

人工智能伦理视角下的英语翻译问题研究

张雪 胡芳

青岛恒星科技学院, 山东 青岛 266100

摘要: 人工智能技术的快速发展为英语翻译领域带来革命性变革, 同时也引发诸多伦理问题。本研究从人工智能伦理视角出发, 系统探讨机器翻译过程中存在的文化偏见、隐私泄露、责任归属等核心议题。通过文献分析与案例研究相结合的方法, 揭示当前主流翻译算法在文化敏感信息处理、用户数据保护等方面存在的局限性。研究发现, 机器翻译系统可能无意识地强化文化刻板印象, 其数据采集过程存在侵犯用户隐私的风险, 而翻译错误导致的法律责任界定尚不明确。

关键词: 人工智能伦理; 英语翻译; 机器翻译; 伦理问题; 跨文化交际

DOI: 10.63887/jc.2025.1.3.7

1 研究背景与研究目的

随着人工智能技术的迅猛发展, 其在英语翻译领域的应用已从简单的工具辅助演变为具有自主决策能力的复杂系统。这种技术变革在提升翻译效率的同时, 也引发了诸多伦理层面的争议。当前主流机器翻译系统普遍采用深度学习算法, 其训练数据的选择与处理过程往往隐含文化偏见, 导致翻译结果可能无意识地强化特定文化刻板印象。例如, 某些语种间的互译常因语料库不平衡而出现性别或种族偏见, 这种现象在跨文化交际场景中可能造成严重的误解。

技术层面之外, 数据隐私问题同样值得关注。神经机器翻译系统需要海量语料进行模型训练, 而用户提交的翻译请求常包含敏感信息。现有研究表明, 部分商业翻译平台存在用户数据二次利用的情况, 且数据脱敏技术尚未完善, 这直接威胁到个人隐私安全。更复杂的是, 当翻译错误引发法律纠纷时, 责任主体往往难以界定——算法开发者、平台运营商与终端用户之间的权责关系缺乏明确的法律规范。

本研究旨在系统梳理人工智能英语翻译中存在的核心伦理问题, 包括但不限于文化偏见、隐私风险与责任模糊等议题。通过构建理论分析框架, 探索技术优化与伦理规范协同发展的可行路径。具体研究目标包括: 揭示当前机器翻译算法在文化敏感信息处理方面的局限性; 评估不同数据保护技术的有效性; 提出

兼顾技术创新与伦理约束的人机协作模式。这些研究不仅有助于完善人工智能伦理理论体系, 更能为开发更具文化包容性的翻译系统提供实践指导, 最终促进跨文化交流的公平性与安全性。

2 人工智能伦理与英语翻译的理论基础

2.1 人工智能伦理的核心概念与框架

人工智能伦理作为技术哲学的重要分支, 其核心概念聚焦于技术应用中价值判断与行为规范的建立。在英语翻译领域, 人工智能伦理主要涉及算法决策的公平性、数据处理的透明性以及人机协作的责任性三大维度。王佑镁在研究中指出, “建立教育人工智能伦理风险分类框架需要系统考量技术应用全流程中的价值冲突”, 这一观点同样适用于翻译技术场景。具体而言, 机器翻译系统的伦理框架需包含文化价值平衡机制、隐私保护机制和权责分配机制, 三者共同构成技术应用的伦理边界。

从理论渊源来看, 人工智能伦理框架的构建主要基于义务论、功利主义和德性伦理三大哲学传统。义务论强调算法设计应遵循“不伤害”原则, 要求翻译系统避免输出具有文化冒犯性的内容; 功利主义则关注翻译结果的社会效益最大化, 主张通过优化语料库结构来提升跨文化沟通效率; 德性伦理则注重开发者的道德自觉, 要求技术团队在模型训练中主动识别并消除潜在偏见。易佳萦认为, “推动人工智能在正确

伦理轨道上发展需要多维度的价值引导”，这种多维性在翻译伦理中表现为对语言多样性、文化敏感性和用户自主性的协同保障^[1]。

在操作层面，人工智能翻译伦理框架包含三个递进层次：基础层关注技术合规性，要求算法符合数据保护法规和行业标准；中间层强调文化适应性，确保翻译模型能够识别并处理文化特定性表达；最高层则涉及价值判断，需要系统在面临伦理困境时（如敏感政治术语的翻译）具备风险预警能力。这种分层结构既体现了技术伦理的普遍性原则，又兼顾了语言服务的特殊性要求。值得注意的是，当前主流翻译平台在基础层建设相对完善，但在文化适应性和价值判断层面的伦理设计仍存在明显不足。

2.2 英语翻译中的伦理问题与挑战

英语翻译领域在人工智能技术应用过程中面临着多维度的伦理挑战，这些挑战深刻影响着翻译质量、文化传播和社会公平。肖志清指出，“翻译技术应用中的伦理问题在人工智能时代日渐凸显，包括准确性和偏见歧视、数据隐私和滥用、责任和法律问题等方面”。这些问题的产生既源于技术本身的局限性，也与算法设计中的价值取向密切相关。

