

五款生成式人工智能译文质量 对比分析及应对策略

——以医学文献汉译为例

刘 龙, 陈丽珠

(武汉工程大学 外语学院, 湖北 武汉 430205)

摘 要: 采用定性和定量分析相结合的方法研究对比了五款生成式人工智能在医学文献汉译中的译文质量。针对生成式人工智能可能给译者带来的不利影响, 从源文本选择、翻译工作模式、译者背景知识、译者综合能力四个层面提出了应对策略。

关键词: 生成式人工智能; 医学文献汉译; 译文质量; 对比分析; 应对策略

中图分类号: H085

文献标识码: A

文章编号: 1009-4970(2025)05-0091-06

DOI:10.16594/j.cnki.41-1302/g4.2025.05.010

0 引言

生成式人工智能是一种可以生成文本和创意内容以及整合来自不同来源数据进行分析的人工智能^[1]。目前生成式人工智能译文对比的相关研究大致可分三类。一是研究对比生成式人工智能和主流神经机器翻译, 如王和私等以生物医学文本为例, 使用目前主流的热门翻译软件, 包括有道翻译、百度翻译、谷歌翻译和 ChatGPT, 进行英汉翻译及汉英翻译, 比较分析了各个软件存在的优缺点^[2]; 孙(Son)等对比研究了大型语言模型和神经机器的翻译性能^[3]。二是探究生成式人工智能译后编辑, 如耿芳等基于翻译实例文本, 从错误修正、句法结构、篇章连贯及语域调整等角度分析了 ChatGPT 在英汉与汉英翻译任务中的译后编辑表现^[4]。三是探究生成式人工智能对外语教学的影响, 如周忠良探讨了 ChatGPT 给翻译教学带来的变革、挑战及相关应对策略^[5], 佩雷斯-努涅斯(Pérez-Núñez)分析了 ChatGPT 在外语教学中的应用及挑战^[6]。

经过仔细梳理, 笔者发现上述鲜有研究同类对比生成式人工智能的译文质量。丁敏娇等强调, 医学文献的翻译难点主要包括三个方面: 其一, 未能

准确判断词的含义; 其二, 在语言句式转换时未能实现句子的通顺; 其三, 未采取学术机构名、课题名的既定译文^[7]。鉴于此, 本文采用定性和定量分析相结合的研究方法, 以文心一言、通义千问、GPT-4、Claude 2、Bard 为载体, 从忠实度、可懂度两个维度对比分析五款生成式人工智能的译文质量。研究期望助力从事医学翻译行业的译者更好地利用生成式人工智能进行医学文献汉译, 并提出相关应对策略, 以帮助译者应对生成式人工智能可能带来的不利影响。

1 研究设计

1.1 研究材料

纽马克(Newmark)按语言功能种类, 划分出表达型、信息型和呼唤型三种文本类型^[8], 其中信息型文本侧重描述和讨论科学事实, 具有较强的客观性、准确性和严密性, 旨在传递信息, 不掺杂个人情感。本研究选择信息型文本供生成式人工智能翻译, 源语节选自 *The New England Journal of Medicine* 期刊 2023 年 12 月至 2024 年 1 月发表的 6 篇医学文献, 译文来自《NEJM 医学前沿》官网译文。

收稿日期: 2024-05-09

作者简介: 刘龙(1999—), 男, 湖北咸宁人, 硕士研究生; 陈丽珠(1977—), 女, 湖北黄冈人, 副教授, 研究方向为外事翻译、文学翻译。

1.2 对比标准

日本科学技术厅将机器翻译的译文评估分为忠实度(共 7 级)和可懂度(共 5 级)^[9]。已有学者选用可懂度和忠实度两个参考标准研究对比了谷歌翻译、微软必应、百度翻译、有道翻译、ChatGPT 五款机器翻译的译文质量^[10]。

本研究以可懂度和忠实度为评估标准,并分别量化忠实度等级、可懂度等级。评价人员按照 0—6 分给忠实度打分,最后忠实度分值是所有评价人员的平均分。评价人员按照 0—4 分给可懂度打分,最后可懂度分值是所有评价人员的平均分。

