

ChatGPT与DeepSeek在经贸文本汉英翻译中的质量对比研究

赵欢焕*, 李华东

上海海事大学外国语学院, 上海

收稿日期: 2025年12月8日; 录用日期: 2025年12月31日; 发布日期: 2026年1月12日

摘要

随着大语言模型的迅速发展, 人工智能在专业文本翻译中的应用日益受到关注, 但其在经贸类文本翻译中的实际表现仍缺乏系统研究。本文选取一篇经贸类专业文本作为研究语料, 基于BLEU值、TER值、Flesch-Kincaid Grade Level与Gunning Fog Index等指标, 比较分析了ChatGPT-4o与DeepSeek-R1两款模型在经贸文本汉英翻译中的表现差异。研究发现, DeepSeek-R1在准确性和术语还原等方面表现更优, 更贴近标准译文风格。研究建议未来可采用“AI初译 + 人工审校”的人机协作模式, 以提升经贸文本翻译的质量与可控性, 也为AI系统在高语用敏感度语境中的翻译应用提供实践参考。

关键词

人工智能, ChatGPT-4o, DeepSeek-R1, 经贸文本, 翻译质量评估

A Comparative Study on the Chinese-English Translation Quality of Economic and Trade Texts Using ChatGPT and DeepSeek

Huanhuan Zhao*, Huadong Li

College of Foreign Languages, Shanghai Maritime University, Shanghai

Received: December 8, 2025; accepted: December 31, 2025; published: January 12, 2026

Abstract

The rapid development of large language models has sparked growing interest in applying artificial

*通讯作者。

文章引用: 赵欢焕, 李华东. ChatGPT 与 DeepSeek 在经贸文本汉英翻译中的质量对比研究[J]. 现代语言学, 2026, 14(1): 445-453. DOI: 10.12677/ml.2026.141058

intelligence to specialized text translation. However, their performance in translation of economic and trade text remains underexplored. This study compares the translation performance of ChatGPT-4o and DeepSeek-R1 using a professional economic and trade text as the corpus. Based on BLEU, TER, Flesch-Kincaid Grade Level, and Gunning Fog Index metrics, the results show that DeepSeek-R1 delivers greater accuracy and more consistent terminology reproduction, yielding translations that more closely align with the reference translation. The study proposes a human-AI collaborative model, wherein AI-generated drafts are refined through human post-editing, to enhance both the quality and controllability of economic and trade text translation. The findings also offer practical insights into deploying AI in pragmatically sensitive contexts.

Keywords

Artificial Intelligence, ChatGPT-4o, DeepSeek-R1, Economic and Trade Texts, Translation Quality Assessment

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

近年来,在人工智能技术特别是大语言模型快速发展的推动下,大语言模型翻译领域取得了显著进展。传统的基于规则或统计的翻译方法正逐渐被深度神经网络驱动的神神经大语言模型翻译系统所替代。新一代大语言模型在语境理解、句法生成和多语种处理方面的表现日益成熟,为实现高效率、高质量的自动翻译提供了新的可能性。而经贸文本具有标准化程度高、指向性明确的特征,对译者在语言表达与语用适切性方面提出了较高要求。尽管当前大语言模型在语言理解和生成方面取得了较大进步,但其优势更多体现于生成表达规范、结构连贯的内容,对于复杂语用现象的识别和理解仍然面临着诸多挑战[1][2]。在此背景下,对大语言模型在经贸文本翻译中的表现进行系统评估具有重要意义。

2024年5月,Open AI发布的ChatGPT-4o在原生多模态处理、响应速度以及语言覆盖方面均实现重大升级,这表明语言模型进入更高的发展阶段。紧随其后,中国本土大模型代表DeepSeek-R1于2025年1月推出,在长文本理解、多语种支持及成本优化方面表现突出,并在国际人工智能评测中取得优异成绩。这些技术进展显著增强了大语言模型在多语种文本生成与风格控制中的能力,尤其在经贸、法律等专业语境下的翻译应用潜力日益受到关注。