文化偏见是当前机器翻译系统最突出的伦理困境。深度学习模型依赖训练语料库进行决策，当语料库中某些文化群体的表达占主导地位时，系统输出的翻译结果会无意识地强化既有的文化刻板印象。例如，在性别角色相关的翻译中，算法可能倾向于将特定职业与单一性别关联，这种隐性偏见在跨文化交际场景中可能造成严重误解。更复杂的是，某些语言间的互译存在语义不对称现象，当算法缺乏文化背景理解能力时，简单直译可能导致文化特定概念被曲解或简化。

数据隐私风险构成另一重伦理挑战。神经机器翻译系统需要持续吸收海量语料优化模型性能，而用户提交的翻译请求常包含个人身份信息、商业机密等敏感内容。现有数据脱敏技术难以完全消除隐私泄露风险，部分平台还存在用户数据二次利用的灰色操作。孙佳在研究教育领域机器翻译时发现，“机器翻译在大学英语教育教学领域的应用揭示了独特的人工智能

伦理风险”，这种风险在医疗、法律等专业翻译场景中表现得更为突出。

责任归属的模糊性进一步加剧了伦理困境的复杂性。当翻译错误引发法律纠纷或文化冲突时，责任主体往往难以明确界定。算法开发者、平台运营商、终端用户之间的权责关系缺乏清晰的法律规范，导致追责机制失效。特别是在“机器翻译+译后编辑”模式下，人类编辑对机器输出的修改程度与最终错误的责任分担比例尚无科学判定标准。这种不确定性不仅威胁用户权益，也阻碍了翻译行业的规范化发展^[2]。

教育领域的应用暴露出特殊的伦理矛盾。孙佳认为，“本文从人文价值、师生关系、情感盲区和数据决策等方面入手，探讨机器翻译的伦理失范问题”。机器翻译的普及可能削弱语言学习过程中的人文价值传递，使师生互动趋于工具化。同时，算法无法识别文本中的情感暗示和文化隐喻，导致翻译结果丧失原文的情感维度，这种“情感盲区”在文学翻译中尤为显著^[3]。

3 人工智能伦理视角下的英语翻译问题分析

3.1 人工智能翻译中的伦理冲突与案例分析

人工智能翻译技术在实际应用中产生的伦理冲突主要表现为文化价值冲突、隐私权冲突和责任归属冲突三个维度，这些冲突在具体案例中呈现出复杂的交互关系。文化价值冲突源于算法对语言文化内涵的处理失当，肖志清在研究中指出，“翻译技术研发者需要融合技术伦理考量，确保技术向善”，但现有系统往往难以平衡技术效率与文化敏感性之间的张力。例如，某主流翻译平台在处理土著语言时，将传统祭祀用语直译为“原始崇拜”，这种简化不仅扭曲了文化本义，更强化了文明等级偏见。类似案例表明，算法对低频语种的文化特定性表达缺乏识别能力，导致翻译结果沦为文化霸权的工具。

隐私权冲突集中体现在数据采集与处理环节。艾尚乐研究发现，“人工智能的应用中还存在一些伦理风险问题，比如数据隐私泄露”，这一现象在专业翻译场景尤为突出。某医疗翻译系统曾因未完全脱敏患者病历数据，导致敏感健康信息在模型训练过程中

被间接泄露。更隐蔽的风险在于，部分平台通过用户协议中的模糊条款获取数据二次使用权，使个人隐私面临商业化滥用威胁。这些案例揭示了当前数据治理框架的缺陷：既缺乏对敏感信息的动态识别机制，也缺少用户数据主权的有效保障措施。

责任归属冲突在翻译错误引发的法律纠纷中表现得尤为尖锐。当某跨境电商因机器翻译错误导致合同条款歧义，并造成重大经济损失时，平台运营商、算法开发商和终端用户三方相互推诿责任。这种困境反映了现有责任划分机制的滞后性——既未明确机器翻译作为“技术产品”还是“服务内容”的法律属性，也未建立人机协作过程中的过错认定标准。王炜琳注意到，“人工智能在英语写作教学中的应用存在伦理公平的问题”，这一观察同样适用于翻译领域，当算法决策缺乏透明性时，责任追溯变得异常困难。

3.2 伦理原则在英语翻译中的应用与优化策略

人工智能伦理原则在英语翻译领域的应用需要构建多层次、动态化的实施框架，其核心在于将抽象伦理准则转化为可操作的技术规范与行业标准。肖志清指出，“翻译技术使用者应强化个人伦理意识，自觉遵循伦理规范”^[4]，这一观点揭示了伦理原则落地的双重路径：既需要技术系统的内在约束，也依赖从业者的外部践行。在文化偏见治理方面，优化策略应聚焦于语料库的多元文化表征，通过建立动态平衡机制确保低频语种和文化特定表达获得公平呈现。具体而言，可采用文化敏感度分级标注技术，对涉及宗教、性别、种族等内容的语料进行伦理风险评估，并在模型训练阶段设置文化过滤层，从算法源头减少偏见输出。

