

法律文本人工智能翻译质量评估与优化路径研究

——基于术语库与译后编辑的实证分析

王奕茗

江苏警官学院, 江苏 南京 210031

DOI:10.61369/SE.2026050002

摘 要 : 基于 140 条法律文本语料, 对比 ChatGP T 与有道翻译 AI 的译文与官方译本, 构建术语错误、语义偏差、句法问题、表达不规范四维误差分类体系进行量化评估。针对原始翻译中句法混乱与术语不一致等高频问题, 引入术语库嵌入与译后编辑优化策略。实验表明, 优化后整体错误率从 97% 降至 6%, 错误率降低 94%。人机协同模式可显著提升法律文本翻译质量, 为涉外法治语言服务提供实践路径。

关 键 词 : 人工智能翻译; 法律条文; 误差分析; 术语库; 译后编辑; 人机协同

Research on Quality Evaluation and Optimization Path of AI Translation for Legal Texts— An Empirical Analysis Based on Term Banks and Post-Editing

Wang Yiming

Jiangsu Police College, Nanjing , Jiangsu 210031

Abstract : Based on 140 legal text corpora, this study compares the translations generated by ChatGPT and Youdao Translation AI with official translations. A four-dimensional error classification system is constructed for quantitative evaluation, encompassing terminology errors, semantic deviations, syntactic issues, and non-standard expressions. To address high-frequency problems such as syntactic confusion and inconsistent terminology in original translations, strategies including terminology database embedding and post-editing optimization are introduced. Experimental results show that the overall error rate decreases from 97% to 6% after optimization, a reduction of 94%. The human-machine collaborative model can significantly enhance the quality of legal text translation, providing practical pathways for international legal.

Keywords : AI-powered translation; legal provisions; error analysis; terminology database; post-editing; human-machine collaboration

引言

近年来, 神经机器翻译快速发展, 以 ChatGPT 为代表的生成式人工智能在法律领域的应用日益增多, 但法律文本术语固定、句式复杂, 人工智能翻译存在术语不一致、语义偏差、句法混乱、表达不规范等问题。本研究选取三部法律共 140 条法条, 对比 ChatGPT 和有道翻译 AI 译文与官方译本, 构建误差分类体系进行量化分析, 并引入“术语库嵌入 + 译后编辑”优化策略, 验证人机协同模式的有效性, 为法律文本智能翻译提供实践参考。

一、法律文本人工智能翻译的误差类型与分布

(一) 整体错误率与错误类型分布

本研究对 140 条法律文本分别使用 ChatGPT 和有道翻译 AI 进行逐句翻译, 并与官方译本对照, 依据术语错误、语义偏差、句法问题、表达不规范四类误差进行标注。统计结果显示, ChatGPT 原始译文总错误数为 136 处, 整体错误率为 97%; 其中

句法问题最多 (85 处, 占 63%), 表达不规范次之 (27 处, 占 20%), 术语错误 22 处 (占 16%), 语义偏差仅 2 处 (占 1%)。有道翻译 AI 原始译文总错误数为 91 处, 整体错误率为 65%; 其中表达不规范占比最高 (56 处, 占 62%), 句法问题 11 处 (占 12%), 术语错误 14 处 (占 15%), 语义偏差 10 处 (占 11%)。数据表明, ChatGPT 在句法结构处理上劣势明显, 而有道翻译 AI 在表达规范性上问题突出; 整体上, 有道翻译 AI 的准确性与流畅

性略优于 ChatGPT。

（二）术语错误典型表现

术语错误是法律翻译中最具代表性的问题，主要表现为同一术语在不同句子中出现不同译法，以及未采用标准法律表达。例如，ChatGPT 将“有组织犯罪集团”简化为“结构性团体”，漏译“依照本公约确立的犯罪”核心要件，导致犯罪构成要件缺失；将“拘留审查”拆分为“detention and examination”，破坏术语整体性，混淆行政强制措施与刑事侦查行为。有道翻译 AI 则将“offences”译为“违法行为”而非“犯罪”，降低了法律严重程度；将“knowledge”译为“知识”而非“明知”，忽略刑法中的主观要件；将“expert”译为“专家”而非“鉴定人”，混淆诉讼参与人角色。此外，ChatGPT 在翻译“应当”时时而用“shall”时而用“should”，缺乏一致性。术语不统一或不规范会直接导致法律定义失真，影响司法合作中的准确理解，甚至可能引发国际法律争议。

