

• 管理纵横 •

DOI: 10.3724/BNSFC-2024.11.15.0002

生成式人工智能在科研项目管理中的风险分析 及应对措施

姚 畅^{1*} 宫 辰^{2*} 李广宇² 邓雨亭³ 万 升² 魏 鹏⁴ 黄俊杰⁴

1. 国家自然科学基金委员会 信息中心, 北京 100085
2. 南京理工大学 计算机科学与工程学院, 南京 210094
3. 南京理工大学 知识产权学院, 南京 210094
4. 南京理工大学 科学技术研究院, 南京 210094

[摘 要] 近年来,生成式人工智能的浪潮席卷全球,深刻地改变了多个行业领域的发展格局。然而,随着生成式人工智能技术的快速发展与广泛应用,其潜在的风险与挑战也逐渐显现,尤以科研项目管理领域为甚。为全面分析生成式人工智能对科研项目管理的影响和风险,本文对全球范围内的相关机构政策和专家观点进行了系统调研,并利用“动机—机会—能力”及“技术—组织—环境”理论框架,从人工智能技术、法律法规和伦理道德等多方面入手,深入分析了生成式人工智能在科研项目管理中的各类风险,包括隐私与安全风险、内容与决策风险、权属与侵权风险、道德与诚信风险、多元与创新风险、文化与伦理风险等。在此基础上,本文进一步从政策引导、管理规范与技术支撑三个层面提出应对策略,以确保科研项目管理工作的健康有序进行,推动生成式人工智能在科研项目管理领域的规范化应用。

[关键词] 生成式人工智能; 科研项目管理; 风险分析; 风险应对; “动机—机会—能力”框架; “技术—组织—环境”框架

1 背景与目的

近年来,以ChatGPT为代表的生成式人工智能(Generative AI)浪潮席卷全球,其强大的文本生成、图像合成、声音模拟及视频创作能力,深刻地改变了多个行业领域的发展轨迹。相较于传统人工智能,生成式人工智能是一种可让计算机自动创建新的内容和想法的人工智能,其在艺术创作、自然语言处理、文本翻译和代码编写等任务中获得了能够和人类智能相媲美的性能。目前,生成式人工智能给人们的工作和生活带来了很大便利,在优化业务流程、改善用户体验和提高工作效率等方面展现出独特的优势,激发了各行各业的广泛关注与应用探索。从艺术创意到金融投资,从教育培训到医疗健康,生成式人工智能正以前所未有的速度融入社会生活的方方面面。

然而,随着生成式人工智能技术的快速发展与广泛应用,其潜在的危害也逐渐显现。例如,生成式人工智能能够轻松创建与真实内容难以区分的虚假信息,如假新闻,以及深度伪造的图像和视频等。这些虚假内容不仅可能误导公众,破坏社会信任,还可能引发社会动荡、经济损失甚至暴力事件。在科学领域,研究者有可能利用生成式人工智能技术轻易制作虚假研究论文并伪造实验数据等,这些虚假成果一旦发表,将严重干扰学术界的正常秩序,误导科研人员的研究方向,严重损害科研诚信,破坏科研生态。此外,恶意利用生成式人工智能进行造假和欺诈等行为也日益增多,如利用深度伪造技术进行勒索、诽谤以及通过生成式人工智能辅助的网络攻击,对数字系统和网络安全构成严重威胁。

科研项目管理是一项系统性强、任务复杂的工作,

收稿日期:2024-11-15; 修回日期:2025-06-08

* 通信作者, Email: yaochang@nsfc.gov.cn; chen.gong@njut.edu.cn

引用格式: 姚畅, 宫辰, 李广宇, 等. 生成式人工智能在科研项目管理中的风险分析及应对措施. 中国科学基金, 2025, 39(4): 1-10.

Yao C, Gong C, Li GY, et al. Risk analysis and countermeasures of generative artificial intelligence in research project management. Bulletin of National Natural Science Foundation of China, 2025, 39(4): 1-10. (in Chinese)

其核心目标在于保障科研资源的合理配置,推动科研活动高效有序地开展,并产出高质量成果^[1]。然而,随着生成式人工智能技术的迅猛发展,科研项目管理也面临着诸多新的风险和挑战。例如,从科研项目管理全流程的角度出发,在项目申请阶段,申请人可能利用生成式人工智能伪造研究内容、项目背景、技术路线、研究基础和答辩视频等材料,以提高申请成功率。这种欺诈行为不仅浪费了宝贵的科研资源,也损害了科研管理部门的公正性和公信力。在项目实施过程中,生成式人工智能可能被用于伪造实验数据和篡改研究结果等。由于这些虚假数据具有高度的迷惑性,传统的监管手段难以有效识别,从而给科研诚信管理带来了巨大挑战。在成果验收阶段,生成式人工智能生成的科研成果可能以假乱真,导致验收专家难以判断其真实性和创新性。这不仅可能使不合格的项目通过验收,还可能使真正有价值的科研成果被埋没。据不完全统计,截至2024年12月,国内外媒体累计报道了79起因生成式人工智能引发的学术不端事件,涉及行为包括但不限于利用人工智能直接生成论文、改写内容、撰写审稿意见、伪造图像、捏造参考文献及生成虚假项目申请书等。例如,2023年国内某高校的作者在向*Frontiers in Cell and Developmental Biology*投稿的过程中使用人工智能生成图像,该论文最终被撤稿处理;同时,一位审稿人在为*International Journal of Molecular Sciences*评审论文时,利用人工智能生成了不当审稿意见。此类行为违背了科学研究中真实与可信的基本原则,影响了学术诚信与科研透明度。此外,生成式人工智能严重依赖大量用户上传的训练数据,由此引发的权属纠纷、信息泄露、数据问题也不容小觑,严重者甚至可能威胁到国家安全。由此可见,生成式人工智能技术具有很强的“两面性”。一方面,其有可能在科研项目管理中的某些环节得到应用,但另一方面,其潜在的风险和挑战也不容忽视。