1.3 研究步骤

第一步,按照文心一言、通义千问、GPT-4、Claude 2、Bard 的顺序,将五款生成式人工智能分别标记为软件 1、软件 2、软件 3、软件 4、软件 5。

第二步,从上文所述的 6 篇文献中选取 6 个例句(源语单句最长包括 120 个单词,最短 70 个单词,平均每句包含 85 个单词)作为研究材料。对于五款生成式人工智能的提示语,研究参照 Jiao

等^[11]研究成果,按照该提示语“Please provide the [Chinese version] translation for these sentences”生成 30 个句子,五款生成式人工智能软件共计 30 个例句。

第三步,采用定性研究方法,从已选定的 6 个例句中选出 1 个例句,基于忠实度和可懂度两个维度探究五款生成式人工智能在医学文献汉译中的表现。

第四步,采用定量研究方法,设计调查问卷。问卷内容包括评价流程说明、源语、参考译文、例句、忠实度等级选项、可懂度等级选项。五款软件共计 30 个句子,忠实度和可懂度两个维度共计 60 个例句。

第五步,邀请来自不同领域的 23 名评价人员进行评价(评价人员均互不认识,并且五款人工智能相关信息均已被隐藏)。笔者要求评价人员阅读完问卷的“评价流程说明”后开始进行评价;同时要求他们以参考译文为标准,严格按照忠实度和可懂度两个维度评价五款软件的例句,时间限时一周。评价人员组成如表 1 所示。

表 1 评价人员组成

评价人员类型	人数	资质
翻译硕士专业教师	4	副教授及以上。
翻译硕士专业研究生	10	5 名通过 CATII 二级笔译;5 名在各类翻译比赛中均获奖。
医学硕士研究生	6	4 名评价人员来自西安交通大学,1 名来自浙江大学,1 名来自北京大学,CET 6 成绩平均分 570。
翻译公司职员(医学翻译)	3	3 名公司职员,均有 1 年以上的医学翻译经验。

第六步,收集整理数据,运用软件 SPSS 27.0 分析数据,计算出每款软件忠实度平均值、可懂度平均值、标准偏差,运用 Kolmogorov-Smirnov 检验和 Shapiro-Wilk 检验对数据进行正态性检验。

第七步,结合定性分析和定量分析的结果,对比分析五款软件在医学文献汉译中的表现,判断哪款生成式人工智能软件在医学文献汉译中表现最优。

2 结果分析

2.1 定性分析

笔者采用定性研究方法,从已选定的 6 个例句中节选 1 个例句,对源语进行编号。然后,笔者从忠实度和可懂度两个维度分析对比五款生成式人工智能的译文。具体例子如下文所示:

例 The diagnoses¹ were made according to the National Institute of Aging-Alzheimer's Association² research framework, which³ includes evidence of pathologic changes in Alzheimer's disease as indicated by

• 92 •

abnormal levels of A β accumulation⁴, a standardized uptake value ratio (SUVr) greater than 1.4 in signature regions of the brain that are associated with Alzheimer's disease (i. e., the middle temporal gyrus, inferior temporal gyrus, temporal pole, superior frontal gyrus, middle frontal gyrus, precuneus, supramarginalgyrus, angular gyrus, and superior parietal lobule) on 18F-florbetaben PET scan⁵, and elevated levels (>21.7 pg per milliliter) of phosphorylated tau⁶ in cerebrospinal fluid (individual levels along with inclusion and exclusion criteria are shown in Table S2 in the Supplementary Appendix, available with the full text of this article at NEJM.org).