为了系统比较分析ChatGPT-4o与DeepSeek-R1两款代表性大语言模型在处理经贸类文本汉英翻译任务中的表现差异,本研究采用BLEU与TER等自动评估指标来分析译文与标准版本在词汇和结构上的差异,并结合Flesch-Kincaid Grade Level与Gunning Fog Index等可读性指标,从多个维度评估两款模型译文的准确性与可读性,同时探讨两款模型在文本处理和语言生成上的优势与不足。

2. 文献综述

2.1. 大语言模型翻译

随着大语言模型的快速发展,大语言模型翻译技术已成为翻译学界与语言服务产业的研究热点。大语言模型在人们学习、生活和工作中拥有非常广阔的前景,可以直接应用于数据驱动的翻译研究。然而,

大语言模型在给翻译实践和翻译研究带来巨大历史机遇的同时,也带来重大挑战[3]。其中一个重要挑战体现在语用处理方面。虽然新一代大语言模型在语用能力上有所提升,但在处理复杂语用现象时仍显不足[4]。例如,虽然大语言模型在翻译中有一定优势,但其生成的译文在话语取向一致性的控制上存在风险,可能导致误导性信息的传播[5],面临着文化差异挑战等问题[6]。此外,自然语言极为复杂,提高自然语言处理水平,让计算机真正理解自然语言,仍是极为困难的任务[7]。尽管人工智能技术在提高翻译效率和降低成本方面具有优势,但也不可避免地存在准确性不足等局限[8]。

由于技术发展的局限性,单纯的人工智能语言服务还不能满足市场对翻译质量的要求,“人工+AI”的混合语言服务模式成为语言服务的主要模式[9]。大语言模型翻译的高效可以运用在信息型文本或初步翻译当中,人类译员转向译后编辑加工,从而实现翻译质量的提升[10]。另外,专业术语输入、跨文化差异以及特定语境共情等问题仍然需要人类译员,特别在处理专业度高的文本时,人工翻译的角色依然关键[6]。因此,人机协作是翻译行业发展的必然趋势[11]。

2.2. 经贸类文本翻译

经贸文本的翻译研究一直是学界的重点方向,它不仅是语言转换的过程,更是跨文化交流和形象呈现的重要方式。此类文本在我国的对外传播活动中占有重要位置,是国际受众了解我国立场与实践的重要渠道。对经贸文本的翻译开展系统研究,有助于提升跨文化传播的效果,促进不同文化之间的理解与沟通[12]。这类文本不仅要求译文语言的准确性和规范性,更强调译者对语用环境与文化差异等的适切把握。

有研究主张在翻译该类文本时采用“内容异化、形式归化”的策略,以避免中式英语对目标语读者理解的障碍,提升译文的地道性与接受度[13]。也有学者进一步强调,这类文本的翻译应遵循释疑解惑、对接目标语文本修辞、彰显文化自信等原则,通过语义明晰化、增补文化背景、调整句法结构等手段,提升译文的可理解性与接受度[14]。

在大语言模型参与经贸文本翻译的背景下,有学者倡导将大语言模型作为外宣翻译的辅助手段,以缓解人力不足与时间压力[15]。但同时也应认识到其在意识形态控制与语境重构方面的局限性,因而需由人工译员进行最终把关与校准。有学者指出,该类语篇的翻译实践应注重信息的构建与议程设置,译者必须结合目标受众的文化背景和认知习惯,灵活调整翻译策略[16]。当前我国在对外传播的组织与协调方面具备一定基础,但在实践能力、传播能力等方面仍有提升空间[17]。

2.3. 翻译质量评估

翻译质量评估长期以来是翻译研究的核心议题之一。随着神经网络大语言模型翻译技术的广泛应用,客观、标准化的自动化评估方法成为研究主流。量化的译文相似度标准与高质量的人工参考译文是大语言模型翻译自动评分系统的两个基本要素[18]。

在大语言模型翻译研究中,BLEU (Bilingual Evaluation Understudy)和 TER (Translation Edit Rate)是当前两个主流的自动化评估指标。BLEU 通过计算机器译文与人工参考译文之间的 n-gram 重合率来衡量翻译准确性,具有计算简便、通用性强等特点[18]。TER 是对 BLEU 的有效补充,衡量将大语言模型翻译结果修改为与人工参考译文完全一致所需的最少编辑操作数[19]。

