

应用人工智能辅助科研的诚信困境与应对

王立元

江西中医药大学中医药与大健康发展研究院 江西南昌

【摘要】人工智能应用于科研辅助工作已经成为较普遍现象，同时也引发全新的学术诚信问题。应对 AI 辅助科研的诚信困境，需要解决学术主体与责任、科研成果可靠性与原创性、群体差异与学术公平失衡、传统治理适配不足等方面的问题。根据国内外现状与研究进展，从制度监管、技术迭代、生态影响与规范认知等角度探究原因，并提出了相应的应对路径，分别从制度、技术、教育和人文层面有效监管缓解压力，为维护健康学术生态、平衡 AI 创新赋能与学术诚信底线提供理论与实践支撑。

【关键词】人工智能；科研；学术诚信；治理路径

【基金项目】江西中医药大学学位与研究生教育教学改革研究项目（jzyjg-2023-12）；江西中医药大学自选项目（20203BAAW208025）

【收稿日期】2025 年 11 月 16 日 **【出刊日期】**2025 年 12 月 17 日 **【DOI】**10.12208/j.sdr.20250297

Integrity dilemmas and countermeasures of applying artificial intelligence to assist scientific research

Liyuan Wang

Institute of Traditional Chinese Medicine and Health Development, Jiangxi University of Chinese Medicine, Nanchang, Jiangxi

【Abstract】 Becoming a common phenomenon, the application of artificial intelligence in auxiliary scientific research work, at the same time, has triggered new academic integrity issues. To address the integrity dilemmas of AI-assisted scientific research, it is necessary to solve some problems such as academic subject and responsibility, reliability and originality of scientific research results, group differences and imbalance of academic fairness, and insufficient adaptation of traditional governance. Based on the current situation and research progress at home and abroad, this paper explores the causes from the perspectives of institutional supervision, technological iteration, ecological impact and normative cognition, and puts forward corresponding countermeasures. It could effectively regulate and alleviate pressures from the institutional, technological, educational and humanistic levels, providing theoretical and practical support for maintaining a healthy academic ecology and balancing AI innovation empowerment with the bottom line of academic integrity.

【Keywords】 Artificial intelligence; Scientific research; Academic integrity; Governance pathway

人工智能（Artificial Intelligence，简称 AI）以 ChatGPT、豆包、腾讯元宝、DeepSeek 等为代表，凭借其强大的文本生成、数据整合与虚拟仿真能力，可以深度嵌入科研工作的全流程。目前，从文献综述、实验设计到成果撰写，AI 的使用显著提升了研究效率，降低了创新门槛，同时也带来了新的诚信问题，甚至形成了系统性学术诚信危机：AI 幻觉导

致数据虚构，隐性剽窃引发关于原创性的争议，责任界定模糊使得追责无能，群体差异诱发公平失衡，等等^[1,2]。AI 辅助科研问题频发，传统的科研诚信面临 AI 引发的问题与挑战。如果不合理应对和有效治理，不仅会侵蚀学术的真实性，也会引发知识创造行为的扭曲，威胁健康的学术生态。

1 监管现状

从国外机构应对 AI 辅助科研的诚信治理方式来看,美国的大学大致有三类应对方式^[3]。第一类是持反对态度,出于对隐私泄露和学术剽窃等风险的担心,严格加强风险防范并优化管理措施;第二类是持支持立场,依托技术优势主动为师生提供 AI 技术相关的资源并开展培训支持,充分发挥在信息汇总、跨语言转换等方面的效能;还有些高校则立场折中,通过设计风险分级管理制度,构建兼顾支持与约束的管理框架。欧盟则是于 2021 年通过了《人工智能法案》,对算法透明性及数据可追溯性作出要求,强调要保护欧盟公民的基本权利、健康与安全免受人工智能技术的威胁。

