

生成式AI背景下的科研伦理问题探究

张 凯, 闫其涛, 张 睿, 郭 娇, 王嘉莹, 徐琳君, 李梅玲[†]

上海市农业科学院农业科技信息研究所, 201403, 上海

摘 要 近年来,随着生成式人工智能(AI)技术的飞速发展,其在科研领域的应用日益广泛。生成式AI通过深度学习、自然语言处理等先进技术,能够高效生成高质量的数据、文本、图像等内容,极大提升了科研效率。然而,该技术的普及带来了诸多科研伦理问题,如数据隐私泄露、知识产权争议以及人类监督缺失等。针对这些问题,本文提出了具体的解决对策与建议,如完善相关法律法规与行业规范,提升科研人员伦理意识,并在技术层面推动算法透明化与可解释性研究,旨在为科研工作者、政策制定者及相关利益方提供理论指导和实践参考,从而促进生成式AI技术在科研领域的健康发展。

关键词 生成式AI;科研伦理;数据隐私;知识产权;人类监督
Research on scientific ethics in the era of generative artificial intelligence//ZHANG Kai, YAN Qitao, ZHANG Rui, GUO Jiao, WANG Jiaxuan, XU Linjun, LI Meiling

近年来,随着生成式人工智能(generative AI)技术的飞速发展,其在科研领域的应用呈现出前所未有的广阔前景。生成式AI技术,尤其是深度学习和自然语言处理领域的突破性进展,使得该技术在科研工作中已具备强大的数据分析、模型构建和预测能力。

生成式AI在科研领域的应用体现在诸多方面。在数据分析领域,生成式AI能够高效处理海量科研数据,通过自动化方式生成高质量的数据报告。在科研模型构建方面,生成式AI技术能够根据已有数据自动构建预测模型,帮助研究人员更好地理解复杂现象。以气候科学领域为例,生成式AI技术已被广泛应用于气候预测模型的开发,显著提高了气候预测的准确性^[1]。生成式AI在科研文献整理和知识图谱构建方面也展现了卓越的优势,即能够通过对大量科研文献的分析,自动生成摘要、提取关键信息并构建知识图谱,为科研人员提供了高效的信息检索工具^[2]。在科研创新方面,生成式AI技术能够助力科研人员快速设计试验方案、优化试验参数并预测试验结果^[3]。这种高效的创新模式不仅加快了科研成果的产出,而且为科研领域的创新发展带来了更多的可能性。生成式AI技术正在深刻改变着科研领域的运作方式,其发展潜力不容忽视。

尽管生成式AI技术在科研领域展现出了巨大的潜力,但其也面临着诸多挑战。例如,生成式AI技术

在数据质量和算法透明度方面的不足可能导致科研结果的可靠性受到质疑。生成式AI技术的广泛应用可能带来数据隐私和安全隐患,尤其在涉及敏感数据的科研工作中。因此,如何在保证科研效率的同时,确保生成式AI技术的可靠性和安全性,成为目前科研领域亟待解决的问题。

1 生成式AI带来的科研伦理挑战

1.1 数据隐私与安全问题

生成式人工智能作为一项前沿技术,其强大的数据处理能力和生成能力极大地改变了科学研究的模式。生成式AI通过对大量数据的训练,能够模拟人类的创造力,生成高度逼真的文本、图像、音频甚至视频内容。然而,该技术的广泛应用也随之伴随着数据隐私与安全问题的加剧^[4]。

为了生成高质量的内容,生成式AI需要访问大量的训练数据,这些数据可能涵盖个人的身份信息、健康记录、地理位置信息等敏感内容^[5]。但是这种数据的采集通常缺乏用户的明确授权,势必导致数据滥用的风险增加。生成式AI的训练过程通常采用深度学习算法,这些算法需要对大规模数据集进行反复迭代和优化。在这一过程中,数据可能会被无意间存储在模型的权重中,从而形成所谓的“记忆效应”。这意味着即使数据在表面上被删除,仍可能通过模型的生成结果间接泄露。即生成式AI的数据处理方式也可能进一步加剧隐私泄露的可能性。此外,生成式AI可以通过分析已有的数据生成看似真实的虚假内容,这些内容可能包含虚构的个人信息或伪造的事件记录。例如,DeepSeek等大语言模型可以生成高度逼真的假新闻或虚假社交媒体帖子,这些内容可能误导公众并对个人或组织造成负面影响^[6]。