因此,为全面了解当前生成式人工智能在科研项目管理中的应用形势及政策动向,本文系统调研了全球范围内主要机构与专家对该技术的使用态度与管理政策。调研范围涵盖13个国家的科研资助机构、10个政府部门、13所国内外高校、11家出版机构、5家咨询机构以及9位知名学者等。此外,为推动生成式人工智能在科研项目管理领域的规范化应用,充分发挥其积极作用,同时有效防范潜在风险,本文利用管理学中的“动机—机会—能力”(Motivation-Opportunity-Ability, MOA)^[2]及“技术—组织—环境”(Technology-Organization-Environment, TOE)^[3]理论框架,从人工智能技术、法律法规和伦理道德等多方面深入分析生成式人工智能在科研项目管理

中可能引发的各类问题,从而为相关管理部门提供科学决策依据,促进科研项目管理体系的健康有序运行。

本文的组织架构如下:第2章调研全球范围内的相关权威机构和专家对生成式人工智能应用于科学研究的态度和政策;第3章基于MOA和TOE理论构建生成式人工智能在科研项目管理中的风险分析框架;第4章对生成式人工智能在科研项目管理中的具体风险进行详细分析和深入讨论;第5章从政策、管理和技术三个方面提出相应的风险应对措施;第6章总结全文并对生成式人工智能在科研项目管理中的应用进行展望。

2 现状调研

为全面掌握生成式人工智能对科学研究的潜在影响,本文系统调研了全球范围内相关权威机构与专家对该技术应用的政策态度,调研对象涵盖各国科研资助机构、政府主管部门、高等院校、国际出版机构、专业咨询机构及知名学者。

2.1 各国科研资助机构

本节系统调研了各国科研资助机构关于生成式人工智能的相关政策,并分别从申请人和评审人的视角分析其在项目申请与评审环节中的应用立场与态度。对于申请人,瑞典研究理事会、荷兰研究理事会、德国研究协会、美国国家科学基金会、英国癌症研究中心、英国维康基金对申请人使用生成式人工智能持开放态度,允许申请人在研究工作中有条件地使用生成式人工智能撰写资助申请,但均要求研究者在公开研究成果时,明确声明是否使用了人工智能以及使用的目的和范围。澳大利亚研究委员会、澳大利亚国家卫生和医学研究委员会、日本学术振兴会对申请人使用生成式人工智能持保守态度。这些机构认为研究者在使用生成式人工智能时,输入到生成式人工智能数据库的信息无法管控,可能会无意中侵犯版权并泄露个人或机密信息,因此建议申请者在撰写基金申请书时谨慎使用生成式人工智能工具。需要补充的是,上述所有机构均要求研究者自觉遵守科研诚信的基本原则,对提交内容的准确性和真实性负责,保证申请内容的原创性、有效性、可靠性和完整性。对于评审人,上述大多数科研资助机构,以及美国国家卫生研究院、瑞士国家科学基金会和比利时国家科学研究基金普遍认为,生成式人工智能的使用违反了基金评审中的保密性与独立性原则,因此明确禁止评审人员在评审过程中使用此类工具。

2.2 各国政府部门

为弥补现有政策法规在应对生成式人工智能相关

问题上的不足,世界各国政府部门纷纷出台针对生成式人工智能的政策法案,将生成式人工智能纳入现有法律体系内,为其合理使用提供指导规范。例如,2025年3月中国国家互联网信息办公室等四部委发布《人工智能生成合成内容标识办法》,欧盟委员会于2025年发布了《科研中负责任使用生成式AI的指南》,2021年1月美国国会正式颁布《2020年国家人工智能倡议法案》。此外,中国、欧盟、美国、英国、日本、韩国、新加坡、联合国及七国集团等主要国际治理主体普遍持开放与审慎并重的态度,鼓励研究人员在科学研究中合规使用生成式人工智能,同时强调保障其使用过程中的透明性、公平性与可追溯性。

2.3 国内外高校

生成式人工智能正深刻影响高校师生的教育和学习方式,本节调研了多所国内外高校在教学场景中对该技术的使用态度与管理政策。哈佛大学、德国柏林洪堡大学、宾夕法尼亚大学、新加坡国立大学、墨尔本大学、悉尼大学和香港大学等高校对学生使用生成式人工智能持鼓励和支持态度,他们认为生成式人工智能不应被视为潜在威胁,更可作为促进学生问题解决能力与批判性思维培养的有力工具。日本东京大学、美国范德比尔特大学、法国巴黎政治学院和德国图宾根大学等高校禁止学生使用生成式人工智能工具,原因是生成式人工智能可能会导致版权和隐私侵犯、作弊、学术剽窃以及信息偏见等问题。此外,北京大学、上海交通大学与康奈尔大学等高校已相继出台人工智能系统使用规范与行为指引,力求在鼓励探索技术潜力的同时,构建合理边界与应对机制,以保障教育公平与学术诚信。

2.4 出版机构

本节系统调研了国际主要出版机构关于生成式人工智能使用的政策规范,并分别从作者与审稿人两个维度分析其使用态度与管理要求。在作者层面,几乎所有出版机构均强调学术论文必须遵循客观、严谨与准确的基本原则,明确要求作者对提交内容的原创性、有效性与完整性承担全部责任,严禁触碰学术不端的底线。IEEE、Elsevier、Springer Nature、SAGE、Wiley、Taylor & Francis、arXiv、Science、Cambridge University Press、American Chemical Society和PLOS ONE等出版机构认为作者可以有条件地使用生成式人工智能工具,前提是使用行为应审慎、可控且具有明确目的。需要指出的是,Elsevier、Springer Nature以及Science明确指出原则上不允许在论文中使用人工智能生成的图像和其他多媒体素材。对于审稿人,Elsevier和SAGE明确禁止评审人员在论文评审过程中使用生成式人工智能