我国则聚焦实践层面问题进行探索。国家网信办等七部委联合出台《生成式人工智能服务管理暂行办法》,设立技术应用的底线要求;部分高校明确 AI 工具的使用限制,如复旦大学出台的《复旦大学关于在本科毕业论文(设计)中使用 AI 工具的规定(试行)》被称作“全球最严 AI 新规”;大部分期刊杂志对使用 AI 写作说不,要求投稿论文必须通过 AIGC 检测。学者们还围绕学术研究、高等教育等工作,探讨了 AI 辅助使用引发的思维惰化、形式主义强化等问题。但是,现有研究对 AI 使用的风险还停留在单一维度,缺乏对诚信困境的系统性分析^[4-6],对引发的诚信治理需求关注较少,难以应对技术变革对科研诚信治理的新需求。

2 面临困境

2.1 学术主体与责任边界困境

学术主体地位弱化和责任链断裂是出现诚信困境的核心原因。一方面,尽管“影子作者”现象日益凸显,但 AI 不具备法律人格,不能作为学术成果的合法作者出现。有研究显示,19%的高影响因子期刊论文存在 AI 生成内容且并未进行声明,用 AI 进行创作的贡献与风险问题普遍存在,现有署名方式不能够界定人类研究者与 AI 工具的权责划分。另一方面,存在着认知外包现象且有加剧趋势,研究者的主体责任不断弱化,部分科研人员甚至依赖 AI 工具来进行核心研究。一旦成果出现问题导致严重后果,显然 AI 作为工具属无责,使用者则将责任推给 AI 工具也可能免责,最终无人承担后果。更糟糕的是,大部分人对使用 AI 的认知是很模糊的,大多数研究者仅能认识到成果代写属于违规,但对 AI 辅助提纲设计、语句润色等场景的是否合规缺乏正确判断,

很多人认为使用 AI 仅是使用普通工具,而忽视了生成内容的原创性与可靠性要求。

2.2 科研成果可靠性与原创性困境

科研成果的真实性与独特性受到双重冲击。可靠性方面,已经暴露出多种 AI 幻觉引发数据虚构的风险,2024 年《Nature》杂志刊文证实 GenAI 至少存在 15%的数据虚构概率^[7],其生成的虚假文献、杜撰案例和实验数据伪造事件已经污染了现有的学术知识体系,导致相当部分的行业资深专家都被 AI 生成的数据与成果所误导。同时, AI 的深度润色会导致研究成果华而不实,特别是那些生成的文本,表面上逻辑连贯、术语规范、描述与形容贴切,实际上却是基于概率统计生成的缺乏实证和思想深度的文字集合,根本不属于学术研究范畴。原创性和反剽窃方面,使用 AI 工具进行洗稿、文字重组等技术能够逃避传统的文本相似度检测手段,实现窃思想观点但不偷具体文字,传统检测手段对以 AI 进行的剽窃成果的识别率不足 40%。使用 AI 还存在知识同质化的风险, AI 工具生成的文本内容实质上是基于概率统计对现有知识进行重组,以 AI 为主体生成科研成果会陷入学术的舒适区,从源头固化和扼杀学术的创新活力。此外,现有检测手段多聚焦于检测文本内容,对 AI 生成的图像、代码、实验数据等非文本内容检测效果较差,进一步降低了对科研成果可靠性的监管。

2.3 群体差异化与学术公平困境

不同学科、不同群体的研究者面临的诚信风险存在差异,可能会加剧学术不公的产生。从学科维度来看,农林类、医学等学科的科研活动一般周期偏长,数据采集过程受到自然环境的制约较大,有匿名访谈发现,相当部分的农林类研究生在工作任务繁忙时考虑过或者已经使用过用 AI 工具代劳,这些情况一旦发生就可能造成伪造数据、编造成果等严重问题。人文社科领域则更易出现使用 AI 生成内容来代替实际研究。从不同科研群体的视角来看, AI 工具的质量水平可能会形成基于数字技术的鸿沟:经济条件好的研究者能够使用高端的 AI 工具和定制化服务,拥有某种意义上的技术特权;而部分条件匮乏的科研人员则缺乏经济能力和 AI 工具使用权限,在与前者的竞争中处于劣势地位。这种数字技术鸿沟,从广义上会导致劣币驱逐良币,让学术研究行为背离其真正价值追求,而异化成使用技术