此外,文本可读性指标在经贸类文本大语言模型翻译后评估中也具有重要意义。Flesch-Kincaid Grade Level 和 Gunning Fog Index 是两种经典的英语可读性指标[20][21]。前者根据句子长度与单词音节数来估算读者理解该文本所需的教育年级水平;后者主要从平均句长和多音节词等方面来计算文本的阅读难度,其得分反映了理解该文本所需的最低教育年级。

3. 研究方法

3.1. 研究问题

本研究旨在讨论以下三个问题：第一，ChatGPT-4o 与 DeepSeek-R1 作为大语言模型翻译工具，在处理经贸类汉英翻译任务中，其译文质量表现如何？第二，这两款模型是否有潜力替代人工译者在经贸文本翻译中的功能？第三，在人工智能持续发展的背景下，经贸类文本翻译的未来发展趋势将如何演变？

3.2. 研究语料与对象

本文选取 2025 年 4 月公开发布的经贸类文本《关于中美经贸关系若干问题的中方立场》作为研究语料。该文本结构严谨、术语密集、语体正式，适合作为评估大语言模型翻译系统在专业经贸语篇汉英翻译中表现的研究对象。本研究使用的 ChatGPT-4o 为 OpenAI 发布的 4o 模型版本，DeepSeek-R1 为 DeepSeek 发布的 R1 版本。

本研究包括三类英文译文：其一为该经贸文本随附的参考译文；其二与其三分别为通过 ChatGPT-4o 与 DeepSeek-R1 在无上下文提示条件下，对中文原文逐段生成的译文，用于与参考译文进行质量差异分析。为确保数据分析的透明性与可比性，本研究对原文及三类英文译文进行了量化统计(见表 1)。所有文本均与中文原文的 221 个段落逐一对应，保证语篇结构一致，便于后续指标计算。中文语料以汉字计数，英文语料以单词(tokens)计数。

Table 1. Overview of the research corpus

表 1. 研究语料基本信息

| 文本类型 | 段落数 | 汉字/单词数(形符数) | 文件格式 |
|------------------|-----|-------------|------|
| 中文原文 | 221 | 26,809 | .txt |
| 标准英文译文 | 221 | 17,455 | .txt |
| ChatGPT-4o 英文文本 | 221 | 16,445 | .txt |
| DeepSeek-R1 英文文本 | 221 | 15,899 | .txt |

3.3. 数据处理与分析

数据采集与处理依托 Python 3.9 环境开展，具体流程如下：

首先，将中文原文输入 ChatGPT-4o 与 DeepSeek-R1，分别生成对应英文译文，并将所有译文整理为.txt 文件，每段内容为一行，与中文原文段落严格对应。接着，采用 sacrebleu 模块分别计算两个模型译文的 BLEU 值与 TER 值，以评估词汇匹配程度与编辑距离，反映翻译的准确性。同时，使用 matplotlib 绘制箱型图，呈现模型在各项指标中的得分分布与波动情况。为进一步分析译文的语言复杂度与可读性，研究采用 textstat 模块，计算三类译文的 Flesch-Kincaid Grade Level (FKGL)与 Gunning Fog Index (GFI)得分。上述操作均通过 Python 脚本批量执行，实现数据的自动处理与可视化呈现。

本研究采用 BLEU 与 TER 用于衡量词汇与结构层面的译文准确性，BLEU 的 n-gram 精度(P1-P4)与惩罚因子(Brevity Penalty)进一步辅助评估整体翻译表现；FKGL 与 GFI 指标则从语言复杂度与可读性角度补充测评体系，形成多维交叉分析框架。