工具。

2.5 咨询机构

本节调研了麦肯锡、埃森哲、德勤、毕马威和贝恩五家国际知名咨询机构对生成式人工智能的看法。这些机构普遍认为,生成式人工智能具有显著的应用潜力,能够提升生产效率、加速知识获取,并在激发人类创造力与增强生产力方面发挥积极作用。然而,现阶段人工智能的监管体系仍不健全,生成式人工智能在透明性与可解释性、公平性与安全性、可追责性与合规性以及知识产权和隐私保护等方面仍面临诸多亟需关注的关键问题。

2.6 知名学者

除各类权威组织外,本文还调研了多位国内外专家学者对生成式人工智能发展的观点与态度。清华大学教授唐杰、中国科学院院士张钹以及特斯拉首席执行官埃隆·马斯克、著名投资人沃伦·巴菲特等业内具有广泛影响力的代表人士普遍对人工智能的未来持开放态度,认为生成式人工智能的快速崛起既带来技术机遇,也伴随诸多挑战,具有广阔的应用前景。与此同时,上述人士也对人工智能技术的不受控扩张表示担忧,呼吁在其研发与应用过程中加强监管与伦理约束,特别关注其可能引发的问题,如大模型理解能力有限、劳动力替代所致的就业冲击、算法造成的各类偏见、数据泄漏风险以及数字伦理困境等。

2.7 调研小结

总体而言,在客观、严谨和诚信的基本原则下,国内外多数机构和专家普遍对生成式人工智能的应用持“有条件的肯定开放”态度。一方面,生成式人工智能展现出广阔的应用前景,在提升人类创造力与生产效率方面具有巨大潜力;另一方面,技术透明性、隐私安全、知识产权、可解释性与可追责性等问题亦成为各界的共同担忧。需要指出的是,不同国家在治理生成式人工智能风险方面展现出显著的制度差异。以欧盟为代表的地区更强调前置监管与伦理规范,侧重通过立法确立通用性原则和强制合规要求;美国则更倾向于技术驱动与自律治理,强调创新激励与市场导向下的责任分担;而中国则呈现出政策引导与制度建设并重的治理路径,致力于在推动技术发展与维护社会稳定之间实现动态平衡。鉴于此,建议相关管理部门在借鉴国际经验的基础上,进一步完善国内法律法规与配套实施细则,明确生成式人工智能的使用边界与合规要求,并构建基于风险的分分类分级监管体系,以实现科研伦理底线的有效守护,推动生成式人工智能在科研领域的健康、安全与可持续发展。

3 MOA-TOE风险因素分析框架

从全球政府机构、科研资助机构与全球知名学者对生成式人工智能的响应与立法措施来看,该技术在科技、经济及社会层面引发的风险治理问题日益凸显,亟需系统应对。科研资助体系在推动国家基础研究与创新发展的过程中扮演着至关重要的角色,关涉教育、科技与人才等国家战略的深入实施。在此背景下,本节将系统分析并分类讨论生成式人工智能在科研项目管理中可能引发的各类风险。

从内容生成视角来看,ChatGPT等生成式人工智能工具的输出并非类智慧体的逻辑思考,而是模型基于文字排列概率预测所产生的内容。然而,技术只是社会风险治理的组成要素之一,从风险治理与科研项目管理视角来看,生成式人工智能产生的风险并非简单的技术风险叠加,而是技术、开发者、用户、技术支持要素以及制度环境等多维要素共同作用的结果。因此,生成式人工智能在科研项目管理中产生的风险亦与科研项目相关主体、人工智能发展水平以及相关制度环境息息相关。基于上述分析,如图1所示,本文从微观行为与宏观发展双重视角出发,构建生成式人工智能在科研项目管理中的风险分析框架。具体而言,采用“动机—机会—能力”模型分析科研项目管理相关主体在使用生成式人工智能过程中的微观风险;同时引入“技术—组织—环境”框架,探讨技术发展态势、组织运行机制与政策环境等宏观层面的潜在风险因素。

3.1 MOA框架下的风险因素分析

MOA理论框架由MacInnis和Jaworski^[2]提出,包含动机层、能力层、机会层三个部分,源自传播学和营销学领域的广告信息选择行为研究^[4,5]。MOA模型自被提出以来,已作为一个重要的理论框架应用于多个学科领域,用于解释一系列行为现象,如消费者信息选择行为^[6]和知识共享行为^[7]等。

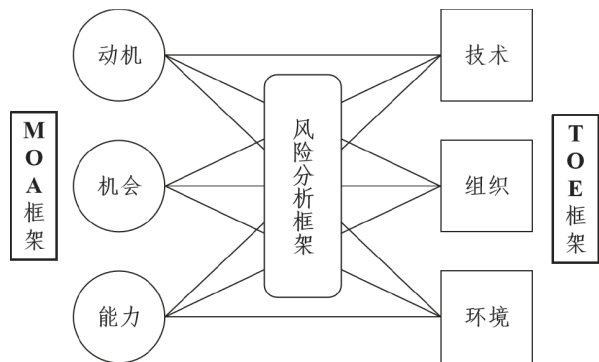


图1 基于MOA和TOE理论的风险因素交叉分析框架

Fig.1 Cross Analysis Framework for Risk Factors Based on MOA and TOE Theories

3.1.1 动机层

动机层是驱动科研项目相关主体使用生成式人工智能的基础,主要包括研发动机、利用动机以及消极动机。(1)研发动机:生成式人工智能模型开发者对于现有模型进行迭代和改进的内在驱动力。(2)利用动机:项目申请人或评审人可能利用生成式人工智能技术提高科研效率和成果质量。(3)消极动机:恶意行为者可能会利用生成式人工智能实施学术不端等恶意行为,从而对科研项目管理流程及科研项目成果造成损害。