（三）语义偏差与句法问题分析

语义偏差主要表现为信息缺失、逻辑关系错误或含义偏离。ChatGPT 将“in a manner consistent with”误译为时间状语“在履行……时应遵循”，将条件关系转化为时间关系，逻辑关系不准确；将“可以不说理由”译为“没有义务说明理由”，将授权性规范误译为义务性规范的否定，未能体现主动拒绝的法律效果。有道翻译 AI 将“disposition”译为“去向”缩小词义范围，遗漏“处分、处置”之意；将“被引渡人”误译为“the person sought for extradition”，混淆引渡前“被请求引渡人”与引渡后“被引渡人”两种法律身份，可能引发管辖权误判。句法问题在法律长句中尤为突出。ChatGPT 无法识别 where 从句的修饰范围，导致跨国要件被割裂于两项犯罪之外；套用 where 条件从句造成主句缺失，句式残缺。有道翻译 AI 同样出现缺少主句谓语的动词的不完整句，以及时间状语嵌套混乱、头重脚轻等结构问题。

二、表达不规范及原始翻译质量对比

（一）表达不规范问题分析

法律文本要求正式、严谨的语言风格，而人工智能译文常出现口语化表达或遗漏法律特有结构。ChatGPT 未使用规范情态动词 shall，例如将“本公约的目的是……”译为“The purpose is……”，缺失国际公约应有的立法严谨性；将“适用本法”译为“This Law applies to……”，遗漏 shall 结构；将“可以不说理由”译为“may not give reasons”，存在歧义且过于简略。有道翻译 AI 将“purpose”译为“目的”而非“宗旨”，将“社会公共利益”译为“the public interests of society”造成冗余，将“may not”译为“可以不”与“不得”的歧义表达。这些不规范现象削弱了法律文本的权威性和庄重感。

（二）ChatGPT 与有道翻译 AI 的对比

两种翻译工具在错误类型上呈现明显差异。ChatGPT 在句法问题上错误率高达 63%，说明其处理复杂法律长句的能力较弱，容易产生结构混乱、逻辑断层；但在表达规范性和术语准确性方

面相对较好，语义偏差仅 1%。有道翻译 AI 在句法层面表现良好（错误率仅 12%），能够较为准确地处理主谓关系和从句嵌套；但其表达不规范问题占 62%，且在术语准确性和语义理解上错误更多。总体而言，ChatGPT 在法律条文的表达风格上更接近法律英语，有道翻译 AI 则在句子通顺度上占优，但两者均无法独立完成高质量法律翻译。

（三）原始翻译质量评估

采用 BLEU 思想进行自动评估并辅以人工标注，原始翻译阶段两种工具均未达到专业法律翻译的可接受标准。ChatGPT 的错误率高达 97%，意味着平均每句话几乎都存在至少一处错误，其中大部分为句法问题，需要大量人工重写。有道翻译 AI 错误率 65%，虽然相对较低，但表达不规范问题突出，术语和语义错误也较多。从实际应用角度看，完全依赖人工智能进行法律文本翻译风险较高，尤其是在引渡、司法协助等涉外法治场景中，小小的翻译偏差可能导致法律后果的误判。因此，必须引入优化策略。

三、术语库嵌入与译后编辑的优化路径

（一）优化方案设计与实施

针对原始翻译中的问题，本研究提出“术语库嵌入 + 译后编辑”的优化策略。首先自建法律术语库，从语料中提取 50 条高频法律术语及其官方标准译文，例如“出境入境—exit and entry”“引渡请求—request for extradition”“强制措施—compulsory measures”等，以 Excel 表格存储。术语库还纳入了“有组织犯罪集团—organized criminal group”“司法协助—mutual legal assistance”“临时引渡—provisional extradition”等核心表达，确保覆盖原文主要法律概念。译后编辑阶段，人工依据术语库统一替换不准确或不一致的术语，避免同一概念出现多种译法；系统调整长句语序、修复主谓关系和从句结构，解决机器翻译中常见的逻辑断层问题；将口语化或非正式表达改为法律文体，如补入情态动词 shall、修正歧义用词（将“may not”改为“is not required to”）。优化过程中重点针对高频错误类型，并保留原始译文与优化译文的对照记录，以便后续追溯和效果评估。

（二）优化后错误率变化

经过译后编辑，ChatGPT 译文的总错误数从 136 处降至 8 处，整体错误率由 97% 降至 6%，错误率降低 94%。具体来看：术语错误从 22 处减至 8 处，下降 64%；语义偏差从 2 处降至 0，下降 100%；句法问题从 85 处降至 0，下降 100%；表达不规范从 27 处降至 0，下降 100%。这表明 ChatGPT 的句法缺陷完全可以通过人工重构予以修复。有道翻译 AI 译文的总错误数从 91 处降至 31 处，整体错误率由 65% 降至 22%，降低 66%。其中：术语错误从 14 处减至 13 处，仅下降 7%；语义偏差从 10 处降至 5 处，下降 50%；句法问题从 11 处降至 10 处，下降 9%；表达不规范从 56 处降至 3 处，下降 95%。数据表明，译后编辑对 ChatGPT 的提升效果更为显著，而对有道翻译 AI 的术语和句法问题改善有限，说明后者在术语一致性方面更加顽固，其原始译文中的错误往往根植