软件 1: 诊断是根据美国国立衰老研究所-阿尔茨海默病协会的研究框架作出的,其中包括阿尔茨海默病病理变化的证据,如 A β 积累的异常水平,标准化摄取值比率(SUVr)大于 1.4,与阿尔茨海默病相关的大脑特征区域(即中颞叶、下颞叶、

颞极、前额叶上皮质、中额叶、背侧内侧前额叶皮质、顶叶、角回和顶上小叶)在 18F-氟-β-淀粉样蛋白正电子发射断层扫描(PET)上的摄取值比率(SUVr)大于 1.4, 以及脑脊液中磷酸化 tau 水平升高(>21.7 pg/ml)(个体水平以及纳入和排除标准见补充附录表 S2)。

软件 2: 这些诊断是根据美国国家老龄化研究所-阿尔茨海默病协会的研究框架进行的, 该框架包括了阿尔茨海默病病理变化的证据, 表现为异常水平的 Aβ 积累, 以及在与阿尔茨海默病相关的脑部特征区域(即, 中颞叶、下颞叶、颞极、额上回、额中回、楔前叶、缘上回、角回和顶上小叶)的标准化摄取值比(SUVr)大于 1.4, 在 18F-氟苯并亚甲基苯乙炔 PET 扫描中显示, 并且在脑脊液中磷酸化 tau 的水平升高(>21.7 pg 每毫升)(个体水平以及纳入和排除标准见本补充附录表 S2, 可在 NEJM.org 网站获取全文)。

软件 3: 诊断是根据国家老龄化研究所-阿尔茨海默病协会研究框架制定的, 该框架包括阿尔茨海默病病理变化的证据, 如 Aβ 堆积的异常水平, 标准化摄取值比率(SUVr)在 1.4 以上(在与阿尔茨海默病相关的大脑特征区域, 即中颞回、下颞回、颞极、上额回、中额回、前楔、额顶缘回、角回和顶上小叶)的 18F-氟贝坦 PET 扫描, 以及脑脊液中磷酸化 tau 的水平升高(>21.7 pg 每毫升)(个别水平以及纳入和排除标准在补充附录的表 S2 中显示, 可在 NEJM.org 的本文全文中查看)。

软件 4: 诊断是根据美国国立老年病研究所-阿尔茨海默氏病协会的研究框架制定的, 其中包括阿尔茨海默氏病的病理变化证据, 如脑内相关区域异常水平的 β 淀粉样蛋白积累, 这些区域与阿尔茨海默氏病相关(即, 颞中回、颞下回、颞极、前额上回、前额中间回、枕叶、楔前部、角回和顶上小叶), 18F-弗贝他宾 PET 扫描标准化摄取值比大于 1.4, 以及脑脊液中磷酸化 tau 蛋白水平升高(>21.7 pg/ml)(各个水平及入选和排除标准详见该文章的补充材料表 S2, 可在 NEJM.org 上获得全文)。

软件 5: 诊断方法遵循美国国立老龄化研究所-阿尔茨海默病协会的研究框架, 该框架包括以下证据: 阿尔茨海默病相关的病理改变证据, 由 F-18 氟贝塔苯 PET 扫描检测到异常的 Aβ 沉积水平, 具体表现为脑内与阿尔茨海默病相关的标志性区域(即中颞回, 下颞回, 颞极, 上额回, 中额回, 前扣

带, 缘上回, 角回, 上顶叶小叶)的标准摄取值(SUVr)高于 1.4. 脑脊液中磷酸化 tau 蛋白的浓度升高(>21.7 pg/mL)(具体浓度值以及纳入和排除标准见 NEJM.org 上这篇全文的补充附录 S2 表)。

参考译文: 诊断依据是美国国立老龄化研究所-阿尔茨海默病学会(National Institute of Aging-Alzheimer's Association)研究框架, 该框架包括阿尔茨海默病的病理变化证据(具体表现为异常水平 Aβ 累积), 18F-florbetaben PET 扫描显示阿尔茨海默病相关特征性脑区(即颞中回、颞下回、颞极、额上回、额中回、楔前叶、缘上回、角回和顶上小叶)的标准化摄取值比(SUVr)大于 1.4, 以及脑脊液中磷酸化 tau 蛋白水平升高(>21.7 pg/mL)(个体水平以及纳入和排除标准见补充附录表 S2, 补充附录与本文全文可在 NEJM.org 获取)。