4. 实验结果与分析

4.1. BLEU 值和 TER 值

在 Python 3.9 的环境中运行相关脚本可得到表 2。

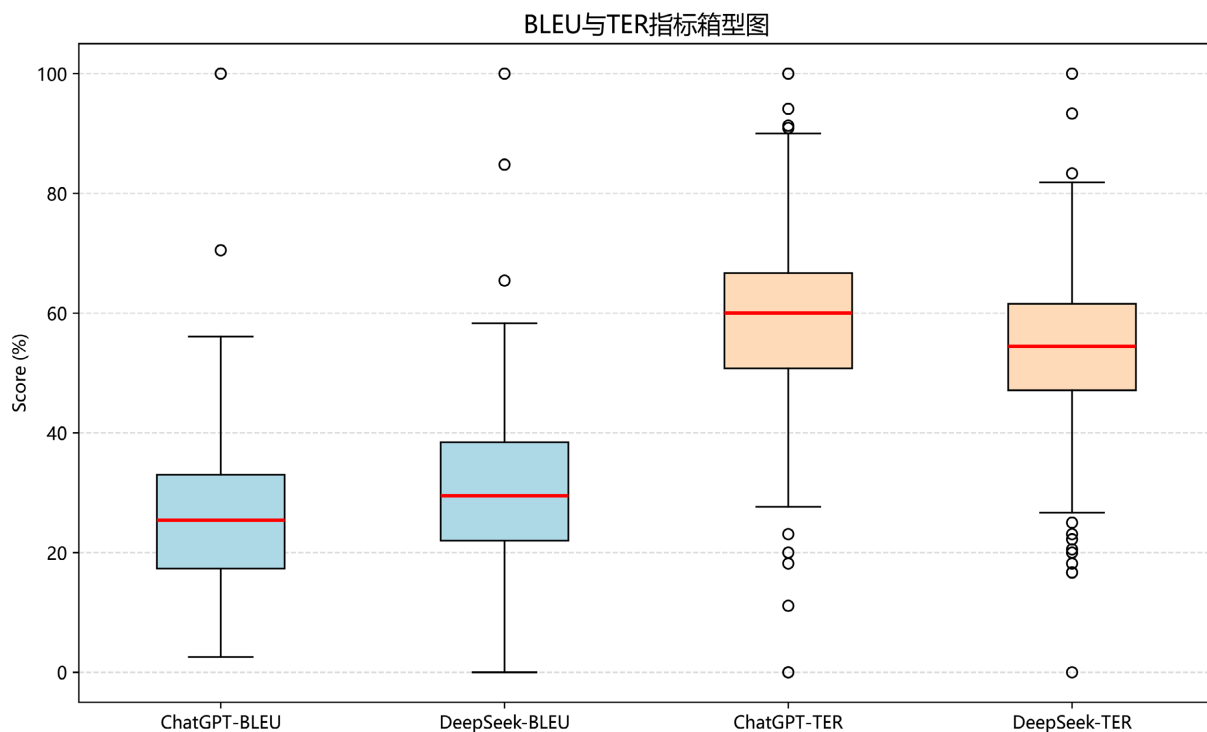
Table 2. BLEU&TER values of the translated texts of ChatGPT-4o and DeepSeek-R1**表 2.** ChatGPT-4o 和 DeepSeek-R1 译文 BLEU&TER 得分

| 翻译工具 | P1 | P2 | P3 | P4 | BP | BLEU | TER |
|-------------|------|------|------|------|-------|--------|--------|
| ChatGPT-4o | 0.62 | 0.36 | 0.23 | 0.15 | 0.995 | 0.2954 | 0.5911 |
| DeepSeek-R1 | 0.68 | 0.41 | 0.27 | 0.19 | 0.950 | 0.3289 | 0.5411 |

BLEU 值越高, 表示译文与参考译文越接近, 翻译质量越好。一个翻译工具生成译文的 BLEU 值达到 0.314, 就表明该译文质量良好, 达到了大语言模型翻译的基本要求[22]。由表 1 可知, ChatGPT-4o 的 BLEU 值为 0.2954, 略低于参考标准, 这表明 ChatGPT-4o 在充分表达原意和语言自然流畅方面都还有很大的提升空间; 而 DeepSeek-R1 的得分为 0.3289, 已达到较为理想的水平, 显示其在译文准确性和语言自然度方面更为优越。进一步观察 n-gram 精度, DeepSeek-R1 在 P1 至 P4 的匹配率上均优于 ChatGPT-4o, 说明其在词汇命中率及高阶短语组合方面具备更高稳定性。在 BP (Brevity Penalty) 方面, ChatGPT-4o 的值为 0.995, 更接近 1, 说明其译文长度与参考译文更为接近, 而 DeepSeek-R1 的 BP 略低, 但未明显影响 BLEU 总得分, 反映出其压缩表达策略较为合理。