生成式AI对数据隐私的影响涵盖多方面。从数据采集到数据处理,再到数据共享和生成,每一步都可能成为隐私泄露的潜在途径。目前在实际生活中,已经有大量案件产生并被起诉至法院^[7]。

1.2 知识产权与原创性争议

1.2.1 AI生成物原创性的争议

随着生成式AI技术的迅速发展,其在科研领域的广泛应用给传统意义上的原创性认定带来了前所未有的

[†] 通信作者

的挑战^[8]。这些挑战主要体现在以下几个方面。

(1) 生成式AI生成的内容通常具有高度的复杂性和多样性,这使得传统的原创性评估标准难以适应新的需求。例如,在文本生成领域,生成式AI能够快速生成大量高质量的学术论文摘要或综述^[9]。然而,这种高水平的生成能力导致了学术界对于生成内容是否具备真正原创性的质疑。

(2) 生成式AI在处理数据时所依赖的大规模训练集增加了原创性认定的难度。生成式AI通常通过学习大量的公开数据来提升自身的生成能力,而这些数据可能包含了其他研究人员多年积累的知识成果。当生成式AI利用这些数据生成新内容时,如何界定其中哪些部分属于原始创新、哪些部分仅仅是已有知识的重新组合成了一个亟待解决的问题。例如,在医学领域,生成式AI可能会基于过去几十年发表的海量文献来生成新的诊疗方案建议。在这种情况下,如果最终提出的方案与已知的最佳实践非常接近甚至完全一致,那么该方案是否应该被视为原创将变得模糊不清。

(3) 生成式AI的自主学习特性进一步加剧了原创性认定的困难。随着模型训练过程的不断优化,生成式AI逐渐具备了一定程度上的自我调整和优化能力,这使得其能够在没有人为干预的情况下持续改进自己的生成效果。然而,这种自主学习的过程往往缺乏清晰的记录机制,导致研究人员很难追踪到某个特定生成结果的具体形成路径。这种情况不仅增加了判断一个生成内容是否符合原创性标准的工作量,还可能导致一些真正具有创新价值的成果被错误地归类为非原创。

1.2.2 AI生成物的知识产权归属争议

生成式人工智能作为一项前沿技术,其生成的内容是否能够享有知识产权保护是一个备受关注的课题。在传统的知识产权法律框架下,作品的创作主体通常被限定为自然人或法人,而生成式AI作为一种机器智能,并不具备法律意义上的主体资格。因此,AI生成内容的知识产权归属问题成为一个复杂的议题,且其涉及多个层面。在国际上,一些国家已经尝试通过立法或司法实践对此进行规范。例如,在美国,根据《版权法》的相关规定,只有由人类创作的作品才能获得版权保护。这一原则在 *Thaler v. Perlmutter* 一案 [*Thaler v. Perlmutter*, Case 1:22-cv-01564-BAH (D. D. C., Aug. 18, 2023)] 中得到了体现,即版权法保护的是人类智力劳动的成果,而非机器的输出结果。然而,随着生成式AI技术的进步,越来越多的研究表明,AI生成内容的质量和复杂性已经接近甚至超越了人类创作者的水平,这使得知识产权归属问题变得更加

突出。另外AI生成的视频是否侵犯他人权利,如“林黛玉倒拔垂杨柳”“孙悟空大战诸葛亮”一系列经典著作经过AI的生成,似乎变得不那么“正经”,这也引发AI创作的内容是自由创作还是侵权的讨论^[10]。