数据隐私保护需贯穿翻译服务全生命周期，形成预防-监测-响应的闭环管理体系。在预处理环节，采用差分隐私和联邦学习技术可在保证模型性能的前提下实现数据最小化采集；在翻译过程中，部署实时敏感信息识别模块，对医疗记录、财务数据等实施自动脱敏；在后期存储阶段，则需建立数据生命周期管理制度，明确不同类别信息的保留期限与销毁流程。值得注意的是，当前隐私保护的技术方案往往与翻译质量存在此消彼长的关系，这要求开发者在设计阶段就进行伦理权衡，正如李红岩所述，“为遵循‘科技创

新伦理先行’的理念，需要对相关政策、法律法规及其现状进行研究”，这种前置性伦理评估能有效避免技术应用中的价值冲突。

责任归属机制的优化需要法律规范与技术手段的协同创新。建议引入区块链技术建立翻译过程的可追溯系统，通过不可篡改的日志记录明确机器输出与人工修改的贡献比例。在制度层面，应制定人机协作责任矩阵，根据任务复杂度划分不同模式下的责任主体：对于常规性翻译，以算法开发者承担主要责任；对于专业领域翻译，则要求终端用户履行结果核验义务；在“机器翻译+译后编辑”模式下，需按照编辑介入程度动态调整责任分配权重。这种精细化的责任划分既能保障用户权益，又能促进技术提供方的质量改进。

教育领域的翻译伦理优化具有特殊要求。针对机器翻译在语言学习中的应用，应当建立“有限使用”原则，通过元数据标注区分适合学生直接使用的安全内容与需要教师指导的敏感材料。开发具备伦理提示功能的智能辅助系统，当检测到可能引发文化误解或情感偏差的翻译结果时，自动触发解释性注释。王高申提出的“构建适应时代需求的伦理道德框架”在此体现为教育场景的特殊适配，例如在文学翻译教学中，系统可对比展示不同文化视角下的译法选择，培养学生的伦理判断能力。

行业协同治理构成伦理原则落地的制度保障。建议成立跨学科的翻译伦理委员会，由语言学家、技术专家、法律人士和伦理学者共同制定动态更新的行业标准。建立第三方伦理认证机制，对翻译平台的数据处理流程、文化包容性表现和社会影响进行定期审计。同时，开发开放性的伦理问题报告平台，鼓励用户和译者反馈算法偏见案例，形成社会监督与技术改进的良性互动。这些策略的实施需要突破传统技术优化的单一视角，在技术创新、制度规范和文化调适三个维度形成合力，最终实现人工智能翻译工具从“能用”到“善用”的伦理跃升。

4 研究结论与未来展望

本研究的系统分析表明，人工智能英语翻译面临的伦理挑战主要呈现三个维度的特征：文化偏见的内生性、隐私风险的隐蔽性以及责任界定的复杂性。研

究发现，当前主流翻译算法在文化敏感信息处理上存在结构性缺陷，其训练数据的不平衡分布导致低频语种和文化特定表达被系统性边缘化。同时，神经机器翻译的数据采集机制与用户隐私保护之间尚未建立有效平衡，现有脱敏技术难以应对专业场景中的敏感信息识别需求。更关键的是，人机协作模式下的责任划分缺乏明确的法律依据和技术支持，这使得翻译错误引发的纠纷陷入追责困境。

针对这些核心问题，研究提出的优化路径强调技术改进与制度创新的协同。在文化包容性方面，建立

动态平衡的多语种语料库和引入文化敏感度分级机制被证实能显著降低算法偏见；隐私保护领域，联邦学习与差分隐私技术的结合应用为数据安全提供了新思路；而基于区块链的可追溯系统和责任矩阵模型则为解决责任归属难题提供了可行方案。这些措施的有效性已在教育、医疗等典型应用场景中得到初步验证，但其大规模推广仍需克服技术成本与行业标准化的障碍^[5]。

参考文献

- [1] 杨婷婷. 司法视角下人工智能生成物著作权归属问题研究[J]. 时代人物, 2025, (7): 0124-0126.
- [2] 孟丽君. 人工智能在康复领域研究应用的伦理审视[J]. 中国医学伦理学, 2025, (2): 166-172.
- [3] 钟鸣. 人工智能时代学术英语写作新视角概述[J]. 海外英语, 2025, (4): 115-117.
- [4] 肖志清. 人工智能时代翻译技术应用伦理研究: 概念、问题与对策[J]. 民族翻译, 2024, (6): 18-25.
- [5] 杨扬. 生成式人工智能背景下的二语写作研究: 现状、问题与展望[J]. 语言与文化研究, 2025, (2): 135-138.