

Routledge
Taylor & Francis Group

Perspectives

Studies in Translation Theory and Practice

ISSN: 0907-676X (Print) 1747-6623 (Online) Journal homepage: www.tandfonline.com/journals/rmps20

Evaluating the translation of deontic and epistemic modal expressions in legal texts: a comparison between human translators and the GPT-4 model

Yanfang Su, Ho Ling Kwok & Kanglong Liu

To cite this article: Yanfang Su, Ho Ling Kwok & Kanglong Liu (12 Jul 2026): Evaluating the translation of deontic and epistemic modal expressions in legal texts: a comparison between human translators and the GPT-4 model, Perspectives, DOI: 10.1080/0907676X.2026.2690390

To link to this article: https://doi.org/10.1080/0907676X.2026.2690390

Published online: 12 Jul 2026.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=rmps20

PERSPECTIVES
https://doi.org/10.1080/0907676X.2026.2690390

Routledge
Taylor & Francis Group

Check for updates

Evaluating the translation of deontic and epistemic modal expressions in legal texts: a comparison between human translators and the GPT-4 model

Yanfang Su [a], Ho Ling Kwok [b] and Kanglong Liu [b]

[a]Department of Translation, Lingnan University, Hong Kong, People's Republic of China; [b]Department of Language Science and Technology, The Hong Kong Polytechnic University, Hong Kong, People's Republic of China

ABSTRACT

The inherent differences between languages and the distinct origins of legal concepts often lead to terminological asymmetries, posing significant challenges for legal translators. The emergence of large language models such as GPT-4 has created new opportunities for legal translation. This study examines how GPT-4 handles modal expressions in legal translation compared to human translators within Hong Kong's bilingual common law system. Modality is a central linguistic feature of legal discourse. It is instrumental in expressing balanced viewpoints, acknowledging legal uncertainty and delineating rights, permissions, obligations and responsibilities. Its translation poses particular challenges due to its polysemous nature and contextual nuances. Using a Chinese-English parallel corpus of 50 Hong Kong judgments with their human and GPT-4 translations, this study analyzed shifts in modal categories, values and orientations during translation. The findings reveal that while GPT-4 maintains basic modal patterns and generally adheres to legal translation norms, its translations exhibit more modality shifts than those produced by human translators. In addition, GPT-4 shows less diversity in its handling of modal expressions. These findings contribute to understanding AI capabilities in legal translation and offer practical insights for optimizing prompt engineering for AI-assisted legal translation while maintaining translation quality in multilingual legal contexts.

ARTICLE HISTORY
Received 30 November 2024
Accepted 11 June 2026

KEYWORDS
Legal translation; modality; GPT-4; bilingual legal system; translation quality

1. Introduction

Legal translation can be defined as a 'communicative act' which operates 'within the mechanism of the law' (Šarcevic, 1997, p. 55; cf. Cheng et al., 2014). It has been recognized as a challenging task due to language-, translation process- and legal system-specific factors (Biel, 2017). The emergence of advanced generative artificial intelligence (GenAI) models has introduced new possibilities for automating this challenging task. Notably, the GPT-4 model has shown significant improvements in context length, creativity, reliability and the ability to handle complex tasks and nuanced instructions

CONTACT Kanglong Liu kl.liu@polyu.edu.hk
© 2026 Informa UK Limited, trading as Taylor & Francis Group

(OpenAI, 2024), demonstrating its potential to translate lengthy legal texts and adapt to diverse requirements in legal translation. Moreover, its top 10% performance ranking on the Uniform Bar Exam further highlights its potential for legal translation (OpenAI, 2024). Despite the numerous studies examining the overall performance of GPT models in legal translation using automatic evaluation metrics, with inconsistent findings (e.g., Ding, 2024; Larroyed, 2023), GenAI's ability to handle nuanced linguistic features, such as modality, has often been overlooked.

Modality refers to the extent of 'commitment' of the speaker to the validity of a statement and represents 'the intermediate degrees' that span the spectrum between positive and negative poles (Halliday, 1994, p. 88). Among the different categories of modality, deontic modality and epistemic modality are most frequently used and play vital roles in legal discourse. Epistemic modality, which concerns the degree of certainty attached to the information, helps judges to provide balanced perspectives, acknowledge reasonable doubts and articulate the validity of the propositions with precision (Koźbiał, 2020, p. 44; Macleod, 2019, p. 473). Deontic modality delineates the rights, permissions, obligations, prohibitions and responsibilities of a legal party (Gong et al., 2020; Lian & Jiang, 2014). The accurate translation of modal expressions is particularly challenging due to their inherent polysemy and context-dependent interpretation. Previous research has documented significant semantic shifts in modal expressions during translation (Biel, 2014; Su et al., 2023), which might influence the argumentative strength of legal judgments and judicial decision-making (Szczyrbak, 2017). This study examines GPT-4's performance compared with that of human translators in handling modal expressions within Hong Kong's bilingual legal system. It aims to deepen our understanding of Large Language Models' (LLMs) capabilities in legal translation and offer practical implications for their integration into multilingual legal workflows.

2. Deontic and epistemic modality in legal discourse

In legal discourse, epistemic and deontic modal expressions are pivotal and extensively used (Cheng & Cheng, 2014; Gong et al., 2020; Su et al., 2023). A substantial body of research has examined the distributional patterns and rhetorical functions of epistemic modality in legal discourse, with a particular focus on judicial texts. For instance, Cheng and Cheng (2014) compared epistemic modality in civil judgments from Hong Kong and Scotland, linking similar patterns to the standards of proof in civil proceedings. Szczyrbak (2017) explored the rhetorical functions of modal adverbs in the opinions of the European Court of Justice, highlighting their role in managing divergent legal interpretations and value-laden assessments. Macleod (2019) found that judges strategically used certainty and uncertainty expressions to maintain professional reputation while advancing truth-seeking objectives. Collectively, these studies reveal that epistemic modality in legal contexts is closely tied to standards of proof, judges' hierarchical roles, confidence in litigants' arguments, the interpretation of laws and judges' caution in safeguarding their reputation (Cheng & Cheng, 2014; Koźbiał, 2020; Macleod, 2019; Szczyrbak, 2017).

Another strand of research focuses on the strategic use of deontic modality in legal discourse, with early studies frequently analyzing the use of 'shall' in legal texts (Garzone, 2013; Williams, 2005). These investigations revealed how the semantic

complexity and ambiguous usage of 'shall' across various jurisdictions often introduce interpretative challenges in legal texts (Garzone, 2013). More recent studies have extended the scope of analysis from 'shall' to other deontic modal expressions. For instance, Jiang (2013, p. 66) found judges favored deontic over epistemic modality to reinforce authority and frequently used projecting clauses to shift modal responsibility to external statutes, portraying themselves as impartial 'spokespersons'. In the context of Hong Kong, Cheng and Sin (2011) pioneered a socio-semiotic approach to examine deontic modality in Chinese judgments, while Li et al. (2016) analyzed its distribution in Chinese legislative discourse in Hong Kong through the lens of modal logic. Contributing to the diachronic perspective, Gong et al. (2020) traced the evolution of deontic modal verbs in Chinese civil and commercial laws between 1949 and 2014. Overall, these studies revealed the crucial functions of deontic modality in conveying the authoritativeness of legal discourse and its role in imposing duties, establishing requirements, conferring rights, granting permissions, or articulating prohibitions (Gong et al., 2020; Lian & Jiang, 2014).

Beyond monolingual analysis, a third line of research focuses on the comparative analysis of modality in different languages. For instance, studies have utilized comparable corpora to examine the use of modality in different legal systems (e.g., Koźbiał, 2020), thus revealing the role of underlying legal philosophies in shaping the distribution patterns of modality. Studies on modality translation highlight several challenges, including the polysemous nature of modal expressions (Li, 2018) and the influence of sociocultural and legal factors on modality usage (Biel, 2014; Su et al., 2023). Despite efforts to harmonize the use of modality in the legal discourse of different languages, inconsistencies persist (Cheng & Sin, 2011).