关于AI生成内容的知识产权归属问题,目前我国司法实践中也存在不同的观点。在北京互联网法院审理的AIGC图片可版权性第一案[李某与刘某侵害著作权纠纷一案,北京互联网法院(2023)京0491民初11279号民事判决书]中,法院认为涉案图片系原告利用AI生成,根据《著作权法》关于作品的构成要件进行判断,因涉案图片体现出原告的独创性智力投入,被认定为作品,相关著作权归属于原告。同时本案判决强调,利用人工智能生成的内容,是否构成作品,需要个案判断,不能一概而论。在武汉东湖法院审理的王某案[王某与武汉某科技有限公司著作权权属、侵权纠纷案,武汉东湖新技术开发区人民法院(2024)鄂0192知民初968号民事判决书]中,用户通过调整关键词、参数和多次筛选生成图片,法院认为其“构思、审美选择和个性化表达”满足独创性要求,认定作品受保护。在张家港市人民法院审理的AI文生图一案[丰某诉东山公司侵害著作权纠纷一案,张家港市人民法院(2024)苏0582民初9015号民事判决书]中,法院认为简单的提示词并非作品。著作权保护的落脚点在于具体表达而非抽象思想。提示词输入相对于生成式人工智能产生的内容只是思想,并非受《著作权法》保护的表达,因此该行为并不构成侵权。

值得注意的是,AI生成内容的知识产权归属问题不仅局限于法律层面,还涉及道德和社会层面的考量。例如,当AI生成的内容被用于商业用途时,如何确保公平竞争环境,避免因知识产权归属不明而导致的市场混乱;当AI生成的内容被用于学术研究时,如何保障研究成果的真实性和可信度等问题,均需要深入思考。

AI生成内容的知识产权归属是一个多维度、多层次的问题,涉及法律、技术、经济和社会等多个领域。尽管目前尚无统一的答案,但可以预见的是,随着生成式AI技术的不断发展,将会不断产生新的问题。

1.3 AI决策与生成内容的伦理困境

生成式AI在科研领域的广泛应用带来了前所未有的机遇,但同时也引发了诸多伦理困境,主要体现在决策透明性不足、责任归属模糊以及伦理监督缺失等方面。

AI决策过程透明性不足与缺乏伦理监督。在科研过程中,生成式AI往往通过复杂的算法模型进行数据分析和预测,然而其内部运作机制通常表现为一种

“黑箱”状态。这种不可解释性使得科研人员难以理解AI如何得出特定结论,从而增加了科研决策的风险^[11]。此外,由于AI系统的自主性和复杂性,现有的伦理审查机制往往难以有效覆盖整个决策流程。传统的人类伦理监督模式依赖于专家评审和同行评议,但在面对高度自动化和动态化的AI决策时却显得力不从心。

AI决策过程中的责任归属问题难以厘清。在传统的学术科研中,研究人员对其工作成果承担完全的责任,而生成式AI的介入则模糊了这一界限。一旦AI生成的结论被采纳并用于实际操作,出现失误时应由谁承担责任成为一个亟待解决的问题。目前,国际上尚缺乏明确的法律框架来界定生成式AI在科研中的法律责任。

AI决策内容缺乏人类监督。在一些案例中,研究人员利用生成式AI生成虚假的研究成果^[12],试图通过伪造数据来获取学术声誉或经济利益。这种行为不仅违背了科学研究的基本原则,而且对整个学术界造成不良影响。为了避免这种情况进一步恶化,有必要建立健全的人类监督体系,定期审查AI生成内容的真实性 and 可靠性。

2 生成式AI科研伦理问题的解决对策与建议

尽管生成式AI为科研工作提供了强大助力,但其同时也带来了诸多伦理问题。面对这些难题,我们除了依靠技术本身去解决外,还需结合人类智慧与经验,构建一个多层次、全方位的监督框架。只有这样,才能真正实现AI技术的安全可控发展,促进科学研究向着更加健康的方向迈进。

2.1 完善内部风险控制制度

生成式AI在科研领域广泛应用的同时,也带来了显著的数据安全风险。数据泄露问题成为一大隐患,其可能涉及敏感的个人身份信息、商业机密或国家重要科研数据。数据篡改是生成式AI背景下的另一大安全隐患。随着生成式AI技术的发展,恶意攻击者可以通过操纵训练数据集来改变AI模型的行为模式,进而影响最终的科研结果。这种数据篡改不仅会影响科研工作的准确性,还可能导致错误的决策制定,从而对社会造成不可估量的危害。为了应对上述数据安全风险,科研机构 and 相关部门应采取一系列综合性的措施。

(1) 加强数据加密技术的应用至关重要。现代加密算法,如AES(高级加密标准)和RSA(公钥加密技术)能够有效保护数据在传输和存储过程中的安全性。例如,采用AES-256位加密算法可以将数据的破解难度提升到前所未有的高度,使得即使数据被截获,攻击