在 TER 方面, ChatGPT-4o 的得分为 0.5911, 高于 DeepSeek-R1 的 0.5411。TER 值越低表示译文修改成本越小, 与人工参考文本更接近。因此, DeepSeek-R1 在结构还原、语序控制与术语匹配方面表现更佳, 整体译文质量更高。

为进一步考察两款模型在翻译稳定性与一致性方面的表现, 本文基于 221 段译文分别计算段落级 BLEU 值与 TER 值, 并绘制箱型图以呈现分布情况, 如图 1 所示。

**Figure 1.** Boxplots of BLEU and TER scores**图 1.** BLEU 与 TER 指标箱型图

图中显示, DeepSeek-R1 在 BLEU 维度上得分中位数更高、分布更集中, 低分波动较少, 表明其译

文在词汇与句法层面更贴近参考译文, 整体稳定性和准确性更优。在 TER 维度上, DeepSeek-R1 的中位数低于 ChatGPT-4o, 且得分更为集中, 表明其译文在整体上更接近参考译文, 所需后期编辑操作更少。相较之下, ChatGPT-4o 在部分段落上出现较高的 TER 值, 反映出其译文在局部结构与语义还原方面的不稳定性。

综上, DeepSeek-R1 在 BLEU 与 TER 两个维度上均优于 ChatGPT-4o, 表现出更高的语言一致性与结构相似度, 体现出其在处理经贸这类高复杂度文本时的适应性与可靠性。

4.2. Flesch-Kincaid Grade Level 和 Gunning Fog Index 指标

Flesch-Kincaid Grade Level 用于评估文本对读者的教育水平要求, 得分越高, 表示文本越复杂, 适合受教育程度较高的读者阅读。Gunning Fog Index 是衡量文本复杂度的重要指标, 一篇文章的得分越低, 读者就越容易读懂。标准译文、ChatGPT-4o 和 DeepSeek-R1 译文的 Flesch-Kincaid Grade、Gunning Fog Index 得分如表 3 所示。

Table 3. Values of Flesch-Kincaid Grade Level and Gunning Fog Index

表 3. Flesch-Kincaid Grade Level、Gunning Fog Index 值

| 指标 | 标准译文 | ChatGPT-4o 译文 | DeepSeek-R1 译文 |
|----------------------------|-------|---------------|----------------|
| Flesch-Kincaid Grade Level | 15.96 | 14.10 | 15.70 |
| Gunning Fog Index | 19.15 | 17.31 | 19.06 |

从整体趋势看, 标准译文的 FKGL (15.96)与 GFI (19.15)得分最高, 反映出其语句结构复杂、术语密集, 符合经贸文本高严肃性、高专业度的语言特点, 适合具备较高教育背景的读者。ChatGPT-4o 的译文在两项指标上得分最低, 其中 FKGL 为 14.10, GFI 为 17.31, 说明其在一定程度上对原文进行了简化, 提升了文本的可读性, 但也可能意味着译文在部分句法转换或逻辑承接上未能完整保留经贸文本应有的庄重与正式风格。相比之下, DeepSeek-R1 在 FKGL 与 GFI 上的得分更接近标准译文, 尤其 GFI 仅相差 0.09, 表明其译文在句法结构、专业词汇和语言风格上更倾向于保留经贸文本本身的专业性与正式性, 更贴合人工翻译的表达习惯。

综上, 两个模型在可读性上均呈现出一定程度的语言简化, 但表现路径有所不同。ChatGPT-4o 倾向于对原文进行通俗化处理, 强调表达清晰与易读性; 而 DeepSeek-R1 则更注重保留原文语言特征, 在语言复杂度与信息密度之间取得相对平衡, 更接近标准译文风格。