Given the vital roles of modality and the high-stakes context of legal discourse, precise modality translation is essential. Modality shifts can transform the rhetorical stance of target texts (Szczyrbak, 2017), impacting both the persuasive efficacy of judgments and judicial decision-making. This underscores the necessity for developing robust approaches to improve modality translation in legal contexts.

3. Affordances and limitations of GenAI models in legal translation

Recent advances in GenAI models have introduced new possibilities for translation (Hendy et al., 2023). Studies comparing GenAI with traditional machine translation (MT) systems have found that GenAI often matches or even outperforms commercial MT systems in high-resource languages, e.g., Chinese, German and English, but struggles with low-resource languages, e.g., Hausa and Icelandic (Gao et al., 2023; Jiao et al., 2024). Notably, GenAI excels at producing more readable and coherent translations (Cai, 2024), with translation quality influenced by factors such as text type, model sophistication and prompting strategies (Jiao et al., 2024; Peng et al., 2023).

Compared to other fields of translation, the legal domain has maintained a relatively cautious stance toward GenAI, due to the complexity involved (Cai, 2024), as well as ethical concerns about data confidentiality, training biases and the potentially severe consequences of translation errors (Quinci & Pontrandolfo, 2023). Studies examining GenAI's performance in legal translation have highlighted both the limitations and strengths of GenAI models in handling the unique demands of legal texts. Some

studies reported that AI translation performance in the legal domain ranged from barely reaching the passing threshold to being unsatisfactory, as seen in translations between Arabic and English or French (Ait ElFqih & Monti, 2024; Altakhaineh et al., 2025), English and Italian (Giampieri, 2024) and English and Chinese (Ding, 2024). For example, in translating Arabic legal texts into English and French, GPT-4 and Gemini excelled in lexical and syntactic accuracy but underperformed in legal terminology compared to traditional MT systems (Ait ElFqih & Monti, 2024). This was attributed to differences in legal systems and cultural nuances, which GenAI models struggle to grasp. However, the study's dataset, which comprises isolated sentences with legal terms, raises validity concerns, as the lack of context makes it harder for AI to choose accurate terminology. Altakhaineh et al. (2025) expanded the dataset into a 20,000-word English-Arabic legal corpus, highlighting GPT-4's strength in simpler texts and its difficulties with legal terminology, complex constructs and cultural nuances in Arabic legal texts. Likewise, Iu et al. (2024) found GPT-4's translation style in translating sentences or short paragraphs from English overly literal and unsuitable for Chinese legal texts, often missing nuances, such as misinterpreting modal expressions of discretion as jurisdictional limitations. This is particularly problematic in legal translation, where even minor inaccuracies can have far-reaching legal and reputational consequences. Researchers emphasized the need for post-editing to address specific issues, such as inaccurate modal expressions, pronominal redundancy (Farghal et al., 2025), misuse of archaisms and collocations and misinterpretation of polysemous legal terms (Giampieri, 2024).

In addition to limitations, the potential of LLMs in legal translation emerges in other studies. For example, Larroyed (2023) found that GPT-4 outperformed traditional MT systems in the overall translation quality of patents translated from English into Portuguese, although those produced by traditional systems had marginally better terminology accuracy. Iu et al. (2024) found ChatGPT performed well in maintaining linguistically and culturally appropriate communication styles when translating testimonials from English to Chinese, generating creative solutions for legal terms lacking direct equivalents based on contextual cues and accurately conveying contextual information occasionally misinterpreted by human translators.

Overall, previous studies reveal both the limitations and the potential of GenAI in legal translation. Several research gaps are also identified. First, most studies focus on sentence-level or short-text translations and rely on small-scale corpora, while largely overlooking GenAI's capabilities in handling longer legal texts and complex translation challenges, e.g., managing complex syntactic patterns and logical coherence across sections. Also, much of the existing literature employs basic prompts, lacking the detailed and context-aware prompt engineering needed to improve translation accuracy. Furthermore, evaluations typically focus on overall performance or terminological equivalents, while other critical aspects of legal translation that require nuanced contextual understanding, such as modality, remain relatively underexplored.

4. Objectives and methods of the study

To address these gaps, this study investigates the effectiveness of GPT-4 compared to human translators in handling modal expressions. Legal judgments were selected as the focus of the analysis because of their central role in Hong Kong's bilingual legal

system. As the only common law jurisdiction operating in both English and Chinese, Hong Kong relies on legal judgments as precedents for future cases. Accurate translation of judgments is thus critical for judicial integrity, legal reasoning and the societal impact of judicial decisions (Cham, 2019).

Specifically, the study focuses on the following two research questions:

(1) What are the distributional patterns of epistemic and deontic modal expressions in human and GPT-4 English translations of Hong Kong legal judgments?
(2) What modality shifts occur in human and GPT-4 English translations of Hong Kong legal judgments, and how do they differ between the two translation versions?

4.1. Methods

4.1.1. The corpus

Given the importance of judgment translation in Hong Kong, a parallel corpus was compiled comprising 50 legal judgments originally written in Chinese, their official English human translations and corresponding GPT-4 translations. Both original Chinese judgments and their English human translations were sourced from the Legal Reference System of the Hong Kong Judiciary.[1] It should be noted that only judgments deemed to have 'jurisprudential value' by judges are selected for translation (Cham, 2019, p. 92). These judgments are translated by translators from the Hong Kong Judiciary, who possess the bilingual competence and translation expertise required by the Judiciary. The translations are then reviewed by an authorized reviewer (termed 'vetter') with a strong legal background to ensure linguistic and legal accuracy. After review, the translation is returned to the original translator for verification and correction.

Since the distribution of these judgments was uneven across time and case types, five civil cases and five criminal cases were randomly selected from each five-year period between 1997 and 2021 to ensure a balanced and representative corpus. Notably, the year 1997 is significant because it marks the handover of Hong Kong to China, and the selected period predates the widespread use of GenAI. This stratified random sampling process yielded a total of 50 judgments over 25 years (see Table 1).

4.1.2. Preparation of GPT-4 prompt

In this study, the Azure OpenAI service, which provides API access to the GPT-4 model, was utilized. GPT-4 (model version: turbo-2024-04-09) was selected for the translation of legal judgments due to its popularity, enhanced ability to handle longer contexts, strong performance in comprehending nuanced instructions and comparatively low rate of translation errors at the time of the study (OpenAI, 2024). Following the recommendations of Jiao et al. (2024) and taking into account token limitations, we designed

Table 1. Corpus composition.

	Time period	Number of texts	Mean words	Total words
Original Chinese texts	1997–2021	50	4,543.54	227,177
English human translations	1997–2021	50	4,182.46	209,123
English GPT-4 translations	1997–2021	50	3,654.90	182,745

four zero-shot prompts that defined GPT-4's role, provided contextual information, outlined translation objectives and integrated legal translation principles. The legal translation principles were adapted from Li's (2012, pp. 43–63) general guidelines for Chinese-English legal translation and refined according to recommendations for judgment translation in Hong Kong (Cham, 2019). The four prompts differed in instruction specificity and phrasing. The temperature was set at 0, as legal texts require consistency and ChatGPT tends to perform relatively better with lower temperature settings (Peng et al., 2023).

To identify the prompt yielding the highest translation quality, four legal judgments were randomly selected from the corpus to test the prompts. Prompt 4 consistently outperformed the others across all machine translation evaluation metrics (BLEU, METEOR and COMET), followed by Prompt 3 (see Table 2). However, manual review revealed that Prompt 4 occasionally restructured the source texts and generated hallucinations. In contrast, Prompt 3 performed well across all metrics without such issues. Therefore, Prompt 3 was chosen for this study (see Appendix).