者仍难以解密^[13]。科研机构应定期更新加密算法,并结合量子计算技术的发展趋势,探索更先进的加密方案,以应对未来可能出现的安全威胁。

(2) 建立完善的数据访问控制机制。通过实施严格的权限管理和身份认证制度,可以确保只有经过授权的用户才能访问特定的数据资源。同时,科研机构还可以引入区块链技术,利用其分布式账本的特点,实现数据的透明记录和不可篡改性,进一步增强数据的安全性。建立健全的数据备份和恢复机制对于防范数据丢失具有重要意义。科研机构应当制定详细的数据备份计划,定期将关键数据备份至多个地理位置不同的存储设备上。一旦发生意外事故,如自然灾害或硬件故障,也能迅速恢复受损的数据。

(3) 提升科研人员的伦理意识。随着生成式AI技术在科研领域的广泛应用,科研人员对于科研伦理的理解和掌握显得尤为重要。科研伦理教育应贯穿科研人员的职业生涯全过程。研究表明,大学生对于生成式人工智能在学习与研究辅助中的认同度较高,其已成为大学生学习与研究辅助的重要帮手^[12]。但这个过程会产生科研诚信问题,如剽窃他人学术成果、伪造虚假数据等^[14]。因此,从本科生阶段到博士生阶段,再到科研工作的实际操作环节,均需要有系统的伦理教育支持,即在科研诚信建设方面,塑造科研伦理价值观至关重要^[15]。特别是在研究生阶段,科研伦理教育应作为必修课的一部分。通过系统的教育培训和科学合理的课程设置,科研人员的伦理意识能够得到显著提升。这种提升不仅有助于解决生成式AI背景下的科研伦理问题,还能促进整个科研生态的健康发展。在未来,科研机构应进一步加大投入力度,完善科研伦理教育体系,为科研人员提供更加全面和深入的伦理指导。

因此,面对生成式AI带来的数据安全风险,科研机构需要从技术、管理和教育等多个维度入手,采取全面而系统的应对措施。只有这样,才能在享受生成式AI带来的便利的同时,最大限度地降低数据安全风险,为科研事业的健康发展提供坚实保障。

2.2 制定并完善相关法律法规

从法律框架的角度来看,当前许多国家和地区已经意识到数据保护的重要性,并制定了一系列法律法规。例如,欧盟早在2018年就出台了《通用数据保护条例》(GDPR),该条例为数据主体提供了强有力的保护,规定了数据收集、存储、传输和使用的详细规则。根据GDPR的规定,数据主体拥有知情权、访问权、更正权、删除权等多项权利。这些权利的设定为科研活动中数据的合法使用提供了明确的指导原则。然而,

尽管GDPR具有较高的权威性和影响力,但其适用范围主要限于欧盟成员国,对于非欧盟国家的科研机构而言,仍缺乏足够的约束力。因此,有必要在全球范围内推广类似的高标准数据保护法规。

(1) 政策法规的制定要结合生成式AI技术的特点更加细化和具有针对性。生成式AI通常依赖于大规模的数据集进行训练,而这些数据往往包含了敏感信息,如个人身份信息、医疗记录、财务数据等。因此,数据保护法规应明确规定生成式AI在数据收集、处理和使用过程中必须遵循的原则。例如,要求数据匿名化处理、限制数据共享范围、设立数据使用审计机制等。此外,应建立数据泄露应急响应机制,确保一旦发生数据泄露事件,能够迅速采取措施减少损害。

(2) 为了保障数据保护法规的有效实施,还需建立健全的监管体系。监管机构应具备独立性和权威性,能够对科研机构的数据使用行为进行有效监督。例如,美国联邦贸易委员会(FTC)在数据保护方面发挥了重要作用,其有权调查和处罚违反数据保护法规的行为。同时,监管机构应定期发布数据保护指南,为科研机构提供具体的合规指引。

