactic boundaries, upon which the translator reconstructs the sentence structure to ensure that the translation conforms to the form and stylistic specification of the target legal text.

5.2 Multimodal Knowledge Fusion of Human-Machine Collaborative Optimization Mechanism

In response to semantically ambiguous terms in legal text, their specific Context and corresponding translations are pre-annotated in the professional corpus, to provide the model with refined training samples for enhancing sensitivity to Context.At the same time, the Knowledge Graph technology is applied to associate ambiguous terms with relevant legal concepts, embed them into a broader legal knowledge framework, and integrate Judicial Judgments, Judicial Precedents, and other Unstructured Data to supplement the model’ s understanding of the practical application of terms, thereby overcoming the limitations of single structured data.

Based on enhanced training output, the model generates multiple candidate translations with confidence scores, from which the translator selects the most contextual-

ly appropriate translation according to the Context, thereby achieving an efficient and accurate machine-assisted legal translation process.

6. Conclusion

Based on a self-constructed corpus of the *Civil Code Contract Book* and incorporating translations produced by GPT-4o, this study systematically compares—through both qualitative and quantitative methods—the textual differences between the GPT version and the officially published FaBao Version across three dimensions: terminological accuracy, syntactic complexity, and the handling of semantically ambiguous expressions. The findings are as follows:

(1) Terminological Accuracy: The FaBao Version demonstrates a higher degree of alignment with authoritative legal dictionaries in terms of lexical selection, reflecting greater terminological precision and legal rigor. In contrast, the GPT version exhibits deviations in terminology, which may result in conceptual confusion and ambiguity in legal interpretation.

(2) Syntactic Complexity: While the GPT version enhances readability by simplifying sentence structures, it does so at the cost of reducing the syntactic rigor required in legal drafting. The human version employs more complex and logically structured syntax, thereby preserving the formal and normative characteristics of legal language.

(3) Semantic Ambiguity Handling: In addressing semantically vague expressions, the human version adopts strategies such as terminological transformation, lexical supplementation, and tense modulation. Conversely, the GPT version tends toward literal translation, which increases the risk of omissions or misinterpretations in legal provisions.

This study contributes to the ongoing development of high-quality legal translation assisted by LLMs, offering insights into how their capabilities may be directionally optimized. Simultaneously, it deepens our understanding of the current limitations of LLM-based translation in the legal domain. The findings underscore the necessity of human–machine collaborative workflows, in which translators actively engage with machine outputs to ensure both legal fidelity and linguistic clarity—ultimately laying the groundwork for the sustainable and effective integration of LLMs into legal translation practices.

Acknowledgements

This research was funded by Major Humanities and Social Sciences Research Projects in Zhejiang higher education institutions, Grant Number: 2024QN165. And it's also supported by Special Research Project of the Foreign Languages Association of Zhejiang Province (No. ZWZD2025022).

References

[1] Ding, L. (2024) A Comparative Study on the Quality of English-Chinese Translation of Legal Texts Between ChatGPT and Neural Machine Translation Systems. Theory and Practice in Language Studies, 14(9), 2823-2833.

[2] Du, Jinbang. (2001) From the Vagueness of Legal Language to the Certainty of Judicial Results. Modern Foreign Languages, (03), 306-310+305.

[3] Li Yiyun and Niu Yunping. (2024) The Ideal and Reality of Rebuilding the Tower of Babel —Exploration Based on ChatGPT Translation Cases. Translation Studies and Teaching, (01), 25-31.

[4] Liang Minghui. (2022) A Preliminary Discussion on the Representation Effect of Machine Translation on Long Legal Sentences—Taking English-Chinese Translation of Contract Texts as an Example. Taste & Classics, (15), 38-40.

[5] Ortega, L. (2003) Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. Applied Linguistics, 24(4), 492-518.

[6] Shi Yujiao. (2019) English-Chinese Translation Strategies for Legal Terms. China Science & Technology Translators, 32(01), 36-39.

[7] Wiesmann, E. (2019) Machine Translation in the Field of Law: A Study of the Translation of Italian Legal Texts into German. Comparative Legilinguistics, 37(1), 117-153.

[8] Wang Huashu. (2025) Research on Ethical Risks of Generative AI Translation Technology in Legal Translation. Journal of Gansu University of Political Science and Law, (01), 1-12.

[9] Xu Jun, Zhou Shanshan. (2024) A Corpus-Based Study on the English Translation of Vague Language in the Criminal Law of the People's Republic of China. Studies in Language and Law, (01), 35-52.

[10] Yang Panzhang. (2023) A Study on the Clarification of Vagueness in Legal Language. Ph.D. Thesis, Heilongjiang University, Harbin.