4.1.3. Data coding and analysis

First, a representative list of English and Chinese epistemic and deontic modality markers used in Hong Kong legal judgments was created by reviewing the word frequency and the n-gram lists (including 2- to 4-grams) for both English and Chinese judgments generated by AntConc version 4.1.4 (Anthony, 2022). This identification process was informed by the English modal expression lists proposed by Perkins (1983) and the Chinese modality markers identified by Cheng and Cheng (2014), Li et al. (2016) and Yang (2021). Markers were selected if they ranked in the top 5,000 entries and appeared in at least five texts in the corpus. This approach yielded a list containing 29 potential Chinese markers and 47 English markers.

CUC_ParaConc version 0.3 (Communication University of China, 2013) was then employed to extract concordance lines from the parallel corpus, with Chinese and English markers serving as keywords. A Python program was developed to remove duplicate concordance lines. The parallel retrieval in the two languages ensured that any addition of modal expressions in the target texts could be identified. The final extracted concordance lines consisted of three elements: the original Chinese segments and their corresponding human and GPT-4-generated translations. A subsequent manual review of the concordance lines was also conducted to exclude any markers not functioning as deontic or epistemic modality devices.

After retrieving and reviewing the concordance lines, the modality markers in the original Chinese texts, human translations and GPT-4 translations were annotated for their category (epistemic or deontic), value level and orientation based on the approach to modality analysis developed within systemic functional linguistics (Halliday &

Table 2. Test results of automatic evaluation metrics.

	BLEU	METEOR	COMET
Prompt 1	0.4100	0.6071	0.8519
Prompt 2	0.4100	0.6100	0.8516
Prompt 3	0.4125	0.6100	0.8527
Prompt 4	0.4175	0.6132	0.8531

Matthiessen, 2004). The scales of modality value were categorized into high, median and low (Halliday & Matthiessen, 2004). For epistemic modality, high-value expressions indicate certainty, median-value expressions suggest probability and low-value expressions convey possibility. For deontic modality, high-value expressions denote obligation, prohibition, or necessity, median-value expressions convey expectation and low-value expressions indicate permission or rights (Cheng & Sin, 2011; Li et al., 2016). The value level of each modality marker was determined by consulting Halliday and Matthiessen (2004) and Yang (2021). For markers not addressed in these sources, supplementary references included the Cambridge English-Chinese Traditional Dictionary (n.d.), the Chinese-English Dictionary of Modern Usage (Lin, 1999), the Chinese dictionary *Xiandai Hanyu Babai Ci [Eight hundred words in Modern Chinese]* (Lv, 1987) and the Hong Kong DoJ Glossary of Legal Terms (Hong Kong Department of Justice, 2021).

The English and Chinese linguistic realizations of different modality orientations (Halliday & Matthiessen, 2004; Li, 2018; Yang, 2021) are presented in Table 3. We conducted the initial coding of all concordance lines, which was subsequently reviewed by three trained research assistants. These assistants underwent five rounds of trial coding before the formal annotation process. An inter-rater agreement of 98.1% was achieved. Any discrepancies were resolved through discussion among the research team.

To address the first research question (see Section 4), descriptive statistical measures were first calculated, including raw frequency (F), normalized frequency per 100,000 words (NF) and the relative frequency (%) of each modality category, value and orientation. Then we applied log-likelihood (LL) tests to compare modality distributions in human vs. GPT-4 translations, with significance thresholds at 3.84 ($p < 0.05$), 6.63 ($p < 0.01$), 10.83 ($p < 0.001$), or 15.13 ($p < 0.0001$). A qualitative comparison of high-frequency modal expressions in human and GPT-4 translations was also performed, focusing on markers collectively accounting for 90% of all modality instances. This approach allowed us to analyze preferences and diversity in human and GPT-4 translations.

To address the second research question (see Section 4), modality shifts in human and GPT-4 translations were identified and compared. In this study, changes in modality category, value level, or orientation, as well as additions or omissions of modality in the target texts are classified as 'modality shifts'. Specifically, 'category shifts' refer to cases where a modal expression is preserved in translation but rendered in a different modality category, either from deontic to epistemic, or from epistemic to deontic. A change in

Table 3. Linguistic realizations of different orientations of modal expressions.

Orientation	Realization	English example	Chinese example
Explicit subjective	Clauses with a first-person subject and modal lexical verbs/adjectives	*I am sure that*	我們認為…
	Phrases with first-person pronoun	*In my opinion*	在我看來,… .
Explicit objective	Modal nominalizations	*The only possibility was …*	… 的可能性較低
	Clauses of relational process with modal/evaluative adjective	*It is possible that …*	… 是必要的
	Phrases without a first-person pronoun	*In all probability …*	由此可見
Implicit subjective	Modal verbs	*might, should*	可以, 應
	Co-occurrence of modal verbs and adverbs	*could possibly*	可能會
Implicit objective	Modal adverbs	*perhaps*	必然, 一定
	Be + adjective/past participle	*be required to …*	/

modality value from low to median or high, or from median to high, is coded as 'strengthening', while a shift from median or high to low, or from high to median, is coded as 'weakening'. Orientation shifts include 'implicitation' (explicit to implicit), 'explicitation' (implicit to explicit), 'objectivization' (subjective to objective) and 'subjectivization' (objective to subjective). In our analysis, descriptive statistics of modality shifts, including the raw frequency (F) and relative frequency of shifts in human and GPT-4 translations (i.e., their ratio to total concordance lines; %), will first be presented. Subsequently, qualitative analyses of modality shift examples are conducted to examine how the original modal expressions were rendered in translations.

5. Findings

5.1. Distribution patterns of modality in human and GPT-4 translations

The distribution of modal expressions in human and GPT-4 translations is shown in Table 4. In both translation types, deontic modality is predominant. A key difference lies in the use of epistemic modality, which is more frequent in human (34.79%) than GPT-4 translations (26.99%). In addition, GPT-4 translations are more prone to omitting modal expressions (19.43%) than human translations (9.90%).

An analysis of modal values reveals that high-value modal expressions are prevalent in both translation types, followed by low-value (Table 5). However, no statistically significant differences are found between human and GPT-4 translations in this regard.

Modal orientation exhibits more varied patterns in human vs. GPT-4 translations (Table 6). In both translation types, implicit subjective expressions are the most common, followed by implicit objective expressions. Human translations, however, use significantly more implicit objective forms (133 instances; 10.90%) compared to GPT-4 translations (88 instances; 8.59%). In addition, while explicit expressions are less common overall, human translations employ significantly more explicit subjectivized expressions (61 instances; 5.00%) than GPT-4 translations (6 instances; 0.59%). Similarly, human translations feature more explicit objective expressions, albeit not to a statistically significant degree.

Tables 7–9 present the distribution of high-frequency deontic modal markers across human and GPT-4 translations. Human translators use a wider range of deontic modal expressions compared to GPT-4 translations. For high-value modality (Table 7), human translators frequently employ modal and semi-modal verbs (e.g., 'have to', 'must' and 'shall'), passive constructions (e.g., 'be required') and objective clauses (e.g., 'it is necessary'). In contrast, GPT-4 translations rely heavily on modals and semi-modals, with a marked preference for 'must', followed by 'need' and 'cannot'.

Table 4. Distribution of modal expressions across different categories.

	Human translations			GPT-4 translations			
Category	F	NF	%	F	NF	%	LL
Deontic	749	358.16	55.32%	681	372.65	53.58%	0.56
Epistemic	471	225.23	34.79%	343	187.69	26.99%	6.65**
Omission	134	64.56	9.90%	247	181.13	19.43%	50.99****
Total	1,354	647.94	100.00%	1,271	741.47	100.00%	3.35

**$p < 0.01$; ****$p < 0.0001$.

Table 5. Distribution of modal expressions across different values.