(3) 相关部门在制定生成式AI伦理准则时需关注知识产权与原创性的问题。即生成式AI伦理准则应提出明确的知识产权框架,包括界定AI生成内容的权利归属、设定人类创作者的责任范围以及鼓励开放共享的创新模式^[16]。生成式AI伦理准则应强调人类在决策过程中的主导地位,要求科研人员在使用生成式AI工具时必须保持高度警惕,并定期审查AI生成结果的有效性和合理性。同时,应建立多层次的监督体系,包括内部审计、外部审查以及公众参与等多种形式,以确保生成式AI的应用始终符合社会价值取向。

加强数据保护法规建设是解决生成式AI科研伦理问题的关键环节之一。通过完善法律框架、细化法规要求、强化监管力度和推动国际合作,可以有效提升数据保护水平,为科研活动创造更加安全和可信的环境^[17]。而且该生成式AI伦理准则的制定应遵循动态调整的原则,及时响应新技术发展的新需求。由于生成式AI技术正处于快速迭代阶段,其应用场景和技术特性也在不断变化。因此,伦理准则不应是一成不变的静态文件,而应是一个开放且灵活的框架,允许根据实际情况适时更新和完善。例如,可以设立专门的伦理委员会,负责跟踪技术发展趋势并对现有准则提出修改建议;此外,可以通过举办研讨会、公开听证会等形式广泛征求社会各界的意见,增强准则的包容性和适应性。

2.3 提高生成式AI算法的透明度

生成式AI技术算法的透明度是科研伦理极为关注的问题。透明度是生成式AI系统中至关重要的特性,其直接关系到科研人员能否理解模型的工作方式,并对其结果进行验证和复现。透明度不足可能导致科研过程中出现不可预测的结果,甚至引发严重的伦理争议。因此,提高生成式AI算法的透明度不仅能够增强科研活动的可信度,而且能够促进科研领域的健康发展。

构建可解释的人工智能系统来提升生成式AI技术算法的透明度。可解释性是指模型能够清晰地表达其决策过程的能力,使用户能够理解模型如何得出特定结论。例如,在生成式AI系统中,通过可视化工具展示模型的训练数据分布、参数选择依据以及推理路径,可以帮助研究人员更好地理解模型的行为模式。此外,应引入注意力机制(Attention Mechanism),让模型在处理输入数据时明确标注哪些部分对最终输出贡献最大,从而提高模型的透明度^[18]。

为了进一步提升透明度,应加强生成式AI算法设计阶段的文档记录。完整的文档记录能够详细描述模型的开发流程、所用数据集的特点、训练过程中的超参数调整等关键信息。这不仅有助于科研人员在后续研究中复现结果,而且能为其他研究者提供宝贵的参考。例如,Google推出BERT模型时,提供了详尽的技术文档,包括模型架构设计、预训练步骤、评估指标等内容,使得全球范围内的研究者均能够基于该模型开展新的研究工作。

开放源代码也能提高生成式AI算法透明度。开放源代码意味着任何人都可以访问模型的原始代码,并对其进行修改和优化。这种方式不仅能够加快技术创新的步伐,而且能有效避免因闭源而导致的信息不对称现象。例如,Facebook开源了其Transformer系列模型,包括RoBERTa、XLM-R等,受到了学术界的高度评价。

3 结束语

生成式人工智能(AI)技术飞速发展,其在科研领域应用广泛,极大提升了科研效率。然而,这种技术的普及带来了诸多科研伦理问题,如数据隐私泄露、知识产权争议以及人类监督缺失等。面对新技术,我们应秉持积极的态度使其为人类服务,在这个过程中,面对生成式AI技术带来的挑战,本文提出了具体的解决对策与建议,如完善相关法律法规与行业规范,提升科研人员伦理意识,并在技术层面推动算法透明化与可解释性研究,从而促进生成式AI技术在科研领域的健康