4.3. 典型翻译实例的质性对比分析

为进一步解释定量指标所反映的模型差异, 有必要结合具体翻译现象进行说明。整体来看, 在涉及政策影响或制度性后果的表述中, ChatGPT-4o 更倾向于使用 “impact” “cause harm” 等概括性表达, 弱化了原文通过程度副词与评价性动词所构建的语义张力; 相比之下, DeepSeek-R1 更注重保留原文的逻辑递进关系与语体正式性, 使译文在信息密度与功能指向上更接近参考译文。这一差异在多个语段中反复出现, 也与前文 BLEU、TER 及可读性指标的整体趋势相互印证。表 4 展示了一个典型经贸文本翻译实例的中英对照情况。

该段文本通过并列结构呈现贸易限制措施对全球经济秩序的多重负面影响。标准译文使用 “severely hinders” “undermines the international economic and trade order” 等表达, 较好保留了原文的语义强度与经贸类文本的严谨性。相比之下, 两款模型的译文虽在信息层面基本完整, 但均不同程度采用较为简化和通用的措辞, 如 “severely undermining” “causing significant harm”, 在提升可读性的同时削弱了原文的

正式性与语体张力。其中, ChatGPT-4o 的通俗化倾向更为明显, DeepSeek-R1 虽整体更为紧凑, 但在语义强度上仍略逊于标准译文。

Table 4. A comparative example of translation for an economic and trade text
表 4. 典型经贸文本翻译实例对照

| 文本类型 | 内容 |
|----------------|--|
| 中文原文 | 美方实施的一系列贸易和投资限制措施不仅增加企业合规成本, 严重阻碍两国正常经贸合作, 还影响全球产业链和供应链的稳定, 严重破坏国际经贸秩序。 |
| 标准译文 | The series of trade and investment restrictions implemented by the US not only increases compliance costs for enterprises and severely hinders normal China-US economic and trade cooperation, but also affects the stability of global industrial and supply chains and seriously undermines the international economic and trade order. |
| ChatGPT-4o 译文 | The series of trade and investment restrictions imposed by the U.S. side have not only increased compliance costs for enterprises and seriously hindered normal bilateral economic and trade cooperation, but also impacted the stability of global industrial and supply chains, severely undermining the international economic and trade order. |
| DeepSeek-R1 译文 | The series of trade and investment restrictions imposed by the U.S. not only increase corporate compliance costs and severely hinder normal bilateral economic and trade cooperation but also disrupt the stability of global industrial and supply chains, causing significant harm to the international economic and trade order. |

从目的论(Skopos Theory)视角看, 经贸类文本翻译的核心目标在于服务制度性信息传递与政策功能实现, 而非单纯追求易读性。该实例表明, 当前大语言模型在强调语言简洁的同时, 仍需进一步协调语体强度与功能忠实之间的关系。本研究中, DeepSeek-R1 的整体表现更接近标准译文, 其翻译策略更符合经贸文本以“功能忠实”为导向的翻译目的。

5. 讨论

5.1. 大语言模型翻译的未来发展路径探析

总体而言, 近几年 ChatGPT、文心一言等生成式人工智能语言大模型, 经过海量语料的预训练, 具备了更强大的语义理解和语言生成能力, 不断刷新人工智能在自然语言处理领域的应用边界[23]。研究结果表明, ChatGPT-4o 与 DeepSeek-R1 在经贸文本翻译中已具备较强的语言生成能力, 能够在整体结构与信息层面较好还原原文内容, 但两者呈现出不同倾向: ChatGPT-4o 更加注重语言的通俗性与流畅性, 而 DeepSeek-R1 在保持原文风格、术语一致性及正式表达方面更为稳健。这种差异反映了模型训练中对于可读性与忠实性的不同权衡策略, 也提示未来模型发展需进一步明确目标定位。这一发现印证了关于“大语言模型翻译在高语用场景中尚不具备独立译能”的判断[24], 同时也呼应了“AI 需提升文体适应性与语境理解力”的建议[11]。因此, 当前阶段 AI 译文虽已具备参与经贸文本翻译的基础能力, 但在模型调优与译文风格控制方面仍有进步空间。