	Human translations			GPT-4 translations			
Value	F	NF	%	F	NF	%	LL
Low	421	201.32	34.51%	356	194.81	34.77%	0.27
Median	348	166.41	28.52%	266	145.56	25.98%	2.72
High	451	215.66	36.97%	402	219.98	39.26%	0.07
Total	1,220	583.39	100.00%	1,024	560.34	100.00%	0.91

As for median-value modal expressions (Table 8), both human- and GPT-4-produced translations consistently prefer the modal verb 'should', in its positive or negative form. Finally, regarding low-value deontic expressions (Table 9), human translations again demonstrate greater diversity than GPT-4 translations.

Compared to deontic modality, epistemic modality exhibits greater linguistic diversity in both human and GPT-4 translations (see Tables 10–12). As shown in Table 10, regarding high-value epistemic modality, neither human nor GPT-4 translations demonstrate a strong preference for specific modal markers. Both show similar tendencies in using modal verbs such as 'must', 'cannot' and 'could not (possibly)'. Additionally, both frequently utilize various modal adjuncts and clauses, e.g., 'it is impossible', 'of course' and 'definitely'.

As for median-value epistemic modal expressions, both human and GPT-4 translations predominantly use the modal verbs 'would', 'will' and their negative forms (Table 11). A key difference is human translators' frequent use of subjective epistemic assessments – as in the expressions 'in our/my view/judgment' and 'we are of the view/take the view' – a feature rarely seen in GPT-4 translations.

For low-value epistemic modality (Table 12), both types of translations exhibit similar patterns, with 'could', 'can', 'might' and 'may' making up about 90% of modal expressions.

5.2. Shifts in modality in human and GPT-4 translations

5.2.1. Shifts in the modal category

Modal expressions generally maintain consistency in modality category across human- and machine-translated versions (cf. Table 13), though human translators achieve a higher consistency (79.32%) compared to GPT-4 (75.85%). Shifts in modality category between source and target texts are rare in both versions (less than 1%).

In Example 1, the Chinese source text uses the modal verb 可以 [could] to indicate the court's authority to exercise certain powers under specific conditions. GPT-4 translates it

Table 6. Distribution of modal expressions across different orientations.

	Human translations			GPT-4 translations			
Orientation	F	NF	%	F	NF	%	LL
Explicit objective	78	37.30	6.39%	52	28.45	5.08%	2.32
Explicit subjective	61	29.17	5.00%	6	3.28	0.59%	45.37****
Implicit objective	133	63.60	10.90%	88	48.15	8.59%	4.16*
Implicit subjective	948	453.32	77.70%	878	480.45	85.74%	1.54
Total	1,220	583.39	100.00%	1,024	560.34	100.00%	0.91

*$p < 0.05$; ****$p < 0.0001$.

Table 7. High-frequency high-value deontic modal expressions.

	Human translations		GPT-4 translations	
Modal expression	F	%	F	%
have to	65	19.40%	13	4.29%
must	65	19.40%	117	38.61%
shall	56	16.72%	17	5.61%
cannot	19	5.67%	46	15.18%
could not	16	4.78%	14	4.62%
be required	12	3.58%	4	1.32%
it is necessary	12	3.58%	7	2.31%
need	10	2.99%	47	15.51%
be to	10	2.99%	8	2.64%
must not	7	2.09%	2	0.66%
shall not	7	2.09%	13	4.29%
may not	6	1.79%	0	0
be not allowed	5	1.49%	0	0
be obliged	4	1.19%	0	0
be not entitled	4	1.19%	0	0
may only	3	0.90%	1	0.33%
might not	3	0.90%	0	0
Total	304	90.75%	289	95.37%

as 'can', maintaining similar connotations of permission. The human translation, however, uses 'will', shifting the meaning to epistemic likelihood.

Example 1: Shift from deontic to epistemic modality

Source text: 但是，只是在申索是清楚及明顯地會敗訴或法庭不應審理案件的情況下，法庭才可以運用該規則的權力，這是確定已久的法律。(CACV354/2001)

Human translation: However, it is well-established that the court *will* exercise the power under that Rule where and only where it is clear and obvious that the claim will fail or the court should not hear the case.

GPT-4 translation: However, the court *can* only exercise the power of this rule when the claim is clearly and obviously destined to fail or the court should not hear the case, which is a well-established legal principle.

A similar number of epistemic modal expressions in the source texts is interpreted as deontic modality in translations, with three instances in human translations and four in GPT-4 translations. In Example 2, the epistemic modal 會 [will] is translated as 'should' and 'shall' in human and GPT-4 translations, respectively, shifting the epistemic meaning into a deontic one. This shift alters the original message from describing the court's likely behavior to prescribing what the court ought to or must do regarding arbitration decisions.

Example 2: Shift from epistemic to deontic modality

Source text: 只有在本條例明文規定的情況下，法院才會干預爭議的仲裁。(CACV7/2006)

Human translation: The Court *should* interfere in the arbitration of a dispute only as expressly provided by this Ordinance.

GPT-4 translation: The court *shall* only intervene in the arbitration of a dispute if expressly provided for by this Ordinance.

Table 8. High-frequency median-value deontic modal expressions.

	Human translations		GPT-4 translations	
Modal expression	F	%	F	%
should	107	68.15%	102	64.15%
should not	47	29.94%	55	34.59%
Total	154	98.09%	157	98.74%

Table 9. High-frequency low-value deontic modal expressions.

	Human translations		GPT-4 translations	
Modal expression	F	%	F	%
could	64	24.90%	62	28.44%
can	48	18.68%	48	22.02%
may	29	11.28%	29	13.30%
be entitled to	21	8.17%	6	2.75%
not need to	14	5.45%	31	14.22%
be allowed	13	5.06%	6	2.75%
it is unnecessary/not necessary	11	4.28%	3	1.38%
not have to	10	3.89%	1	0.46%
be permitted	8	3.11%	7	3.21%
we/I allow	6	2.33%	0	0
it is open to	7	2.72%	0	0
there is no need	3	1.17%	7	3.21%
Total	234	91.04%	200	91.74%

Table 10. High-frequency high-value epistemic modal expressions.

	Human translations		GPT-4 translations	
Modal expression	F	%	F	%
must	13	11.21%	9	9.09%
of course	11	9.48%	5	5.05%
cannot	10	8.62%	11	11.11%
could not possibly	8	6.90%	6	6.06%
could not	8	6.90%	6	6.06%
it is impossible	7	6.03%	7	7.07%
definitely	6	5.17%	6	6.06%
obviously	4	3.45%	0	0
clearly	4	3.45%	14	14.14%
the only/irresistible inference is	4	3.45%	4	4.04%
would definitely	3	2.59%	0	0
it is clear	3	2.59%	1	1.01%
could/can only	3	2.59%	2	2.02%
there is no doubt	2	1.72%	0	0
it is apparent	2	1.72%	0	0
shall	2	1.72%	1	1.01%
it is an undisputed fact	3	2.59%	0	0
certainly	2	1.72%	3	3.03%
be very likely to	2	1.72%	0	0
plainly	2	1.72%	0	0
be impossible	2	1.72%	0	0
it is no proof beyond reasonable doubt	2	1.72%	0	0
it is obvious	2	1.72%	0	0
necessarily	1	0.86%	4	4.04%
the undisputed evidence shows	2	1.72%	3	3.03%
inevitably	0	0	3	3.03%
it is evident	0	0	2	2.02%
naturally	0	0	2	2.02%
be certain	0	0	1	1.01%
Total	108	93.08%	90	90.90%

Table 11. High-frequency median-value epistemic modal expressions.