分析: 软件 1 得分(2, 2): 序号 1、2、3、6 等 4 处需修改, 并且源语最后一句话(available with the full text of this article at NEJM.org)漏译; 序号 5 核心意思为“阿尔茨海默病相关特征性脑区的标准化摄取值比(SUVr)大于 1.4”, 译文基本译出, 然而序号 5 括号内相关术语还需修改。

软件 2 得分(4, 3): 序号 1、2、6 等 3 处需修改; 序号 5 译文只出现些许错误, 不影响阅读, 很容易更正, 这部分需要调整“在 18F-氟苯并亚甲基苯乙炔 PET 扫描中显示”的位置, 以实现译文可懂度。

软件 3 得分(3, 2): 序号 1、2、4、6 等 4 处需修改; 序号 5 译文有误, 未能准确处理“18F-florbetaben PET scan”的位置, 且这部分译文意思模棱两可。

软件 4 得分(4, 1): 序号 1、3、4 等 3 处需修改; 序号 5 译文错误较多, 需重新翻译。

软件 5 得分(4, 2): 序号 1、4、6 等 3 处需修改; 序号 5 译文基本译出, 但序号 5 括号内相关术语还需修改, 并且需调整“Aβ 沉积水平”的位置。

经上述定性对比分析, 五款生成式人工智能生成的译文均未达到参考译文标准, 无法直接使用, 译者需进行不同程度的译后编辑。综上, 生成式人工智能在专业文本翻译上存在一定缺陷, 生成式人工智能的专业性不如人工译者, 这也正是人工译者的优势所在。

2.2 定量分析

笔者采用定量研究方法, 利用软件 SPSS 27.0 进行描述统计。就医学文献汉译而言, 生成式人工

智能翻译总共有三种得分：一是忠实度平均值；二是可懂度平均值；三是忠实度 + 可懂度平均值。相关数据结果如图 1 所示。

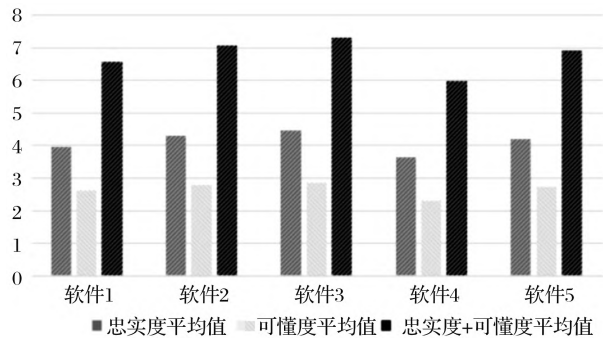


图 1 5 种软件的忠实度、可懂度、忠实度 + 可懂度平均值分布

从忠实度平均值来看：软件 2、软件 3 得分最高，分别为 4.30、4.45；软件 1、软件 5 得分次之，分别为 3.95、4.20；软件 4 得分最低，仅有 3.65。从可懂度平均值来看：软件 2、软件 3 得分最高，分别为 2.78、2.86；软件 1、软件 5 得分次之，分别为 2.62、2.72；软件 4 得分最低，仅有 2.32。

进行正态性检验后，发现 Kolmogorov-Smirnov 检验和 Shapiro-Wilk 检验二者显著性均小于 0.001，数据不符合正态分布。由于部分网页版机器翻译（软件 2 和软件 4）的得分不呈正态分布，故而使用 Kruskal-Wallis 非参数检验法检验各大网页版机器翻译汉英翻译总体得分之间的差异是否存在显著性^[12]。由于本文的相关数据不符合正态分布，笔者也使用 Kruskal-Wallis 非参数检验法检验五款生成式人工智能翻译软件忠实度和可懂度得分是否存在显著差异。结果显示在 0.05 的显著性水平上，忠实