3.1.2 机会层

机会层关注生成式人工智能技术在科研项目管理中的实施基础与应用环境,主要包括模型自主可控性、人工智能发展基础(主要指数据基础)以及模型可追溯性三方面影响因素。(1)模型自主可控性:当前许多大模型是在美国企业主导的技术基础上进行改良和定制而得,其采用的人工智能技术可能面临技术依赖和技术封锁等风险。(2)人工智能发展基础:生成式人工智能在模型训练及推理时往往面临数据来源受限等问题,如果数据集存在系统性的偏差,那么基于这些数据训练出的模型也会继承这种偏见,可能导致某些类型的项目被不公平地对待。(3)模型可追溯性:目前已公开的大模型有很多,项目申请人和评审人等主体所用大模型来源往往不可追溯,同时缺少相应的安全审查机制。

3.1.3 能力层

能力层主要包括技术开发能力、用户参与能力以及数据输入质量三个方面。(1)开发能力:技术的快速迭代需要高水平开发能力,在确保涉及敏感信息的科研项目系统的安全性与合规性的同时,开发过程还需要充分考虑伦理和社会影响。(2)用户参与能力:缺乏研究经验的项目申请人以及未尽到注意义务的评审专家对项目研究内容的主观判断可能会导致对人工智能生成内容的误用,影响科研项目评审的公正性和透明度。(3)输入质量:就生成式人工智能的内容生成机制而言,低质量的数据输入可能会导致模型学习到错误的模式或偏见,影响科研项目分配的公平性和有效性,此外还需要确保数据输入的安全性和保密性。

3.2 TOE框架下的风险因素分析

TOE分析框架是一种用于技术应用行为分析相关领域的理论模型,该框架最初由Tornatzky等人提出,随后被广泛应用于创新管理领域。TOE分析框架认为,新兴技术的应用主要受三方面因素影响,即技术特征、组织特征和环境特征,这三个维度共同决定了技术应用的可能性和效果^[7]。

3.2.1 技术层

技术层风险主要来自于构成生成式人工智能核心要素的数据、算法和算力三个方面。(1)数据:在科研项目管理背景下,数据风险主要涵盖数据质量风险与数据隐私风险。数据若存在偏差或缺失,可能会导致生成式人工智能在基金评审过程中产生错误的建议或结论。(2)算法:算法风险主要包括算法偏见风险和可解释性风险。前者指算法设计中存在的偏见可能导致对特定群体不公平的评价,后者指缺乏透明度和可解释性的算法可能导致决策过程难以被理解和验证。(3)算力:算力风险主要包括资源分配风险与能源消耗风险。国外主导的大模型服务提供商有可能对特定的IP地址或区域实施算力限制,影响我国科研竞争力。此外,人工智能服务器往往消耗大量能源,不符合我国可持续发展战略。

3.2.2 组织层

组织层因素主要来自个人、科研机构以及管理主体三个方面。(1)个人(项目申请人、负责人和评审专家等):项目申请人和评审专家对生成式人工智能的滥用以及科技伦理道德问题是在其科研项目管理中的主要风险点。(2)科研机构(高校院所相关科研管理处室):当项目申请人或评审专家所在的科研管理部门缺乏明确指导性意见时,可能会导致生成式人工智能技术在科研工作中被不当使用,从而引发数据伪造、研究原创性丧失、科技伦理风险和评审标准差异等问题。(3)项目管理主体:其对科研项目管理中生成式人工智能使用的评估与应用是其风险的核心内容,包括评估现有工作流

程与新兴人工智能技术的兼容性、可用性与安全性,并在必要时调整流程以适应新技术的要求。

3.2.3 环境层

环境层风险的宏观影响因素主要包括主管部门监管力度、制度保障以及国际外部影响因素三个方面。(1)主管部门监管力度(科技部等科技管理部门):主管部门需要确保人工智能技术在科研项目管理中的应用有效合规,政策执行的力度适当,监管框架的灵活性合理。(2)制度保障:知识产权制度和数据安全制度是保护科研项目良性运转的关键禀赋,缺乏适当的法律框架可能会导致数据泄露以及知识产权侵权等问题。(3)国际外部影响:可能存在西方国家对于生成式人工智能技术的封锁与后门风险,甚至可能有盗取数据、恶意输出虚假科研信息以迷惑我国科研人员研究思路的安全风险。

4 生成式人工智能在科研项目管理中的风险分析

本节综合分析MOA框架下微观风险因素与TOE宏观风险因素,研究生成式人工智能在应用期间的不同微观风险因素在宏观条件下可能对科研项目管理产生的具体风险。分析结果如表1所示。通过整合分析结果,本文总结生成式人工智能在科研项目管理中的六大风险,包括:(1)隐私与安全风险,即生成式人工智能系统可能导致敏感数据泄露,威胁个人隐私和国家科技安全;(2)内容与决策风险,即生成式人工智能系统可能生

表1 基于MOA-TOE框架的生成式人工智能在科研项目管理中的主要风险分析表
Table 1 Main Risk Analysis of Generative Artificial Intelligence Based on MOA-TOE Framework in Scientific Research Project Management

		技术			组织			环境		
		数据	算力	算法	个人	科研机构	管理主体	监管力度	制度保障	外部影响
动机	研发动机	低质量数据	资源分配	算法安全	数据隐私	资源分配	资源分配	监管不足	权属争议	技术封锁
	利用动机	隐私泄露 内容风险	算力不足	情报窃取	决策分析 道德诚信	能源消耗	数据安全 社会舆论	道德诚信	知识产权 侵权	科技安全
	消极动机	虚假数据 恶意引导	国际打压	算法后门	道德诚信 科技安全	创新多元性	创新多元性	监管不足	数据知识 产权	恶意供给 情报窃取
机会	自主可控	低质量 内容生成	资源分配	低质量 内容生成	阻碍原创	阻碍原创	安全审查	技术审查	法律及 政策支撑	技术后门
	模型基础	/	/	/	创新多元性	资源分配	资源分配	/	知识产权 侵权	技术瓶颈
	可追溯性	低质量 内容生成	资源分配	恶意代码 开源安	创新多元性	安全审查	安全审查	安全审查	数据安全	模型盗用 安全审查
能力	开发水平	内容偏见	算力不足	算法偏见	技术成熟度	技术黑箱	文化风险	技术审查	算法侵权	算法侵权
	用户参与	定向反馈	资源消耗	算法黑箱	低质量反馈	道德诚信	道德诚信	内容侵权	内容侵权	单一来源
	数据输入	内容偏见	资源分配	定向输入	定向输入	创新多元性	创新多元 虚假数据	虚假数据	数据侵权	恶意输入