	Human translations		GPT-4 translations	
Modal expression	F	%	F	%
would	75	39.47%	53	49.53%
will	30	15.79%	13	12.15%
in our/my view/judgment	28	14.74%	0	0
would not	21	11.05%	18	16.82%
we are of the view/take the view	10	5.26%	0	0
will not	9	4.74%	5	4.67%
should	8	4.21%	5	4.67%
be likely	1	0.53%	6	5.61%
Total	182	95.79%	100	93.45%

Table 12. High-frequency low-value epistemic modal expressions.

	Human translations		GPT-4 translations	
Modal expression	F	%	F	%
could	48	29.27%	43	31.16%
can	41	25.00%	41	29.71%
might	35	21.34%	20	14.49%
may	20	12.20%	17	12.32%
may not	4	2.44%	3	2.17%
it can be said/seen	0	0	2	1.45%
Total	148	90.24%	126	91.30%

Regarding the addition of modal expressions, both translation versions favor deontic modality, with human translators introducing more modal expressions overall. In comparison, GPT-4 exhibits more omissions than additions. Another striking difference emerges in the patterns of omissions: human translators omit more deontic modal expressions (108 instances) compared to epistemic modality (26 omissions). GPT-4 translations show comparable deontic modality omission (128 instances) to human translations, while featuring far more omissions of epistemic expressions (119 instances). In Example 3, the Chinese source text features two modal verbs: 可 [can], indicating permission to rectify the defect and 不應該 [should not], expressing that striking out the Statements of Claim is not appropriate. The human translation preserves both modal verbs. The GPT-4 translation renders 可 [can] as 'could' but omits the modal meaning of 不應該 [should not] by restructuring the final clause into a non-finite contrastive construction 'rather than being struck out'. Both versions add a modal expression – 'can' in

Table 13. Shifts in modality category.

	Human translations		GPT-4 translations	
Type of shift	F	%	F	%
From deontic to epistemic	8	0.59%	4	0.31%
From epistemic to deontic	3	0.22%	4	0.31%
Addition of epistemic modality	41	3.03%	13	1.02%
Addition of deontic modality	94	6.94%	39	3.07%
Omission of epistemic modality	26	1.92%	119	9.36%
Omission of deontic modality	108	7.98%	128	10.07%
No shift in category	1,074	79.32%	964	75.85%
Total	1,354	100.00%	1,271	100.00%

human translation and 'should' in GPT-4 translation – to clarify litigation continuation. In Example 4, the human translation retains the median-value epistemic modal 會 [will] as 'will'. In contrast, the GPT-4 translation omits the modal, transforming the probabilistic statement into a factual assertion.

Example 3: Omission/addition of modality

Source text: … 也可修改妥當而繼續訴訟，不應該被剔除。(CACV380/2002)

Human translation: … the defect *can* be rectified and the litigation *can* continue, and the Statements of Claim *should not* be struck out.

GPT-4 translation: … it *could* be amended appropriately and the litigation *should* continue, rather than being struck out.

Example 4: Omission of modality

Source text: … 由於Y患有慢性肺病，預期壽命與一般人口相比會稍為低。(CACV559/2019)

Human translation: … as Y is suffering from chronic lung disease, her life expectancy *will* be shorter than that of the general population.

GPT-4 translation: … due to Y's chronic lung disease, her expected lifespan is slightly lower compared to the general population.

5.2.2. Shifts in modal value

Value shifts are more common than category shifts, as testified by the data in Table 14. Human translators show relatively comparable frequencies of strengthening and weakening modal values, whereas GPT-4 translations lean toward weakening them.

For instance, in Example 5, the source text employs the median-value modal verb 應當 [should] to convey a normative judgment regarding the cancellation of an order. The human translation intensifies the modal value by using 'must'. In contrast, the GPT-4 translation retains the original modal force with 'should'. In Example 6, 不能 [could not] in the source text is translated into the high-value 'could not' in the human translation. Conversely, the GPT-4 translation weakens this modal force by employing 'should not', which suggests a recommendation or advisability rather than prohibition.

Table 14. Shifts in modal value.

	Human translations		GPT-4 translations	
Type of shift	F	%	F	%
Strengthening	36	2.66%	22	1.73%
Weakening	30	2.22%	36	2.83%
Addition of low value	52	3.84%	21	1.65%
Addition of median value	36	2.66%	11	0.87%
Addition of high value	47	3.47%	20	1.57%
Omission of low value	32	2.36%	73	5.74%
Omission of median value	31	2.29%	96	7.55%
Omission of high value	71	5.24%	78	6.14%
No shift in value	1,019	75.26%	914	71.91%
Total	1,354	100.00%	1,271	100.00%

Example 5: Strengthening of modal value

Source text: 他的命令應當予以撤銷。 (CACV100/2016)

Human translation: His order *must* be set aside.

GPT-4 translation: His order *should* be set aside.

Example 6: Weakening of modal value

Source text: 首先兩件案件是在不同日期觸犯, 不能當作一件案件來處理 … (CACC469/1997)

Human translation: First, the two offences in question were committed on different dates and therefore *could not* be treated as one case …

GPT-4 translation: Firstly, the two offenses were committed on different dates and *should not* be treated as one case …

Regarding the addition of modal expressions, both types of translation tend to add low- and high-value modality. However, their omission patterns differ: while both human and GPT-4 translations demonstrate similar frequencies in omitting high-value modal expressions, GPT-4 translations are more likely to omit low- and median-value expressions compared to human translations.

5.2.3. Shifts in modal orientation

Both human and GPT-4 translations show a tendency towards objectivization and explicitation, though to differing extents (Table 15). Explicitation is more frequent in human translations, with 17 shifts to explicitness alone and 33 to explicit-objective forms. This pattern is less frequent in GPT-4 translations, which produced 11 shifts toward explicitness alone and 18 shifts to explicit objective expressions.

Table 15. Shifts in orientation.

	Human translations		GPT-4 translations	
Type of shift	F	%	F	%
Implicitation	7	0.52%	15	1.11%
Explicitation	17	1.25%	11	0.81%
Objectivization	3	0.22%	2	0.15%
Subjectivization	0	0.00%	0	0.00%
Explicitation + Objectivization	33	2.44%	18	1.33%
Explicitation + Subjectivization	1	0.07%	0	0.00%
Implicitation + Subjectivization	2	0.15%	2	0.15%
Implicitation + Objectivization	1	0.07%	2	0.15%
Addition of implicit objective	42	3.10%	21	1.55%
Addition of implicit subjective	91	6.72%	30	2.21%
Addition of explicit objective	1	0.07%	1	0.07%
Addition of explicit subjective	1	0.07%	0	0.00%
Omission of implicit objective	2	0.15%	13	0.96%
Omission of implicit subjective	129	9.52%	182	13.43%
Omission of explicit objective	1	0.07%	2	0.15%
Omission of explicit subjective	2	0.15%	50	3.69%
No shift	1,021	75.35%	922	68.04%
Total	1,354	100.00%	1,271	100.00%

Consider, for instance, Example 7, in which both human and GPT-4 translations shift the implicit modal expression 很明顯 [quite obviously] into explicit modal assessment using impersonal structures.

Example 7: Shift from implicit objective to explicit objective orientation

Source text: 很明顯他的理由完全是基於他對'住宅租賃'（domestic tenancy）及該條例第115A條的理解。(CACV754/2000)

Human translation: *It is obvious* that these reasons rested entirely upon his understanding of 'domestic tenancy' and section 115A of the Ordinance.

GPT-4 translation: *It was clear* that his reasons were entirely based on his understanding of 'domestic tenancy' and section 115A of the Ordinance.

Modal additions in translation show distinct patterns: while explicit modal additions are rare, implicit modality is frequently added in both translation versions. Yet, human translations show markedly higher frequencies in adding both implicit subjective (91 cases) and objective (42 cases) expressions than GPT-4 translations (30 and 21 cases, respectively).