度分值的显著性小于 0.001，五款软件忠实度分值存在显著差异；可懂度显著性为 0.002，五款软件可懂度分值存在显著差异。

表 2 Kruskal-Wallis 非参数检验

	忠实度	可懂度
克鲁斯卡尔 - 沃利斯检验 H	23.664	17.539
自由度	4	4
渐近显著性	<0.001	0.002

笔者在检验前已通过 Bonferroni 校正法调整显著性值(0.05)，再运用 Kruskal-Wallis 非参数检验法检验五款生成式人工智能之间忠实度分值是否存在显著性差异以及五款生成式人工智能之间可懂度分值是否存在显著差异。

结果表明：从忠实度来看，软件 4 和软件 5 存在显著差异(0.044)，软件 4 和软件 2 存在显著差异(0.004)，软件 4 和软件 3 存在显著差异(0.000)，软件 1 和软件 3 存在显著差异(0.042)，其余生成式人工智能之间不存在显著差异。从可懂度来看，软件 4 和软件 2 存在显著差异(0.012)，软件 4 和软件 3 存在显著差异(0.001)，其余则不存在显著差异。五款生成式人工智能的成对比较如表 3 和表 4 所示。

结合图 1、表 2、表 3、表 4，笔者将五款生成式人工智能译文质量划分为三个梯度：表现最好组（通义千问、GPT - 4）；表现较好组（文心一言、Bard）；表现较差组（Claude2）。由上述分组可见，通义千问和 GPT - 4 译文表现较好，这两款人工智能的译文只需简单修订，译文兼顾忠实度和可懂度，便于目标读者阅读和理解。

表 3 五种生成式人工智能译文忠实度的成对比较

Sample1 - Sample2	检验统计	标准误差	标准检验统计	显著性	调整后的显著性*
4 - 1	34.051	22.898	1.487	0.137	1.000
4 - 5	-65.178	22.898	-2.846	0.004	0.044
4 - 2	80.533	22.898	3.517	<0.001	0.004
4 - 3	99.587	22.898	4.349	<0.001	0.000
1 - 5	-31.127	22.898	-1.359	0.174	1.000
1 - 2	-46.482	22.898	-2.030	0.042	0.424
1 - 3	-65.536	22.898	-2.862	0.004	0.042
5 - 2	15.355	22.898	0.671	0.502	1.000
5 - 3	34.409	22.898	1.503	0.133	1.000
2 - 3	-19.054	22.898	-0.832	0.405	1.000

注：经 Bonferroni 校正后的显著性水平，下同

表 4 五种生成式人工智能译文可懂度的成对比较

Sample1 - Sample2	检验统计	标准误差	标准检验统计	显著性	调整后的显著性*
4 - 1	47.174	22.861	2.063	0.039	0.391
4 - 5	-62.435	22.861	-2.731	0.006	0.063
4 - 2	73.870	22.861	3.231	0.001	0.012
4 - 3	87.989	22.861	3.849	<0.001	0.001
1 - 5	-15.261	22.861	-0.668	0.504	1.000
1 - 2	-26.696	22.861	-1.168	0.243	1.000
1 - 3	-40.815	22.861	-1.785	0.074	0.742
5 - 2	11.435	22.861	0.500	0.617	1.000
5 - 3	25.554	22.861	1.118	0.264	1.000
2 - 3	-14.120	22.861	-0.618	0.537	1.000

3 应对策略

3.1 转向专业性更强、要求更高的文本

综上分析,生成式人工智能在专业文本翻译上存在一定缺陷,生成式人工智能的专业性不如人工译者,这也正是人工译者的优势所在。技术进步和自动化普及的同时也使得人类可以专注于更有意义、更为复杂的任务;通常情况下,那些更为复杂的任务可以带来更高的成就感,因为此类复杂任务绝非像可自动化任务那样具有可重复性^[13]。就具体源语而言,未来人工译者主要面向对译文质量要求较严格的文本,比如法律文书、医学专著、政治文献等专业内容,以及需要高度创造力和想象力的领域,如文学、艺术、哲学等人文方面^[14]。