于更深层的语料偏差，单纯的人工替换难以彻底根除。

（三）优化效果分析

实验结果显示，术语库的引入显著提升了术语一致性，减少了重复误译问题，尤其是对于高频固定表达如“应当—shall”“不得—may not”“依照—in accordance with”等起到了强制规范作用。译后编辑则在修正语义偏差、重构句法结构、规范法律语言风格方面发挥了关键作用，使原本松散或歧义的表达转变为严谨的法律措辞。值得注意的是，表达不规范是两种工具优化前后改善最明显的类别（ChatGPT 下降100%，有道翻译 AI 下降95%），说明人工干预对于法律文体的把握具有不可替代性，机器难以自主习得“宗旨”与“目的”、“明知”与“知识”等细微但关键的文体差异。然而，有道翻译 AI 优化后仍存在13处术语错误和10处句法问题，提示部分错误的修正需要更深层的法律知识介入，简单的术语替换和结构调整不足以完全解决，例如涉及不同法系概念对译时需进行解释性翻译。总体而言，“术语库+译后编辑”能够大幅提升译文质量，使人工智能翻译从“不可用”或“低质”状态提升至接近可用水平，但距离完全替代人工仍有差距。

四、人机协同翻译模式的有效性与局限性

（一）人机协同模式的有效性分析

人机协同模式并非简单的“机器翻译+人工校对”，而是基于分工优化的协作机制。机器负责初稿生成，将重复性劳动自动化，大幅提高效率；人工译者专注于术语统一、逻辑调整、风格规范等关键问题；术语库则作为中介机制，为双方提供标准化依据。本研究验证了该模式在法律文本中的适用性：优化后整体错误率可降至6%~22%，同时翻译时间远低于纯人工翻译。尤其在涉外执法、国际司法协作等需要快速处理多语种法律文本的场景中，人机协同能够在保证质量的前提下满足时效要求。核心逻辑可以概括为：机器负责效率，人工负责质量，术语库负责规范。

（二）人工智能在法律翻译中的局限性

尽管优化后效果显著，但人工智能在法律翻译中的固有局限性仍然存在。第一，缺乏对法律语境的深层理解能力。机器翻译主要基于统计规律和语料学习，难以把握法律概念背后的制度背景与逻辑关联。第二，处理文化负载词能力弱。不同法域之间存在制度差异，一些概念在目标语言中没有完全对应的表达，需要人工进行解释性翻译而非简单对等转换。第三，术语一致性不稳定。即使在同一文本中，同一术语也可能出现不同译法，法律文本对此零容忍。第四，翻译结果具有“不可解释性”，难以追溯其生成逻辑，在法律领域这一特点尤为敏感，因为翻译结果需要具备可追溯性与可解释性。

（三）研究局限与未来启示

本研究存在一定局限。首先，样本规模有限（仅140句），虽能反映一定规律，但在更大规模文本中的适用性有待验证。其次，实验工具仅选取ChatGPT和有道翻译AI两种，代表性有限。再次，误差标注过程存在主观性，尽管构建了分类体系，但仍依赖研究者的判断。最后，本文主要从语言层面评估质量，未涉及译文在真实司法场景中的可接受性。未来研究应扩大语料规模并引入更多类型的法律文本；结合领域自适应技术，通过法律语料微调模型；探索将知识图谱引入翻译系统，增强概念关系理解；并开展实际应用层验证，如涉外执法中的表现评估。

五、结语

实验证明，人工智能翻译在法律文本中效率优势显著，但术语严谨性与句法规范性存在明显短板。“术语库嵌入+译后编辑”的优化策略使整体错误率从97%降至6%，验证了人机协同模式的有效性。该模式以机器负责效率、人工负责质量、术语库负责规范，能够在保证翻译速度的同时显著提升译文专业水准。未来应扩大语料类型，引入知识图谱与领域自适应技术，推动法律翻译向智能化、精准化方向发展。

参考文献

- [1] 张舟, 刘文清. 人工智能背景下法律推理模型建立探究 [J]. 南方论刊, 2024, (01): 70-72.
- [2] 周俊琛. 政策工具视角下我国人工智能高质量发展的法律政策文本分析——基于我国86份政策文本的量化分析 [J]. 三门峡职业技术学院学报, 2025, 24(01): 82-90.
- [3] 陈太龙. 人工智能背景下侦查翻译活动的困境与对策研究 [J]. 产业与科技论坛, 2025, 24(06): 247-250.
- [4] 俞璐璐. 人工智能知识产权侵权案件起诉状英汉翻译实践报告 [D]. 广东外语外贸大学, 2025.
- [5] 常亮, 王成山, 田海松. 人工智能赋能机器翻译背景下职教科翻译专业的挑战与发展方向 [J]. 河北石油职业技术大学学报, 2025, 27(06): 1-5.