As for the omission of modal expressions, both human and GPT-4 translations mainly omit implicit subjective expressions. Human translators seldom omit other orientations, whereas GPT-4 translations fail to retain 50 explicit subjective and 13 implicit objective expressions. The frequent omissions of explicit subjective modal expressions are closely tied to the translation of 本庭 [this court], a formal self-reference in Chinese legal texts that is more ceremonial than 'we' but more subjective than 'this court'. In Example 8, the Chinese text uses 本庭認為 [this court believes] to explicitly convey the judges' conviction regarding Magistrate Leung's mistakes. The human translation preserves this subjective assessment using 'in our view'. In contrast, GPT-4 shifts to the more objective expression 'this court'.

Example 8: Omission of explicit subjective expression

Source text: 本庭認為梁裁判官判處答辯人罰款是在量刑原則上犯了錯誤。(CACV1712/2001)

Human translation: *In our view*, the magistrate erred in principle when he imposed a fine on the respondent.

GPT-4 translation: *This Court* believes that Magistrate Leung erred in sentencing principles by imposing a fine on the Respondent.

6. Discussion

6.1. Similarities between human and GPT-4 translations

Similar modality translation patterns are found in human and GPT-4 translations, primarily reflecting the influence of the source text (Su et al., 2023). Both sets of translations demonstrate the predominance of deontic over epistemic modal expressions, reflecting the essential function of deontic modality in conveying legal obligations, prohibitions and authorizations in judicial discourse (Gong et al., 2020). This distribution pattern

aligns with previous studies (Cheng & Cheng, 2014; Jiang, 2013; Su et al., 2023). In addition, the prevalence of high-value modal expressions in both human and GPT-4 translations reflects judges' strong conviction in establishing reliable precedents for future proceedings (Hinkle et al., 2012). As for the orientation of modality, implicit expressions dominate in both sets of translations, corroborating previous studies on English and Chinese modality in legal texts (Gong et al., 2020; Li et al., 2016; Su et al., 2023). In cases of explicit modality, both human and GPT-4 translations favor objectivized expressions, reinforcing judicial impartiality (Dudek & Stępień, 2021). The relatively infrequent use of first-person pronouns or self-reference reflects the judicial practice of attributing decisional authority to institutions rather than individual judges (Dudek & Stępień, 2021, p. 105).

Regarding modality shifts, both human and AI translation approaches exhibit greater accuracy in preserving modal categories, value levels and orientations compared to other contexts, such as conference interpreting (Li, 2018) and translation of government documents (Yu & Wang, 2023). This aligns with the emphasis on 'consistency and accuracy' in legal judgment translation (Poon, 2006, p. 553). The competitive performance of GPT-4 in modality translation stems from its advanced contextual understanding ability (Hendy et al., 2023) and effective domain-specific prompting strategies (Peng et al., 2023).

6.2. Differences between human and GPT-4 translations

Nonetheless, human and GPT-4 translations also exhibit notable differences, especially in modal value and orientation shifts. Firstly, human translators produced fewer modality shifts compared to GPT-4, which indicates that LLMs still struggle with capturing subtle contextual modal meanings. This aligns with Farghal et al. (2025), who observed that while LLMs generally capture Arabic modal meanings, they occasionally shift orientations, potentially affecting faithfulness to the source. This gap likely stems from LLMs' reliance on pattern recognition and probabilistic word predictions, leading them to prioritize probable interpretations over semantic precision (Dalayli, 2023). The inherent translation ambiguities in the Chinese-English language pair (Wen & van Heuven, 2017), such as the polysemous nature of modality, the availability of multiple synonymous translation options and the inherent differences in the Chinese and English modal systems, may further exacerbate the challenges for GPT-4.

Secondly, GPT-4 translations exhibit less diversity in modal expressions than human translators. This difference may be attributed to the fact that human-translated judgments are often completed by multiple translators, whose individual characteristics, such as language proficiency and personal style (Wen & van Heuven, 2017), contribute to a greater variability in the translation of modality. Additionally, the divergence may stem from differences in the translation processes. Human translation involves complex cognitive mechanisms, like source text comprehension, mental representation and cross-linguistic decision-making (Dimitrova, 2010), whereas GPT-4 relies on training data and probabilistic word predictions. The zero-temperature setting used in this study has likely reinforced dominant patterns from its training data, further reducing variation in modal expressions.

Thirdly, the two sets of translations differ in terms of added or omitted modality. The frequent additions of modal expressions in human translations reflect the translators' mediating role (Fu & Chen, 2019) in clarifying implicit legal meanings while balancing clarity and fidelity to the source text. This tendency also arises from cultural differences (Yu & Wang, 2023): high-context Chinese communication relies on implicit understanding and shared cultural knowledge, while low-context English favors explicitness (Hall, 1976). As a result, human translators often make implicit Chinese modal expressions more explicit in English to conform to target-language norms. In contrast, the frequent omission of modality in GPT-4 translations appears to stem from the model's tendency toward condensed expressions, particularly when handling parallel structures in the source text, which may reduce the structural positions available for modal markers (see Example 3). Moneus and Sahari (2024) also found that ChatGPT might render modalized statements as unmodalized statements in translation. As Iu et al. (2024) reported, ChatGPT performed well in producing fluent translations but occasionally failed to capture the subtle modal meanings conveyed in the source texts. While this emphasis on fluency can enhance overall readability, it often compromises fidelity to the source text and may inadvertently alter the legal force or interpretive flexibility of a statement.

Another trend differentiating the approaches to translation concerns the shifts in value and orientation. Human translators show no marked preference for strengthening or weakening modal value, which might reflect a deliberate decision to maintain a balance between reinforcing authoritativeness (Li, 2018) and avoiding over-commitment to preserve interpretive flexibility for legal professionals (Biel, 2014). In comparison, the weakening trend observed in GPT-4 translations is consistent with recent findings from Jiang and Hyland (2026) that GPT-4-generated argumentative texts tend to soften assertions and acknowledge potential counterarguments, rather than emphasizing conviction and certainty. Moreover, translators frequently shift toward explicit objective expressions, a pattern that better aligns with the conventional modality features of non-translated English legal judgments (Su et al., 2023). Considering that Chinese legal texts primarily rely on modal verbs for modal assessment (Lian & Jiang, 2014), such conscious adaptation by human translators reflects their attention to achieving 'textual fit' with English legal conventions (Biel, 2014). On the other hand, GPT-4 shifts less frequently toward explicit expressions, likely reflecting its tendency to preserve the source text's clausal structure rather than reformulating the text to fit target-language conventions. While this pattern could be partly attributed to the prompt used in this study, which asked the model to 'adhere closely to the language style in the source text', the fact that modal shifts still occur in GPT-4 translations suggests that prompt design alone cannot fully account for the model's translational behavior. A complementary explanation, following Jiang (2025), may lie in the absence of cognitive constraints in LLMs. Unlike human translators, who both attend to target-text conventions and reorganize complex source-text structures to ease cognitive load, GPT-4 is 'unburdened' by such pressures and therefore lacks intrinsic motivation to globally restructure its output (Jiang, 2025, p. 14). Together, these factors help explain the closer mirroring of source-text patterns observed in GPT-4 translations and its more limited flexibility in adapting to target-language norms.

6.3. Implications for legal translation

Based on these findings, several implications for legal translation emerge. First, given GenAI's ability to maintain basic modal patterns and generally adhere to legal translation norms, a two-tier translation process can be pursued to enhance translation efficiency. In this process, GenAI would generate initial drafts, which human translators would then review with particular attention to preserving modal expressions during structural adjustments and calibrating modal values appropriately.