3.2 选择生成式人工智能 + 人工译后编辑翻译模式

首先,生成式人工智能可以最大限度地仿拟和集成人类已有翻译经验,却无法自主生成基于自身体验的原创性文本,因而也无法完全再现作为人的原作者(源语发出者)的经验世界^[15]。其次,“ChatGPT + MTPE”交互式译后编辑能力培养模式的有效性,为 ChatGPT 时代生成式人工智能赋能翻译能力培养提供理论指导和实证支持^[16]。典型大模型平台包括 GPT 系列、Claude 系列、PaLM 系列、Bard、文心一言、讯飞星火、通义千问等^[17]。由此可见,文心一言、通义千问、GPT-4、Claude2、Bard 属于同一时代下的产物,隶属同一类别。综上所述,生成式人工智能翻译 + 人工译后编辑翻译模式已经过理论和实证研究,该模式能为译者所运用。

3.3 补充专业背景知识

尽管生成式人工智能能快速生成译文,从而极大减轻译者负担,但生成式人工智能还是会犯不同程度的错误。无论机器多么智能,也无论译者花多长时间做译前编辑工作,未成功翻译的文本还是占

有一定比例,还需进行译后编辑^[14]。换言之,人工智能也会出现翻译错误乃至误译,这些错误需人工译者在译后编辑环节修改。在译后编辑环节,译者只有具备丰富的专业背景知识,才能做到既发现生成式人工智能的翻译错误,又能准确修订错误,产出兼具忠实度和可信度的译文。对于缺少专业背景的译者,译者可以查找和阅读大量与某一领域有关的平行文本。平行文本指与源语主题相同的非翻译性质的目标语文本^[18]。汪洋强调,平行文本的意义包括两方面:一方面,平行文本可以帮助译者获取相关领域的专业知识;另一方面,平行文本也可以获取目标语在专业领域的表达方式^[19]。

3.4 掌握除翻译之外的综合能力

在生成式人工智能时代下,译者更需要掌握除翻译之外的综合能力,诸如熟练运用翻译技术的能力、排版能力,以提升翻译或译后编辑效率。传统型翻译人才已逐渐丧失优势,译者应成为复合型专职翻译人才^[20]。复合型专职翻译人才不仅要有过硬的双语能力,也要熟悉翻译技术,掌握术语管理、译前和译后编辑等一系列能力。一方面,翻译记忆库能够在翻译过程中提供重复或相似的语料,避免译者再次翻译重复内容,以此提高翻译效率^[21]。同时,术语库也能帮助译者解决重复出现的术语,避免重复查找术语。另一方面,在真实翻译项目中,基于客户反馈和出版要求进行文本格式和排版编辑的现象时有发生,这也说明排版和格式也是译后编辑工作的一个重要部分^[16]。由此说明,译者更需牢牢掌握除翻译之外的综合能力。

4 结语

研究以文心一言、通义千问、GPT-4、Claude2、Bard 为载体,采用定性和定量结合的研究方法,从忠实度和可懂度两个维度对比分析五款生成式人工智能译文质量。研究期望助力从事医学翻

译行业的译者更好地利用生成式人工智能进行医学文献汉译, 及提出相关应对策略以帮助译者应对生成式人工智能带来的不利影响。然而, 研究只探究五款生成式人工智能在医学文献汉译中的表现, 未探究五款生成式人工智能在英译中的表现, 未来需继续探究五款生成式人工智能在医学文献互译中的译文质量。人工智能时代应是人机合作的时代。生成式人工智能赢在“量”和效率, 人工翻译赢在“质”和专业。生成式人工智能的优势与人工翻译的强项二者相互耦合, 以此共同实现“质量”, 促进翻译行业交流和发展、提升翻译质量、翻译效率。