成低质量或误导性内容,导致科研项目管理者 and 专家做出错误决策,误用科研资金;(3)权属与侵权风险,即生成式人工智能系统生成的内容可能触及版权问题,引发知识产权的法律争议;(4)道德与诚信风险,即生成式人工智能系统可能被用于伪造数据或研究结果,损害科研诚信和机构声誉;(5)多元与创新风险,即过度依赖人工智能工具可能导致研究内容同质化,抑制创新思维和研究多样性;(6)文化与伦理风险,即生成式人工智能系统的输出内容可能与社会文化和伦理价值观不符,引发公众反感和信任危机。

4.1 隐私与安全风险

隐私与安全风险主要是指生成式人工智能系统在执行生成任务时可能会造成敏感数据泄露,从而侵犯个人隐私,并威胁企业和科研机构的发展安全以及国家的科技安全。

4.1.1 隐私泄露风险

首先,生成式人工智能系统对数据的保护力度不足可能会导致隐私泄露。具体来讲,生成式人工智能的训练和运作依赖于大量敏感数据。在缺乏严格的数据保护和匿名化措施的情况下,敏感信息可能在生成任务执行过程中被无意泄露。其次,由于生成式人工智能模型的数据来源多样且规模庞大,信息管理存在较大难度,导致信息滥用或泄露的风险增加。最后,生成式人工智能的应用还可能面临第三方数据共享和个人数据被二次利用的风险。

4.1.2 安全风险

安全风险主要是指生成式人工智能系统在执行生成任务时,因内在数据问题或外在恶意攻击,对企业、科研机构及国家科技安全造成的潜在威胁。(1)内在数据外泄:在使用生成式人工智能系统时,科研人员往往需要将研究内容上传至国外的服务器,容易造成科研情报的外泄,进而威胁企业及国家科技安全。(2)外在恶意攻击:生成式人工智能系统的安全性是一个复杂但关键的问题。许多技术供应商在保护数据安全方面的投入可能不足,难以充分保障系统抵御高级网络攻击的能力,这使得生成式人工智能系统容易成为网络攻击的目标,引发一系列安全风险。

4.2 内容与决策风险

内容与决策风险是指生成式人工智能系统由于技术局限可能会生成低质量、误导性甚至虚假的内容,从而误导科研项目管理部门及相关专家,引发决策错误问题。

4.2.1 内容风险

内容风险主要是指生成式人工智能系统在执行生成

任务时由于自身的技术限制(如幻觉现象、不正确推理、以及“一本正经地胡说八道”)而产生低质量、误导性或虚假错误的内容。基于生成式人工智能完成的项目申请书在形式上可能非常完美且合理,但可能存在技术机理、研究路线和技术可行性等缺陷,从而给评审过程带来困难,误导科研项目的评审过程,导致科研经费被错误分配给不合适的项目,引发项目评审决策失误。更严重的是,如果生成式人工智能系统被技术供应商恶意篡改、引导或利用,可能会故意生成误导或错误的信息,这种情况可能直接影响到国家层面的科技发展决策。

4.2.2 决策风险

在科研项目管理中,评审过程的公正性和科学性至关重要,这能够确保资金分配给最有潜力的研究项目。个别项目评审专家可能利用生成式人工智能系统生成评审意见,但生成式人工智能模型目前普遍存在泛化能力弱、可解释性差和输出不稳定等问题,使得生成的评审意见可能存在错误,进而导致评审决策不合理、不公正。更严重的是,如果生成式人工智能系统被不当使用或恶意利用,可能会生成故意误导或偏颇的评审意见,从而操控评审结果。这种情况不仅损害了评审过程的公正性,还可能对科研领域的整体诚信造成长远的负面影响。

4.3 权属与侵权风险

权属与侵权风险主要是指生成式人工智能生成的内容可能涉及知识产权问题,从而引发版权侵权和权属争议等风险,进一步产生归责难题。

4.3.1 权属风险

随着生成式人工智能技术在科研领域的广泛应用,其产生的权属风险正日益受到学术界与法律界的高度关注。这些风险主要涉及作品资格认定和权利归属两个核心问题。在作品资格认定方面,人工智能生成物的法律地位存在较大争议。传统著作权法以人类创造性智力劳动为保护基础,而人工智能生成内容则模糊了原创和抄袭的边界^[8,9]。目前,从外观形式上难以区分科学作品是由人类还是人工智能创作的,这不仅影响了作品的法律保护,还可能引发学术诚信问题。权利归属问题则涉及多个利益相关方:首先是人工智能系统开发者,他们投入大量资源开发和训练模型;其次是用户,他们提供输入并指导人工智能生成特定内容;还有训练数据的原始权利人,他们的作品被用于人工智能训练。这些主体对人工智能生成物都有不同程度的贡献,如何合理分配权益成为一个复杂的法律和伦理问题。