工具的比拼。

2.4 传统治理体系适配性困境

传统科研诚信治理体系在 AI 技术的快速迭代变革面前呈现明显的不适应。在技术检测层面, AI 模型迭代速度远超检测工具的更新, DeepSeek-R1 等新模型的漏检率达 32%, 为应对 AIGC 检测手段甚至出现了反 AI 检测工具, 呈现矛与盾之间军备竞赛的奇怪局面。同时, 检测工具还存在严重的误判风险, 比如国内部分高校要求对成果进行 AIGC 检测, 朱自清的《荷塘月色》被判定 AI 生成内容疑似度高达 62.88%, 唐代诗人王勃的《滕王阁序》甚至被检出 AI 率接近 100%, 结果令人啼笑皆非。在制度规则体系层面, 不同国家对 AI 辅助科研的应对治理存在显著的立场差异, 欧盟要求严格监管, 美国基本实行各自独立的分级管理, 我国则是提出底线管理要求, 这些监管策略的显著差异使得跨国的科研合作面临 AI 监管标准衔接的错位。从我国的相关规定来看, 多数高校和机构尚未出台明确的 AI 使用规范, 缺乏监管使用 AI 辅助科研工作的实操性条款, 制度供给明显跟不上实践中的 AI 工具治理需求。在教育评价层面, 唯论文的量化评价变相“激励”研究者追求技术捷径, 忽视甚至主动去钻 AI 工具使用的监管漏洞。据统计, 42% 的高校没有或无力开展针对 AI 工具使用的科研伦理监管培训, 没有指导或监管 AI 技术的使用。

3 原因探究

3.1 AI 特性与检测技术的对抗失衡

AI 工具的技术特性成为学术不端行为的数字温床, 而检测手段的更新滞后加剧了治理的困难。AI 拟真性强, 能够很好的模拟学术写作规范格式与学术术语体系, 几乎不能通过表面特征来识别 AI 生成内容; 过程隐匿性则让知识生产的逻辑推演、数据处理等核心环节处于算法黑箱, 无法观察到 AI 工具是如何进行思考与推理, 传统的过程性监管手段失效; AI 工具还具有低门槛、易操作性, 这进一步降低了学术不端行为的实施成本, 只要输入指令、筛选结果, 就能生成一份看似严谨的研究结果, 而且常规手段难以辨别其真伪。与之相对的是, 现有监管检测手段多以文本相似度比对作为基本功能, 对语义内涵、思维逻辑与多模态内容的深度识别能力普遍缺乏, 难以应对 AI 引发的窃思想、伪原创等新型学术不端行为, 形成技术迭代与监管滞后的恶性

循环。

3.2 规则供给与执行的双重缺位

制度体系的滞后与执行乏力, 可能是导致诚信困境的关键因素。规则供给方面, 现有科研诚信制度的设计都默认是以人为主体, 面对人机协同、AI 主体的新型科研关系, 就出现了严重不适应。针对 AI 生成的内容, 其著作权归属、文献引用规范, 均缺乏具体监管细则, 如何区别无意违规与蓄意违规也缺乏操作标准, 行为背后的法律与道德监管都存在缺失。国际层面, 不同国家和地区间的监管逻辑可能并不一致, 欧盟侧重风险预防, 美国强调创新激励, 我国聚焦底线约束, 三者难以形成有机衔接, 造成监管过程中的错位与碎片化。执行层面, 部分高校一禁了之, 有些则放任自流, 还有些高校主动作为却有心无力, 真正能够针对性适度监管的案例少之又少。现有报道中, 部分高校与科研机构已经出台监管规范, 但是又无法提供针对性好、能够精准配套的技术支撑, 导致政策执行流于形式, 实际发挥效果欠佳。

3.3 学术评价与伦理共识的扭曲

学术生态的价值扭曲成为学术违规的滋生土壤。评价体系方面, 存留的唯论文、唯影响因子量化导向还未根本改变, 科研人员面临巨大的考核压力、晋升压力, 部分人追求短平快的研究成果, 使用 AI 工具成为规避深度思考、缩短研究周期的捷径, 至于探索知识、追求真理的行为只能往后排。伦理共识方面, 学术共同体还没形成统一的 AI 科研应用伦理准则, 如何监管 AI 工具的使用, 对生成内容的原创性、责任划分、隐私保护等核心问题尚未形成权威准则或默认做法, 部分人直接将技术便利等同于合理使用, 忽视了学术研究的诚信底线。此外, 训练 AI 工具的数据中可能潜藏有偏见与价值扭曲, 通过使用者的数据选择倾向, 输出的成果进一步放大了这种不平等与偏见, 造成学术价值与伦理认知的混乱。