参考文献

- [1] DASBOROUGH M T. Awe-inspiring advancements in AI: The impact of ChatGPT on the field of organizational behavior[J]. *Journal of Organizational Behavior*, 2023(2): 177 – 179.
- [2] 王和私, 马柯昕. 人工智能翻译应用的对比研究: 以生物医学文本为例[J]. *中国科技翻译*, 2023(3): 23 – 26.
- [3] SON J, KIM B. Translation Performance from the User's Perspective of Large Language Models and Neural Machine Translation Systems [J]. *Information*, 2023 (10): 574 – 590.
- [4] 耿芳, 胡健. 人工智能辅助译后编辑新方向: 基于 Chat GPT 的翻译实例研究[J]. *中国外语*, 2023(3): 41 – 47.
- [5] 周忠良. ChatGPT 在翻译教学中的应用: 变革、挑战与应对[J]. *北京第二外国语学院学报*, 2023(5): 134 – 146.
- [6] PÉREZ-NÚÑEZ A. Exploring the Potential of Generative AI (ChatGPT) for Foreign Language Instruction: Applications and Challenges[J]. *Hispania*, 2023(3): 355 – 362.
- [7] 丁敏娇, 吴健敏. 医学英文文献翻译难点及处理策略[J]. *温州医科大学学报*, 2016(4): 310 – 312.
- [8] NEWMARK P. *Approaches to Translation* [M]. Shanghai: Shanghai Foreign Language Education Press, 2001: 12 – 15.
- [9] 张政, 王贵明. 论机器翻译系统的评价体系[J]. *北京理工大学学报(社会科学版)*, 2008(2): 112 – 117.
- [10] 毛佳敏, 张继光. 国内外五种机器翻译平台对比: 以特斯拉公司简介中文译文为例[J]. *译苑新谭*, 2023(2): 69 – 77.
- [11] JIAO W X, WANG W X, HUANG J T, et al. Is ChatGPT a good translator? Yes with GPT-4 as the engine[EB/OL]. (2023-11-02) [2024-02-01]. <https://www.arxiv.org/abs/2301.08745v4>.
- [12] 李奉栖. 基于神经网络的在线机器翻译系统英汉互译质量对比研究[J]. *上海翻译*, 2021(4): 46 – 52.
- [13] DESJARDINS R, LARSONNEUR C, LACOUR P. When Translation Goes Digital: Case Studies and Critical Reflections[M]. Cham: Palgrave Macmillan, 2021: 1 – 16.
- [14] 宋仕振. 试论机器翻译与人工翻译的未来关系[J]. *未来与发展*, 2019(2): 25 – 30.
- [15] 杜安. ChatGPT 时代人工翻译的价值空间与翻译教育转型[J]. *外国语言与文化*, 2023(4): 90 – 103.
- [16] 王律, 王湘玲. ChatGPT 时代机器翻译译后编辑能力培养模式研究[J]. *外语电化教学*, 2023(4): 16 – 23, 115.
- [17] 陶建华, 吴飞, 黄民烈, 等. 中国人工智能系列白皮书: 大模型技术(2023 版)[R]. 北京: 中国人工智能学会, 2023: 41 – 65.
- [18] PYM A. *Exploring Translation Theories* [M]. 3rd ed. London: Routledge, 2023: 50.
- [19] 汪洋. 运用现代技术手段英译专门科技文本: 以航海技术领域论文为例[J]. *上海翻译*, 2013(2): 37 – 41.
- [20] 朱一凡, 管新潮. 人工智能时代的翻译人才培养: 挑战与机遇[J]. *上海交通大学学报(哲学社会科学版)*, 2019(4): 37 – 45.
- [21] 张法连. 法律翻译中的机器翻译技术刍议[J]. *外语电化教学*, 2020(1): 53 – 58, 8.

[责任编辑 赵永亮]

A Comparative Analysis on Translation Quality of Five Generative Artificial Intelligence Translations and Countermeasures —A Case Study on E-C Translation of Medical Literature

Liu Long, Chen Lizhu

(School of Foreign Languages, Wuhan Institute of Technology, Wuhan 430205, China)

Abstract: This paper employs a combined qualitative and quantitative analysis method to compare the translation quality of five generative artificial intelligence (AI) tools in E-C translation of medical literature. In response to the potential adverse impacts of generative AI on translators, this study proposes countermeasures from four aspects including source text selection, translation working mode, translators' background knowledge, and translators' comprehensive abilities.

Key Words: generative artificial intelligence; E-C translation of medical literature; translation quality; comparative analysis; countermeasure