4.3.2 侵权风险

根据使用者身份以及人工智能生成物的使用方式,

科研项目管理过程主要涉及三类侵权风险:复制侵权风险、改编侵权风险和主体认定风险。(1)复制侵权风险:科研项目管理各方主体对人工智能的使用主要集中在生成物的输出环节,人工智能可能会生成与已有研究权利人作品基本表达相同或者相似的内容,即存在使用人工智能生成物侵犯已有研究权利人作品复制权的风险。(2)改编侵权风险:当生成物保留了原作品的基本表达,但又与原作品存在实质性区别时,就会引发改编侵权风险。(3)侵权主体认定风险:在人工智能生成物问题上,其中一个很难绕过去的问题是,究竟是认定使用人工智能侵权生成物的用户侵权,还是认定生成式人工智能投资者侵权?或者是二者都侵权?人工智能生成的过程涉及用户与人工智能之间的互动,传统的侵权检验标准难以适用。

4.4 道德与诚信风险

目前人工智能技术呈现高速发展态势,但也存在政策价值取向不明晰、法律法规机制不健全和技术伦理约束滞后等问题,存在着巨大的不确定性和道德信任风险。从总体上看,生成式人工智能的使用有可能存在如下问题。(1)破坏科研项目管理过程中的公平正义:人工智能可能本身存在数据偏见,导致资源分配不公平;(2)冲击科研项目管理过程中的责任伦理:生成式人工智能的使用可能会模糊责任边界,引发责任归属的争议^[10];(3)降低科研项目管理过程中的信任:生成式人工智能的决策过程可能缺乏透明度,使得决策结果难以解释和理解。具体来看,生成式人工智能的发展逐渐贯穿于科研项目管理的全生命周期,从项目申请、项目评审、项目执行、项目结题到成果发布,主要表现为自动生成项目申请书、评审意见自动撰写、数据和资料伪造以及虚假研究成果发布等。在这些环节中,生成式人工智能的使用为科研项目管理带来了诸多道德与诚信方面的风险和挑战,具体如图2所示。

4.5 多元与创新风险

多元与创新风险是指科研人员过度依赖生成式人工智能工具,可能导致研究内容同质化、研究方法平庸化和评价标准单一化,使得原创性研究严重缺乏、原创性项目被忽略,从而影响科学发展的多样性和独特性。

4.5.1 多元化风险

多元化风险主要涵盖研究内容同质化、研究方法平庸化以及评价标准单一化三个方面。(1)研究内容同质化:生成式人工智能依靠已有训练数据来生成内容,导致输出内容具有高度的相似性。(2)研究方法平庸化:生成式人工智能在生成研究方法和实验设计时,往往依赖于已有的成熟方案,缺乏独特性和前瞻性。(3)评价

标准单一化:生成式人工智能在项目评审环节应用的过程中,可能会依赖于预设的算法和模型,仅凭借固定的评分标准进行评估,往往会忽视项目的独特性和创新性。

4.5.2 创新风险

创新风险主要指原创性科研缺失和原创性项目被忽视。(1)原创性科研缺失:指科研人员严重依赖生成式人工智能寻找学术思路 and 方案,减少对未知领域的独立思考和主动探索,从而影响原创性成果的产出。(2)原创性项目被忽视:在评审过程中,生成式人工智能可能对某些非主流研究项目产生偏见,从而影响评审结果的公正性和科学性,限制科学研究的创新性。

4.6 文化与伦理风险

文化与伦理风险是指生成式人工智能系统可能输出与主流科学文化价值观相悖、有违公序良俗和科技伦理的不良内容,从而对社会舆论产生恶劣影响。

4.6.1 文化风险

文化风险包括违背主流科学文化风险和违背公序良俗风险两方面。(1)违背主流科学文化风险:生成式人工智能系统的输出内容往往不可控,很可能生成具有技术偏见、歧视、非客观以及其他有损公平公正性的内容,对人们科学文化的认知构成威胁,还可能威胁科研人员正确的价值取向。更有甚者,生成式人工智能可能会生成违背社会主义核心价值观的内容,增加意识形态冲突风险。(2)违背公序良俗风险:生成式人工智能在生成内容时可能会无意中强化或放大种族、性别、年龄、地域和群体等社会偏见。此外,虚假信息 and 误导性内容可能破坏科技行业的信任基础。

4.6.2 伦理风险

生成式人工智能在科技伦理上有多方面潜在风险,主要包括以下方面:(1)抄袭和剽窃:生成式人工智

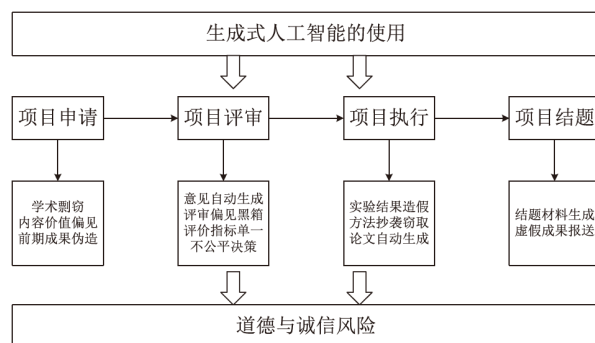


图2 生成式人工智能在科研项目管理全生命周期中的道德与诚信风险

Fig.2 Moral and Integrity Risks of Generative Artificial Intelligence in the Whole Life Cycle of Scientific Research Project Management

能技术生成的论文和报告等材料可能会存在未经授权或允许的非原创内容。(2)数据篡改:生成式人工智能在数据处理方面有着高效性和精确性,但也存在数据被恶意篡改的风险。(3)伪造实验结果:生成式人工智能技术可以生成逼真的虚假数据或实验结果,使得研究人员可以轻松地伪造科研成果。(4)影响评价体系:部分机构和期刊审稿人使用生成式人工智能来评估论文、研究成果及科研基金项目,影响评价体系的公平性。(5)知识不当归因:在使用生成式人工智能系统生成文本时,研究人员需要特别注意正确引用和归因。(6)数据使用合规风险:用户在使用生成式人工智能时可能会输入包括个人信息和商业秘密等在内的敏感内容,若这些内容被不当使用,或者超出了合理使用的范围,都可能导致法律纠纷。对项目管理部门而言,不当使用生成式人工智能可能严重损害其公众形象和信誉,特别是当相关内容被媒体曝光后,相关管理部门不仅会面临来自社会各界的批评,还可能承受政策和舆论的双重压力,进而严重影响其日常运作和未来的资助决策。