3.4 主体素养与伦理认知的不足

科研人员的 AI 素养与伦理认知短板也是造成困境的内在根源。在素养层面, 多数研究者缺乏对 AI 技术原理、运行边界的系统性认知, 对于 AI 输出内容中的逻辑漏洞、数据偏差与偏见倾向不能有效识别, 造成研究体系过度依赖技术结果而丧失批判性思维。部分研究者甚至不能按照基本的 AI 工具

使用规范操作，操作不当陷入无意识的剽窃或者被动的数据污染。在伦理认知层面，技术变革对传统学术诚信观念造成巨大的冲击，部分研究者对“原创性”的理解仅限于文本层面，忽视了思想创新的核心价值。还有人将 AI 辅助与学术不端的边界认知模糊，将 AI 生成内容简单视为自身成果，弱化了学术劳动的责任意识。种种认知不足和缺失，导致研究者不但无法正确发挥 AI 工具的优势，还容易陷入 AI 使用带来的诚信困境。

4 治理路径

4.1 制度方面明确边界与全程管控

建立分层分类、全程覆盖的制度规范，有效破解 AI 使用引发的诚信困境。使用规范层面，制定 AI 辅助科研的分级分类详细标准，明确允许使用场景，如文献检索、格式优化、语言润色等，禁止用 AI 工具生成核心观点和编写实验数据等场景，要求 AI 生成的内容占比不超过 20%或更低比例。严格禁止 AI 工具代写论文甚至是生成学术成果，要求在研究成果中详细标注 AI 使用的细节，包括工具类型、应用环节、修改轨迹等具体信息。监管机制层面，尽快建立全流程管控体系，要求研究者事先提交 AI 使用计划书，明确 AI 的应用范围与预期作用；研究中主动接受智能工具检测和监管复核，对高风险学科、重点研究环节开展专项审查；实行分类追责制度，区分无意违规与蓄意造假，对前者以教育引导为主，对后者依法依规严肃处理。国际协同方面，寻找不同立场与原则下的最大公约数，推动建立跨国、跨区域的科研领域 AI 使用通用标准，破解协同合作中的合规困境。

4.2 技术层面精准检测与溯源并重

以技术对抗技术，加强对 AI 使用的多维度、全链条的监管。在检测技术方面，除了传统文本相似度的检测，还要开发出多模态识别的 AI 检测集成平台，整合自然语言处理技术识别语义价值剽窃，利用知识图谱技术追溯观点来源，应用图像指纹识别、代码逻辑比对等技术手段，实现对非文本成果的有效甄别^[8]。同时，建立动态更新的学术不端案例库，以 AI 监测 AI，为检测工具持续提供新案例新资料新数据，提升检测工具的监管能力。在溯源技术方面，推广区块链与数字水印技术，要求 AI 工具嵌入可追溯标记，对科研过程中的提示词、输出结果、修改记录等行为进行留痕。参考 Science 等部分期刊要

求上传原始科研记录的规定，建立科研数据与 AI 交互记录的分布式存储系统，确保研究过程可验证、可追溯，为责任界定提供数据支撑。

4.3 教育范畴提升素养与内化伦理

强化 AI 素养教育与伦理教育，从根本上提升研究者的诚信意识与风险防控能力。在分层培训方面，构建覆盖不同群体的教育体系，针对高校教师开展 AI 技术原理、科研应用规范与伦理审查能力培训，提升其指导学生合规使用 AI 的能力^[9]。针对研究生及青年研究者，开设 AI 伦理、批判性思维培养等培训课程，系统传授 AI 工具合规使用方法、AI 生成内容甄别技巧，强调学术诚信基本准则。针对不同学科的研究人员，开展针对性强的差异化培训，比如为农林类研究者重点讲解实验数据的 AI 辅助处理规范，为人文社科研究者强化价值导向与原创性思维训练。在伦理共识方面，通过学术研讨会、案例警示教育等形式，重塑 AI 时代的学术诚信观念，将 AI 工具使用诚信纳入师风、学风和科研道德范畴，明确 AI 价值定位于辅助研究的原则，强化研究者的责任意识与底线思维。推动学术共同体更新引文规范与原创性认定标准，总结形成常用的人机协同科研模式与案例。