发展,为人类科研做出更大的贡献。

4 参考文献

- [1] 国家气候中心:人工智能助力气候预测新范式[EB/OL]. (2024-10-10) [2025-06-08]. https://www.cma.gov.cn/ztbd/2024zt/20241009/2024100910/202410/t20241010_6630587.html
- [2] 范睿祺,许梦鑫,文国兴.大学生应用生成式AI的现状、风险与对策研究:以河南省为例[J].信息与电脑,2025,37(10):108
- [3] 马银琦,黄恒,毋磊,等.“技术赋能”还是“工具依赖”:生成式人工智能对研究生科研创新力的影响研究[J].电化教育研究,2024(12):58
- [4] 张爽.金融领域生成式AI技术的应用、风险与应对[J].债券,2025(5):69
- [5] 曾铁英,吴梅利洋,胡斌春,等.护理科研伦理困境和负责任-合伦理研究行为的思考[J].护理与康复,2025,24(3):1
- [6] 丁大琴,徐宇飞.人工智能技术应用的科研伦理风险及其对策[J].未来与发展,2025,49(4):102
- [7] ChatGPT又惹官司:两作家起诉OpenAI,称其侵犯版权[EB/OL]. (2023-07-04) [2025-06-10]. https://www.thepaper.cn/newsDetail_forward_23722065
- [8] 马弘.面向生成式人工智能(AIGC)的学术诚信教育转型策略研究[J].山东图书馆学刊,2025(2):85
- [9] 安彤,薛德军,周哲,等.基于多维特征分析的AI生成期刊论文内容识别与应对策略研究[J].知识管理论
- 坛,2025,10(2):99
- [10] 赵丽,殷增梓.AI“魔改”:创作自由还是侵权行为?[N].法治日报,2025-04-26(4)
- [11] 罗仟合.伦理法视域下医用人工智能的治理研究[J/OL].云南大学学报(社会科学版),2025,1[2025-06-07]. <https://doi.org/10.19833/j.cnki.jyu.20250605.002>
- [12] 刘若辰,李建霞,侯彪.生成式AI对高校教育教学的影响与思考[J].高教学刊,2025,11(15):7
- [13] 李雪芹.AI生成营销内容对消费者趋近/回避行为的影响机制与应对策略[J].哈尔滨师范大学社会科学学报,2025,16(3):104
- [14] 沈雅晨.生成式人工智能应用于高校学术写作对学术诚信建设的影响研究[J].南方论刊,2025(5):76
- [15] 庞娜.日本高校教师科研诚信体系建设及对我国的启示[J/OL].大连理工大学学报(社会科学版),2025,1[2025-06-07]. <https://doi.org/10.19525/j.issn1008-407x.2025.03.013>
- [16] 何星辰.DeepSeek中AI生成内容属性认定及著作权归属问题研究[J].中国品牌与防伪,2025(4):31
- [17] 张爱军,陈瑞琪.DeepSeek等生成式人工智能赋能政治传播的伦理风险与规制路径[J].公共治理研究,2025,37(3):5
- [18] 李伟棠,AI人工智能翻译中基于注意力机制和联合训练的QE任务分析[J].自动化与仪器仪表,2024(9):258

(2025-06-15收稿;2025-10-27修回)

(上接第51页)

- [6] 柳斌杰.完善中国特色学术期刊评价体系[J].青年记者,2022(10):4
- [7] 刘普,孙婉婷,梁鑫渊.社科学术期刊编辑人员的职业认知及其改善研究:基于104名社科学术期刊编辑人员职业自述的文本分析[J].上海师范大学学报(哲学社会科学版),2025,54(2):141
- [8] 李慧,苏莉娜,张郁.基于CiteSpace分析的我国科技期刊数字化转型研究的现状及趋势[J].传播与版权,2024(22):17
- [9] 汪汇源,赵云龙,欧梨成,等.我国农业科技期刊党建工作与业务工作深度融合发展策略[J].编辑学报,2023,35(2):135
- [10] 王元杰,薛春璐,齐秀丽,等.中文期刊开放获取转型后可持续发展策略[J].编辑学报,2023,35(增刊1):84
- [11] 张志强.出版学科专业共建:重要意义、核心问题与目标指向[J].中国出版,2024(3):10
- [12] 金鑫荣.加快建设一流出版学科的时代要求与实践探索[J].中国编辑,2023(7):56
- [13] 邵郭华.新质生产力赋能图书馆服务创新研究:以“易本书”平台为例[J].江苏科技信息,2024,41(19):88
- [14] 朱瑞琦,樊艳平,安娜.学术期刊青年编辑系统化人才培养体系的构建[J].科技传播,2024,16(21):63
- [15] 杨晋.基于岗位胜任力模型的复合型出版人才培养模式[J].出版广角,2023(21):20

(2025-06-12收稿;2025-10-27修回)