This study also offers valuable insights for optimizing ChatGPT prompts in legal translation. Given the identified limitations in GPT-4's performance, strategic prompt engineering, including temperature adjustments and specific guidelines, is necessary. In particular, the guidelines should address modality-specific challenges identified in this study. In light of our finding that GPT-4 translations tend to weaken the modal force of the source text, effective prompts should instruct the model to preserve source-text modality during structural adjustments and to calibrate modal values carefully. Few-shot prompting can also be leveraged by including representative examples of error-prone modal translations identified in the study, to help the model recognize and avoid such patterns in its output. Beyond the modality-specific concerns, the guidelines should also consider general requirements such as the balance between translation accuracy and acceptability, while considering purpose, context and intended audience.

7. Conclusion

This study compared Chinese-to-English human and GPT-4 translations of modal expressions in legal texts. Both human and GPT-4 translations share similar patterns in aspects shaped by the legal context, including the predominance of deontic over epistemic modality, the prevalence of implicit expressions, the preference for objectivized forms in explicit modality and a relatively high accuracy in preserving modal categories, values and orientations. This study also identified several differences between human and GPT-4 translations. GPT-4 translations generally align with legal translation norms but lack the semantic precision and adaptability of human translators. Human translations exhibit no marked preference for adding or omitting modal expressions, or for strengthening or weakening value levels, or shift orientations toward explicit objectivity, whereas GPT-4 uses less diverse modal expressions, omits more expressions and often weakens modal values. These findings underscore both the potential and limitations of GPT-4 in legal translation workflows, suggesting the value of GPT-4 as a supportive tool while reaffirming the indispensable role of human expertise.

The limitations of this study also need to be acknowledged. First, the bias in the model's training data and human translators' preferences could have influenced the results. Being publicly available online, the English translation of the selected source text may be part of GPT-4's training data. Nonetheless, since the model generates translations based on general patterns learned from extensive training data, any potential overlap between the analyzed texts and the training data would likely have a negligible impact on translation outcomes. Second, this study focused on deontic and epistemic modal expressions but excluded other types. Future research could expand the analysis to a broader range of modal expressions to provide more comprehensive insights into

AI translation performance. Third, this study did not undertake an in-depth exploration of the functional roles of modality within legal texts. Future research may consider drawing on perspectives from semiotics and functional logic to better understand how shifts in modality during LLM-based translation could influence legal interpretations, especially with regard to pragmatic and contextual dimensions. Despite these limitations, this study contributes to understanding GenAI capabilities in legal translation and offers insights into developing practical guidelines for incorporating GenAI into legal translation workflows.

Note

1. https://legalref.judiciary.hk/lrs/common/ju/judgment.jsp, last accessed on October 13, 2023.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study was supported by the General Research Fund of the Research Grants Council of Hong Kong [Project No. P0051774].

Notes on contributors

Yanfang Su is an Assistant Professor at Lingnan University. Her research interests include corpus-based legal translation studies, computer-assisted language learning and the use of LLMs in translation and translation pedagogy.

Ho Ling Kwok is currently a PhD Student at The Hong Kong Polytechnic University. Her research interests include corpus-based translation studies and the application of LLMs in translation.

Kanglong Liu is an Associate Professor at The Hong Kong Polytechnic University. He specialises in corpus-based translation studies and his main interests include empirical approaches to translation studies, corpus-based translation research and translation pedagogy. He is currently the Associate Editor of Translation Quarterly, the official publication of the Hong Kong Translation Society.

ORCID

Yanfang Su http://orcid.org/0000-0002-4526-4187
Ho Ling Kwok http://orcid.org/0000-0002-9701-1181
Kanglong Liu http://orcid.org/0000-0003-3962-2563

References

Ait ElFqih, K., & Monti, J. (2024). Large language models as legal translators of Arabic legislation: Do ChatGPT and Gemini care for context and terminology? In N. Habash, H. Bouamor, R. Eskander, N. Tomeh, I. Abu Farha, A. Abdelali, S. Touileb, I. Hamed, Y. Onaizan, B. Alhafni, W. Antoun, S. Khalifa, H. Haddad, I. Zitouni, B. AlKhamissi, R. Almatham, & K. Mrini (Eds.), *Proceedings of the Second Arabic Natural Language Processing Conference* (pp. 111–122). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.arabicnlp-1.10

Altakhaineh, A. R. M., Alghathian, G. A., & Jarrah, M. M. (2025). A comparative study of accuracy in human vs. AI translation of legal documents into Arabic. *International Journal of Language & Law (JLL)*, *14*, 63–80. https://doi.org/10.14762/jll.2025.063

Anthony, L. (2022). *Antconc (version 4.1.4) [Computer Software]*. Waseda University.

Biel, Ł. (2014). The textual fit of translated EU law: A corpus-based study of deontic modality. *The Translator*, *20*(3), 332–355. https://doi.org/10.1080/13556509.2014.909675

Biel, Ł. (2017). Researching legal translation: A multi-perspective and mixed-method framework for legal translation. *Revista de Llengua i Dret*, *68*, 76–88. https://doi.org/10.2436/rld.i68.2017.2967

Cai, L. (2024). How does ChatGPT compare with conventional neural machine translation systems in performing a Chinese to English translation task? *Journal of Translation Studies*, *4*(1), 25–45. https://doi.org/10.3726/JTS012024.02

Cambridge English–Chinese traditional dictionary. (n.d.). https://dictionary.cambridge.org/

Cham, E. S. K. (2019). Discussion on the skills and methods of translating judgments. In E. S. K. Cham, & K. H. Lee (Eds.), *Bilingual legal system in Hong Kong: Language and translation* (pp. 89–110). Hong Kong University Press.

Cheng, L., & Sin, K. K. (2011). Courtroom language and discourse. *International Journal of Law, Language & Discourse*, *1*(1), 1–26.

Cheng, L., Sin, K. K., & Cheng, W. (2014). Legal translation: A sociosemiotic approach. *Semiotica*, *2014*(201), 17–33. https://doi.org/10.1515/sem-2014-0019

Cheng, W., & Cheng, L. (2014). Epistemic modality in court judgments: A corpus-driven comparison of civil cases in Hong Kong and Scotland. *English for Specific Purposes*, *33*(1), 15–26. https://doi.org/10.1016/j.esp.2013.07.006

Communication University of China. (2013). *CUC_ParaConc (Version 0.3)* [Computer Software]. Communication University of China. https://ling.cuc.edu.cn/tools/f8b7291025863577

Dalayli, F. (2023). Use of NLP techniques in translation by ChatGPT: Case study. In *Proceedings of the Workshop on Computational Terminology in NLP and Translation Studies (ConTeNTS) Incorporating the 16th Workshop on Building and Using Comparable Corpora (BUCC)* (pp. 19–25).

Dimitrova, B. E. (2010). Translation process. In Y. Gambier, & L. van Doorslaer (Eds.), *Handbook of translation studies: Volume 1* (pp. 406–411). John Benjamins.

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. https://doi.org/10.17507/tpls.1409.18

Dudek, M., & Stępień, M. (2021). *Courtroom power distance dynamics*. Springer International Publishing.