4.4 人文视角缓解压力与弥合鸿沟

塑造更加包容的科研环境，消除群体差异带来的不公平，从源头减少诚信风险。纾解压力方面，优化科研评价体系，破除唯论文、唯影响因子的错误导向，增加研究过程、创新质量、实践价值等过程性考核权重，缓解对青年学者的考核压力。针对实验性学科领域，建立实验数据共享平台、缩短成果评价周期，提供专业的数据处理辅助服务，缓解实验探索的压力。在公平保障方面，加快推动 AI 工具资源普及，高校统一采购合规 AI 工具并向师生免费开放，为基层科研人员提供更先进的 AI 技术和培训支撑。开展针对性技能培训，重点提升弱势群体的 AI 工具使用能力与批判性思维，弥合数字鸿沟。在人文关怀方面，注重对学术人员的生活关怀与心理疏导，缓解研究人员的内生焦虑。构建以学术团体、行业协会为纽带的学术共同体，通过经验分享、同行互助等形式，营造求真务实的学术氛围。

5 未来展望

使用 AI 辅助科研的诚信问题，可以说是由其技术特点、制度滞后、生态扭曲与认知局限等多方面

因素交织而形成的结果，其核心矛盾在于技术创新与监管治理应对的适配失衡。尽管 AI 辅助科研仍属于新鲜事物，但其应用价值深远、潜在影响范围巨大，对现有科研工作的生态环境可能造成巨大冲击，甚至引发非传统的学术诚信困境。破解这些困境需从多维视角正确应对，构建制度、技术、教育、人文等层面的协同治理与应对措施，在正确发挥 AI 技术创新赋能作用的同时，坚守学术诚信底线，维护科学知识库的真实性与有效性。

结合 AI 工具的发展趋势与监管需求，未来关于 AI 辅助科研的研究可能有两个方向值得关注与聚焦，具体包括细分学科与场景的 AI 使用差异化监管，使用 AI 工具的法律与伦理体系研究，AI 工具使用的监管技术迭代优化等。应对 AI 辅助科研引发的诚信危机，要求科研工作者持续跟进探索，动态推动科研诚信治理体系不断完善，有效应对 AI 技术创新带来的学术诚信挑战。

参考文献

- [1] 吕品,李晓虹.人工智能时代研究者科研规范的现实挑战与实践进路[J].重庆开放大学学报,2025(3).
- [2] 张跃.生成式 AI 在研究生学术写作中的应用:优势、风险和纾解对策[J].呼伦贝尔学院学报,2025(4).
- [3] 田芬,卢春燕.美国顶尖研究型大学应对生成式人工智能的立场与策略[J].电化教育研究,2025(11).
- [4] 吴彦强,杨永甲.生成式人工智能背景下高校研究生科研诚信治理研究——以农林类高校为例[J].智慧农业导刊,2025(19).
- [5] 王艳超,闫欣彤,张涛.生成式人工智能对大学生学术诚信的影响机理及教育引导路径研究[J].高校辅导员,2025(3).
- [6] 王明华,林修杰,杨晓峰,等.生成式人工智能赋能高校教师科研创新研究[J].质量与市场,2025.
- [7] Pividori M .Chatbots in science: What can ChatGPT do for you?[J].Nature, 2024.
- [8] 冯永刚,吕鑫源.生成式人工智能嵌入教育的伦理检视[J].中国电化教育,2025(11).
- [9] 冯凯瑞,杜静.生成式人工智能时代研究生学术道德失范治理的新挑战与应对策略[J].黑龙江高教研究,2025(10).

版权声明: ©2025 作者与开放获取期刊研究中心(OAJRC)所有。本文章按照知识共享署名许可条款发表。

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



OPEN ACCESS