5 生成式人工智能在科研项目管理中的风险应对

由以上分析可知,生成式人工智能在提升科研效率与质量的同时,也为科研项目管理带来了诸多风险与挑战。当前面临的风险主要集中在隐私与安全、内容生成与决策、权属与侵权、道德与科研诚信、多元性与创新性以及文化与伦理六个方面。这些风险不仅可能导致科研人员隐私泄露、学术不端行为频发、知识产权纠纷增多,还可能引发科研成果同质化、伦理争议加剧以及科研资源的不合理消耗,从而严重影响科研项目管理的公正性与公信力。尽管相关政策、技术和管理手段已在一定程度上对上述问题进行了初步探索,但在应对生成式人工智能生成内容的真实性、合规性及伦理正当性方面仍需进一步完善与强化。为充分发挥生成式人工智能的技术优势,同时有效规避其潜在风险,本文从政策、管理和技术三个维度提出生成式人工智能在科研项目管理中的风险应对措施。

5.1 政策措施

为实现生成式人工智能在科研项目管理中的规范化应用,相关管理机构应从培训教育、规范使用、技术支持与伦理建设等多个维度综合施策。首先,必须强化部门内部的生成式人工智能知识普及,可通过定期组织专题培训、制定人工智能使用管理办法以及建立分级风险控制机制,确保工作人员正确理解和使用相关工具。对

于项目申请人,应提前进行教育宣传,并要求签署承诺函,以明确其使用人工智能工具时的责任和规范。同时,在申报过程中,申请人需真实声明人工智能使用情况,避免隐瞒或滥用。此外,有必要建立事后惩戒机制,对违规行为严肃处理,以此保障申请流程的公正性和科学性。

在项目评审和执行环节,需明确各参与主体的行为边界。评审专家需事先承诺不在评审中使用生成式人工智能工具,保证评审结果的独立性和公平性;项目负责人则需签订使用规范协议,并定期接受项目检查,确保研究过程的合规性和成果的真实性。同时,应进一步推动生成式人工智能内容检测与鉴定技术的研发,提升对生成内容的甄别能力,确保人工智能输出内容的可靠性和真实性。此外,针对可能的内容侵权问题,建立涵盖法律咨询、风险预警与危机响应的全链条管理机制,全方位降低潜在的法律风险。

此外,需建立快速响应机制,应对因人工智能使用不当引发的负面舆论,及时化解公众疑虑,维护机构公信力。通过完善科技伦理相关立法和政策,制定明确的操作指南和行为标准,为科研人员和管理者提供明确的操作指引。在伦理教育方面,应加强对科研人员的学术道德培训,提升其对生成式人工智能伦理风险的认知能力,并培养良好的学术行为规范。

5.2 管理措施

为有效管理生成式人工智能在科研领域的应用,相关管理部门应从组织、制度、技术、培训和伦理多层面入手,实施一系列综合性管理措施。首先,在机构内部成立专班和工作组,由领导牵头统筹战略规划,设置学术诚信管理办公室负责处理生成式人工智能相关的学术不端案件,并引入具备人工智能背景的专业人才以提升效率和专业性。其次,出台针对生成式人工智能的专门管理办法,明确科研人员在项目申请及研究中使用人工智能的范围和规范,防止学术诚信问题。同时,在知识产权方面,应制定详细的管理制度,厘清版权归属与法律责任,防范产权争议。

此外,强化评审专家的人工智能相关培训,提升其鉴别能力及对相关法律、伦理问题的理解,保障评审质量。为防止数据泄露和滥用,制定严格的数据安全管理规定,确保生成式人工智能应用过程中涉及的敏感数据得到充分保护。为推动公平资助,应关注不同研究团队的技术与资源差异,避免人工智能技术不平等导致的资源分配不均,并通过支持创新与多样性促进科研生态的健康发展。最后,加强科技伦理教育,在科研人员、项目负责人和依托单位中强化生成式人工智能伦理风险的

认识,弘扬科学道德,营造公正、诚信的学术环境。通过一系列有力的政策和措施,确保生成式人工智能在科研领域的规范应用,推动技术进步和学术诚信的协调发展。

5.3 技术措施

为确保生成式人工智能技术安全、高效地应用于科研项目管理,相关机构需从技术基础设施、数据安全、使用权限和内容质量等多方面采取措施。首先,部署符合我国科技体系的国产生成式人工智能模型,并依托封闭式计算集群,确保运算高效和数据可控。同时,设立严格的信息审查流程,对用户上传的数据进行安全审核及脱敏处理,保护敏感信息的安全性。在网络环境方面,应将生成式人工智能模型部署在内网中,断开外部网络连接,从根源上减少数据泄露的风险。通过严格的访问权限管理,仅允许经过认证的人员使用该技术,杜绝未经授权的操作。此外,建立详细的操作日志记录和定期审核机制,涵盖输入数据、生成内容以及用户行为,以便快速识别潜在风险并及时应对。

同时,应开发先进的人工智能内容检测工具,构建人工智能生成内容鉴别系统,以帮助项目管理人员和评审专家识别申请材料中的人工智能生成内容,确保评审的公正性和科学性。另外,为进一步增强安全性,可在管理流程中的低风险环节先行试点运行生成式人工智能工具,在积累一定经验后再稳步推广到更多应用场景。最后,建设实时监控体系,对生成内容进行持续监督,发现违规行为及时整改,确保生成式人工智能应用符合安全、法律和伦理标准。