Farghal, T., Shraideh, K., & Al-Omari, A. M. (2025). Evaluating free legal translation tools between Arabic and English: A comparative study of Google Translate, ChatGPT, and Gemini. *International Journal for the Semiotics of Law – Revue Internationale de Sémiotique Juridique*, *38*(6), 2083–2110. https://doi.org/10.1007/s11196-025-10282-z

Fu, R., & Chen, J. (2019). Negotiating interpersonal relations in Chinese-English diplomatic interpreting. *Interpreting. International Journal of Research and Practice in Interpreting*, *21*(1), 12–35. https://doi.org/10.1075/intp.00018.fu

Gao, Y., Wang, R., & Hou, F. (2023). How to design translation prompts for ChatGPT: An empirical study. https://doi.org/10.48550/arXiv.2304.02182

Garzone, G. (2013). Variation in the use of modality in legislative texts: Focus on shall. *Journal of Pragmatics*, *57*, 68–81. https://doi.org/10.1016/j.pragma.2013.07.008

Giampieri, P. (2024). AI and the BoLC: Streamlining legal translation. *Comparative Legilinguistics*, *58*, 67–90. https://doi.org/10.14746/cl.2024.58.1

Gong, M., Cheng, W., & Cheng, L. (2020). Development of deontic modality in Chinese civil laws. *Pragmatics and Society*, *11*(3), 337–362. https://doi.org/10.1075/ps.16058.gon

Hall, E. T. (1976). *Beyond culture*. Knopf Doubleday Publishing Group.

Halliday, M. A. K. (1994). *An introduction to functional grammar (2nd edition)*. Hodder Education Publishers.

Halliday, M. A. K., & Matthiessen, C. M. (2004). *An introduction to functional grammar (3rd edition)*. Routledge.
Hendy, A., Abdelrehim, M., Sharaf, A., Raunak, V., Gabr, M., Matsushita, H., Kim, Y. J., Afify, M., & Awadalla, H. H. (2023). How good are GPT models at machine translation? A comprehensive evaluation. https://doi.org/10.48550/arXiv.2302.09210
Hinkle, R. K., Martin, A. D., Shaub, J. D., & Tiller, E. H. (2012). A positive theory and empirical analysis of strategic word choice in district court opinions. *Journal of Legal Analysis*, *4*(2), 407–444. https://doi.org/10.1093/jla/las014
Hong Kong Department of Justice. (2021). *Combined DoJ glossaries of legal terms [Online glossary]*. https://www.glossary.doj.gov.hk/index.php
Iu, K. I., Tu, L., & Liu, Y. (2024). AI-Powered legal translation: Can ChatGPT facilitate the development of Chinese common law? https://doi.org/10.2139/ssrn.4920531
Jiang, F., & Hyland, K. (2026). Taking a stance: AI generated versus student written argumentative essays. *Applied Linguistics Review*. Advance online publication. https://doi.org/10.1515/applirev-2025-0188
Jiang, T. (2013). Modality and identity: An analysis of the judges' discourse from the perspective of systemic functional linguistics. *Yuyan Wenzi Yingyong*, (2), 63–71.
Jiang, Z. (2025). Does LLM translation align with translation universals? A cross-genre simplification study on English-Chinese translation based on dependency grammar. *PLoS One*, *20*(6), e0324830. https://doi.org/10.1371/journal.pone.0324830
Jiao, W., Wang, W., Huang, J. T., Wang, X., Shi, S., & Tu, Z. (2024). Is ChatGPT a good translator? Yes with GPT-4 as the engine. https://doi.org/10.48550/arXiv.2301.08745
Koźbiał, D. (2020). Epistemic modality: A corpus-based analysis of epistemic markers in EU and Polish judgments. *Comparative Legilinguistics*, *41*(1), 39–70. https://doi.org/10.14746/cl.2020.41.3
Larroyed, A. (2023). Redefining patent translation: The influence of ChatGPT and the urgency to align patent language regimes in Europe with progress in translation technology. *GRUR International*, *72*(11), 1009–1017. https://doi.org/10.1093/grurint/ikad099
Li, J., Cheng, L., & Cheng, W. (2016). Deontic meaning making in legislative discourse. *Semiotica*, *2016*(209), 323–340. https://doi.org/10.1515/sem-2016-0002
Li, K. (2012). *Legal translation: A theoretical and practical approach*. Peking University Press.
Li, X. (2018). *The reconstruction of modality in Chinese-English government press conference interpreting: A corpus-based study*. Springer.
Lian, Z., & Jiang, T. (2014). A study of modality system in Chinese-English legal translation from the perspective of SFG. *Theory and Practice in Language Studies*, *4*(3), 497–503. https://doi.org/10.4304/tpls.4.3.497-503
Lin, Y. (1999). *Lin Yutang's Chinese-English dictionary of modern usage [Online dictionary]*. Chinese University of Hong Kong. https://humanum.arts.cuhk.edu.hk/Lexis/Lindict/
Lv, S. (1987). *Xiandai hanyu babai ci (3rd edition)*. The Commercial Press.
Macleod, J. A. (2019). Reporting certainty. *Brigham Young University Law Review*, *2019*(2), 473–532.
Moneus, A. M., & Sahari, Y. (2024). Artificial intelligence and human translation: A contrastive study based on legal texts. *Heliyon*, *10*(6), e28106. https://doi.org/10.1016/j.heliyon.2024.e28106
OpenAI. (2024). GPT-4 technical report. https://arxiv.org/html/2303.08774v6
Peng, K., Ding, L., Zhong, Q., Shen, L., Liu, X., Zhang, M., Ouyang, Y., & Tao, D. (2023). Towards making the most of ChatGPT for machine translation. https://doi.org/10.48550/arXiv.2303.13780
Perkins, M. R. (1983). *Modal expressions in English*. Frances Pinter Publishers.
Poon, W. E. (2006). The translation of judgments. *Meta*, *51*(3), 551–569. https://doi.org/10.7202/013559ar
Quinci, C., & Pontrandolfo, G. (2023). Testing neural machine translation against different levels of specialisation. An exploratory investigation across legal genres and languages. *Trans-kom*, *16*(1), 174–209.
Šarcevic, S. (1997). *New approach to legal translation*. Kluwer Law International BV.

Su, Y., Liu, K., & Cheung, A. K. (2023). Epistemic modality in translated and non-translated English court judgments of Hong Kong: A corpus-based study. *The Journal of Specialised Translation*, *40*, 56–80. https://doi.org/10.26034/cm.jostrans.2023.525

Szczyrbak, M. (2017). Modal adverbs of certainty in EU legal discourse: A parallel corpus approach. In K. Aijmers, & D. Lewis (Eds.), *Contrastive analysis of discourse-pragmatic aspects of linguistic genres* (pp. 91–115). Springer.

Wen, Y., & van Heuven, W. J. (2017). Chinese translation norms for 1,429 English words. *Behavior Research Methods*, *49*(3), 1006–1019. https://doi.org/10.3758/s13428-016-0761-x

Williams, C. (2005). Vagueness in legal texts: Is there a future for shall? In V. K. Bhatia, J. Engberg, M. Gotti, & D. Heller (Eds.), *Vagueness in normative texts* (pp. 201–224). Peter Lang.

Yang, S. (2021). *A systemic functional study of modality in modern Chinese*. Springer.

Yu, H., & Wang, N. (2023). What happened to the modal expressions? Modality in translating Chinese texts into English. *Lingua*, *294*, Article 103606. https://doi.org/10.1016/j.lingua.2023.103606

Appendix

Prompt used for translating legal judgments in this study.

Context information

You are a professional legal translator working on the translation of judgments in Hong Kong, where common law jurisdiction applies. The judgments originally written in traditional Chinese are extracted from the High Court and District Court in Hong Kong.

Judgment translation principles

(1) Accuracy and precision: use accurate and precise legal concepts and terms that are defined in existing laws or relevant legal documents of Hong Kong.
(2) Consistency and identity: maintain consistent and identical usage of terms throughout the translation, including necessary repetition of sentence structures.
(3) Clarity and concision: ensure clear translations and prioritize clarity over conciseness when conflicts arise.
(4) Professionalism: apply specialized legal knowledge effectively in the translation.
(5) Style of language: adhere closely to the language style in the source texts during the translation process and give due consideration to the disparities between the standardized written language and colloquial Cantonese as presented in the original materials.

Please consider the contextual information, adhere to the judgment translation principles and translate the following judgment from traditional Chinese into English: {source_text}