在此基础上, 大语言模型翻译正加速融入新闻采编、国际机构文稿处理、跨文化沟通与出版传播等多类跨语际场景[25][26]。尽管相关技术提升了跨语言传播效率, 本研究发现, 即便表现较优的 DeepSeek-R1, 在语义复杂或语境依赖度较高的文本中, 仍可能出现结构冗余与词语选择偏差, 并在语用要求较高的文本中, 影响理解准确性。例如, 2018 年谷歌翻译将西班牙短语“自由、平等”误译为“奴役”, 引发讨论[27]。就目前大语言模型翻译技术的应用而言, 大语言模型翻译只是突破了部分不复杂、对译文质量要求不高的场景, 远未达到足以取代专业译员的翻译水平[24]。因此, 在当前阶段, 大语言模型尚

不足以独立承担语境依赖度较高的专业翻译任务, 其更适宜被定位为翻译流程中的辅助工具而非最终决策主体。

基于上述表现与风险分析, 经贸文本翻译宜采用“AI 初译 + 人工审校”的协作模式: 大语言模型负责初步生成, 译者侧重语用与立场把控; 其中, 专业性较强的内容应由人工重点审校。此种协作逻辑不仅回应了“实现工具 - 合作者 - 伙伴 - 相互依存”人机关系演进模式[28], 也进一步推动人机协作模式向更深层次的自主型协作演进[11]。译员应培养人机共生意识, 落实协同发展, 才能充分利用智能工具提升翻译生产力[5]。此外, 可通过“术语统一 - 结构解析 - 语用调控”三层机制增强 AI 译文的可控性, 从而提升译文质量并夯实人机协作的信任基础。

综上, 大语言模型在经贸文本翻译中的应用潜力已逐步显现, 未来有必要通过完善人机协作机制与流程设计, 推动其在专业翻译场景中的规范化应用。

5.2. 研究局限性

尽管本研究在模型对比与多指标评估方面力求保持方法一致性, 但仍存在一定局限性。首先, 受语料规模与文本类型限制, 研究对象主要集中于单一语言方向的经贸文本翻译, 这在一定程度上影响了结论向其他语种和多语言场景中的适用范围。其次, 大语言模型处于快速迭代阶段, 不同版本在训练规模与推理机制等方面可能存在差异, 本文结果主要反映特定时间节点下模型的翻译表现。此外, 受限于模型训练数据与过程的不透明性, 本文无法对不同模型的训练语料构成进行实证比较, 相关讨论仅基于输出特征与既有研究作出合理推测。未来研究可通过扩展语料规模、翻译方向及模型类型, 对相关结论进行进一步验证。

6. 结语

本研究对比分析了 ChatGPT-4o 与 DeepSeek-R1 在经贸文本汉英翻译任务中的表现。结果表明, 在 BLEU 与 TER 指标下, DeepSeek-R1 在准确性与结构一致性方面优于 ChatGPT-4o; 在可读性层面, 两款模型译文均较标准译文有所简化, 其中 DeepSeek-R1 在风格模仿与术语保留方面更接近人工翻译。研究进一步发现, 在处理经贸类文本时, 尽管大语言模型在词汇与句法层面的生成能力已相对成熟, 但在语篇层面的语境重构、态度调节与跨文化语用适配方面仍存在不足, 短期内难以完全替代专业译者。总体而言, 大语言模型更适宜作为翻译流程中的初译工具, 通过“AI 初译 + 人工审校”的协作机制, 在提升翻译效率的同时实现质量可控, 为高语境敏感的经贸文本翻译提供一种可行路径。

参考文献

- [1] Hu, J., Floyd, S., Jouravlev, O., Fedorenko, E. and Gibson, E. (2023) A Fine-Grained Comparison of Pragmatic Language Understanding in Humans and Language Models. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers), Toronto, 9-14 July 2023, 4194-4213. <https://doi.org/10.18653/v1/2023.acl-long.230>
- [2] Mahowald, K., Ivanova, A.A., Blank, I.A., Kanwisher, N., Tenenbaum, J.B. and Fedorenko, E. (2024) Dissociating Language and Thought in Large Language Models. *Trends in Cognitive Sciences*, **28**, 517-540. <https://doi.org/10.1016/j.tics.2024.01.011>
- [3] 胡开宝, 李晓倩. 大语言模型背景下翻译研究的发展: 问题与前景[J]. 中国翻译, 2023, 44(6): 65-74.
- [4] 刘海涛, 元达. 大语言模型的语用能力探索——从整体评估到反语分析[J]. 现代外语, 2024, 47(4): 439-451.
- [5] 王贇, 张政. ChatGPT 人工智能翻译的隐忧与纾解[J]. 中国翻译, 2024, 45(2): 95-102.
- [6] 文旭, 田亚灵. ChatGPT 应用于中国特色话语翻译的有效性研究[J]. 上海翻译, 2024(2): 27-34+94-95.
- [7] 冯志伟, 张灯柯. 大语言模型翻译与人工翻译相辅相成[J]. 外国语, 2022, 45(6): 77-87.