6 总结与展望

在科研项目管理中,生成式人工智能犹如一把双刃剑,虽然给项目管理带来了显著的效率提升,但同时也引入了诸多风险和挑战。本文首先调研了全球范围内的相关权威机构和专家对生成式人工智能应用于科学研究的政策立场与治理态度。随后,借助MOA-TOE框架从微观行为与宏观发展的双元视角系统性地分析了生成式人工智能在科研项目管理中可能引起的风险,揭示了多种因素如何通过相互作用来影响科研项目管理的安全与效率,从而为相关管理部门制定科学合理的风险治理政策提供了理论支撑。最后,本文逐一分析了生成式人工智能在科研项目管理中面临的六类主要风险类型,并从政策、管理和技术三个角度给出相应的风险应对措施,以确保科研项目管理工作的健康有序进

行,推动生成式人工智能在科研管理领域的规范化应用。总之,生成式人工智能在赋能科研项目管理的同时,也对科研资助体系的公正性、透明性与伦理性提出了更高要求。科研管理机构应主动开展全方位的风险评估,完善配套政策制度,强化科技伦理规范,推动生成式人工智能在科研项目管理中的规范、安全与可持续应用。

参 考 文 献

- [1] 黎常,金杨华. 科技伦理视角下的人工智能研究. 科研管理,2021,42(8):9—16.
Li C, Jin YH. A research on artificial intelligence from the perspective of science and technology ethics. Science Research Management, 2021, 42(8):9—16. (in Chinese)
- [2] MacInnis DJ, Jaworski BJ. Information processing from advertisements: Toward an integrative framework. Journal of Marketing, 1989, 53(4):1—23.
- [3] Awa HO, Ojiabo OU. A model of adoption determinants of ERP within T-O-E framework. Information Technology & People, 2016, 29(4):901—930.
- [4] Petty RE, Cacioppo JT, Goldman R. Personal involvement as a determinant of argument-based persuasion. Journal of Personality and Social Psychology, 1981, 41(5):847—855.
- [5] Batra R, Ray ML. Situational effects of advertising repetition: The moderating influence of motivation, ability, and opportunity to respond. Journal of Consumer Research, 1986, 12(4):432—445.
- [6] Xu XY, Gao YX, Jia QD. The role of social commerce for enhancing consumers' involvement in the cross-border product: Evidence from SEM and ANN based on MOA framework. Journal of Retailing and Consumer Services, 2023, 71:103187.
- [7] Kettinger WJ, Li Y, Davis JM, et al. The roles of psychological climate, information management capabilities, and IT support on knowledge-sharing: An MOA perspective. European Journal of Information Systems, 2015, 24(1):59—75.
- [8] 肖红军. 算法责任:理论证成、全景画像与治理范式. 管理世界, 2022, 38(4):200—226.
Xiao HJ. Algorithmic responsibility: Theoretical justification, panoramic portrait and governance paradigm. Journal of Management World, 2022, 38(4):200—226. (in Chinese)
- [9] 吴汉东. 人工智能生成作品的著作权法之问. 中外法学, 2020, 32(3):653—673.
Wu HD. Rethinking the copyright of works generated by artificial intelligence. Peking University Law Journal, 2020, 32(3):653—673. (in Chinese)
- [10] 刘宪权. 人工智能时代机器人行为道德伦理与刑法规制. 比较法研究, 2018(4):40—54.
Liu XQ. Robot ethics and criminal regulations in the age of artificial intelligence. Journal of Comparative Law, 2018(4):40—54. (in Chinese)

Risk Analysis and Countermeasures of Generative Artificial Intelligence in Research Project Management

Chang Yao^{1*} Chen Gong^{2*} Guangyu Li² Yuting Deng³ Sheng Wan² Peng Wei⁴ Junjie Huang⁴

1. Information Center, National Natural Science Foundation of China, Beijing 100085, China

2. School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

3. School of Intellectual Property, Nanjing University of Science and Technology, Nanjing 210094, China

4. Office of Science and Technology Research, Nanjing University of Science and Technology, Nanjing 210094, China

Abstract In recent years, the wave of generative artificial intelligence has swept across the globe, profoundly transforming the development trajectory of multiple industries. However, as generative artificial intelligence technology rapidly develops and finds widespread applications, its potential risks and challenges have gradually emerged, particularly in science fund project management. To comprehensively understand the application orientation of generative artificial intelligence in science fund project management, this paper conducts a systematic review of policies from related global institutions and expert viewpoints, providing a basis and reference for the decision-making of the National Natural Science Foundation of China. Furthermore, by employing the “Motivation-Opportunity-Ability” and “Technology-Organization-Environment” frameworks, this paper conducts an in-depth analysis of the various risks that generative artificial intelligence might introduce to science fund project management. These risks include privacy and security risks, content and decision-making risks, ownership and infringement risks, ethical and integrity risks, diversity and innovation risks, as well as cultural and ethical risks. This analysis encompasses perspectives on artificial intelligence technology, legal regulations, and ethical considerations, with the aim of ensuring the healthy and orderly progression of science fund project management and promoting the standardized application of generative artificial intelligence in research management. To deal with the challenges and risks brought about by generative AI in research project management, this paper further proposes corresponding risk mitigation measures from the perspectives of policy, management, and technology to ensure the healthy and orderly conduct of research project management, thereby promoting the standardized application of generative AI in the field of research management.

Keywords generative artificial intelligence; research project management; risk analysis; risk mitigation; Motivation-Opportunity-Ability (MOA); Technology-Organization-Environment (TOE)

姚 畅 国家自然科学基金委员会信息中心副研究员, 主要研究领域为科技政策与科研管理信息化。

官 辰 南京理工大学计算机科学与工程学院教授、博士生导师, IET Fellow。研究方向为人工智能、机器学习。入选中组部“万人计划”青年拔尖人才、中国科协“青年人才托举工程”; 获吴文俊人工智能优秀青年奖、中国电子学会自然科学二等奖、江苏省杰出青年基金等。

(责任编辑 陈 鹤 张 强)

* Corresponding Authors, Email: yaochang@nsfc.gov.cn; chen.gong@njust.edu.cn