- [8] 周兴华, 王传英. 人工智能技术在计算机辅助翻译软件中的应用与评价[J]. 中国翻译, 2020, 41(5): 121-129.
- [9] 崔启亮. 人工智能时代的语言服务行业发展趋势[J]. 北京第二外国语学院学报, 2025, 47(1): 62-72.
- [10] 于天博. ChatGPT 汉译英 2023 年政府工作报告翻译特征研究[J]. 语言与文化研究, 2024, 32(4): 154-158.
- [11] 李焱, 齐心圆. 人机协作模式下翻译行业的转型与创新路径[J]. 外语电化教学, 2025(2): 31-36+46+101.
- [12] 林榕, 林大津. 政治文献英译: 文化翻译观——以《习近平谈治国理政》英译本为例[J]. 东南学术, 2016(6): 235-240.
- [13] 张佑明. 政治文献汉译英译文“地道性”探微[J]. 上海翻译, 2017(2): 30-34+93.
- [14] 司显柱, 曾剑平. 对外政治话语翻译: 原则、策略、成效——以《习近平谈治国理政》的英译为例[J]. 上海翻译, 2021(2): 18-24.
- [15] 雷鹏飞, 张浮凌. 基于大语言模型翻译软件的外宣文本翻译质量评估研究[J]. 未来与发展, 2024, 48(6): 41-48.
- [16] 任文, 赵田园. 国家对外翻译传播能力研究: 理论建构与实践应用[J]. 上海翻译, 2023(2): 1-7+95.
- [17] 胡开宝. 国家外宣翻译能力: 构成, 现状与未来[J]. 上海翻译, 2023(4): 1-7, 95.
- [18] Papineni, K., Roukos, S., Ward, T. and Zhu, W. (2002) BLEU: A Method for Automatic Evaluation of Machine Translation. *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, Philadelphia, 7-12 July 2002, 311-318. <https://doi.org/10.3115/1073083.1073135>
- [19] Snover, M., Dorr, B., Schwartz, R., Micciulla, L. and Makhoul, J. (2006) A Study of Translation Edit Rate with Targeted Human Annotation. *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers*, Cambridge, 8-12 August 2006, 223-231.
- [20] Gunning, R. (1952) *The Technique of Clear Writing*. McGraw-Hill.
- [21] Kincaid, P., Fishburne, R.P., Rogers, R.L. and Chissom, B.S. (1975) Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel. Research Report, University of Central Florida.
- [22] 周成彬, 刘忠宝. 基于语义信息共享 Transformer 的古文大语言模型翻译方法[J]. 情报工程, 2022, 8(6): 114-127.
- [23] 刘松. 人工智能翻译与国际传播: 现状、问题与展望[J]. 外语学刊, 2025(3): 20-25.
- [24] 张法连. 法律翻译中的大语言模型翻译技术刍议[J]. 外语电化教学, 2020(1): 53-58+8.
- [25] 姬德强, 李喆. 数字公域: 数字文化视野中的国际传播新秩序[J]. 对外传播, 2024(4): 29-33.
- [26] 李强辉. 出版产业视域下的 AI 赋能与冲击研究[J]. 科技与出版, 2024(4): 71-79.
- [27] Titone, D.A. and Tiv, M. (2023) Rethinking Multilingual Experience through a Systems Framework of Bilingualism: Response to Commentaries. *Bilingualism: Language and Cognition*, **26**, 247-251. <https://doi.org/10.1017/s1366728922000785>
- [28] 秦颖. 人机共生场景下的外语教学方法探索——以 ChatGPT 为例[J]. 外语电化教学, 2023(2): 24-29+108.