文章类型：论文|刊号（ISSN）：2972-4236(P) / 2972-4244(O)

生成式人工智能赋能高等教育的风险透视与预警机制构建

姚苗
乌鲁木齐职业大学信息工程学院
DOI:10.12238/acair.v3i3.15594

[摘　要] 生成式人工智能(AI-Generated Content,AIGC)的迅猛发展为高等教育数字化转型注入新活力,但在赋能过程中也潜藏着诸多风险。本文基于AIGC赋能高等教育的六大智能应用场景(助教、助学、助评、助育、助研、助管),系统地剖析了如教学内容同质化、学习路径固化、大模型幻觉、认知偏差等风险,并从“制度-技术-管理-教育”四个维度构建风险预警机制。本研究旨在为高等教育领域规范、安全地应用生成式人工智能提供理论参考与实践路径。
[关键词] 生成式人工智能；高等教育；风险
中图分类号：TP18　**文献标识码：**A

Risk Analysis and Early Warning Mechanism Construction of Generative Artificial Intelligence Empowering Higher Education

Miao Yao
School of Information Engineering, Urumqi Vocational University

[Abstract] The rapid development of Generative Artificial Intelligence (AI-Generated Content, AIGC) has injected new vitality into the digital transformation of higher education, but it also hides many risks in the process of empowerment. Based on the six intelligent application scenarios of AIGC empowering higher education (assisting teaching, assisting learning, assisting evaluation, assisting education, assisting research, and assisting management), this paper systematically analyzes risks such as the homogenization of teaching content, the solidification of learning paths, large model hallucinations, and cognitive biases. Furthermore, a risk early warning mechanism is constructed from four dimensions: "system - technology - management - education". This study aims to provide theoretical references and practical paths for the standardized and safe application of generative artificial intelligence in the field of higher education.
[Key words] Generative Artificial Intelligence; Higher Education; Risk

引言

人工智能是引领新一轮科技革命和产业变革的战略性技术,具有溢出带动性很强的“头雁”效应。随着ChatGPT、GPT-4、Deepseek-R1等大模型的迭代升级,大模型和生成式人工智能已从技术探索阶段迈入规模化应用阶段,在智能交互、决策辅助、知识问答等方面展现出强大的创新能力和广泛的应用前景。随着算力水平与数据量的跨越式提升,以多模态大模型为代表的生成式人工智能模型与系统在教育领域展现出较大的应用潜力,不但可以为师生提供智能化实用工具,更是为高等教育的数字化转型注入新动能。然而,技术赋能的背后暗藏风险[1]。

当前,学界对AIGC的研究多聚焦于应用潜力,如个性化学习[2]、智能辅导[3],对风险的探讨仍停留在碎片化层面[4,5],缺乏系统性的风险识别与预警机制研究。高等教育作为知识生产与人才培养的核心场域,其稳定性与规范性直接影响社会发展根基。因此,亟需从技术特性与教育规律的交叉视角,剖析AIGC应用的风险图谱,并构建科学的预警体系,为高等教育的健康发展提供保障。

1 生成式人工智能发展现状

生成式人工智能是一类基于深度学习模型的技术,其核心特征在于通过对海量数据的训练,能够自主生成符合逻辑且贴近人类表达的文本、图像、音频等内容。2022年11月,美国OpenAI实验室发布的ChatGPT,在自然语言理解与生成领域实现显著突破。此后,该机构通过持续迭代推出GPT-3.5、GPT-4等系列模型,在上下文窗口长度、语义理解与推理能力、事实准确性等核心性能上不断提升；其中,GPT-4的多模态版本进一步拓展了模型的视觉理解与分析能力。2025年1月,中国人工智能企业深度求

索(DeepSeek)推出的新一代大语言模型Deepseek-R1,凭借128K上下文窗口、多语言深度对齐及复杂推理能力跃升等突破性进展,迅速引发全球学术界与产业界的高度关注。

从技术风险维度来看,生成式人工智能的风险主要集中于数据、算法与系统三个层面。在数据层面,存在数据偏差、信息泄露及数据垄断等问题[6],这些问题直接影响教育公平的实现与用户隐私安全的保障；算法层面则面临黑箱效应、隐性偏见及生成内容失控等风险[7],可能导致价值观偏差及大模型“幻觉”等风险；系统层面的风险则表现为技术漏洞及风险传导效应,易引发连锁性问题[8]。上述风险的形成,根源在于深度学习技术的固有局限、模型训练机制的结构性缺陷。

随着技术应用的快速推进,截至2025年7月,中国已有474个生成式人工智能服务通过国家网信办备案[9]。中国互联网络信息中心2025年7月21日发布的第56次《中国互联网络发展状况统计报告》显示,截至2025年6月,我国网民规模达11.23亿人,其中利用生成式人工智能产品回答问题的用户比例高达80.9%[10]。在此背景下,准确识别与分析生成式人工智能大模型应用的风险,构建科学的风险预警机制以保障认知安全,已成为亟待解决的重要课题。

2 AIGC赋能高等教育风险透视

2.1 AIGC赋能高等教育场景分类

随着生成式人工智能技术的快速发展,AIGC工具已融入教育教学及学生管理全过程,其应用场景可概括为以“智”助教、以“智”助学、以“智”助评、以“智”助育、以“智”助研、以“智”助管六个核心维度。

以“智”助教聚焦于为教师教学提供全流程技术支持,具体涵盖教育资源的智能检索与精准推荐、教学内容的自动化生成、多维度教学评价分析、实时学情追踪与诊断、智能题库构建、个性化试卷组编、作业自动化批改及在线答疑辅导等功能,通过技术赋能推动教师教学模式创新与教学质量提升。以“智”助学旨在为学生学习提供个性化支持服务,包括基于学习特征的资料精准推送、动态学习路径规划、沉浸式情境化学习环境构建、多语种学习辅助及智能编程指导等,助力学生提升自主学习效率与知识掌握质量。以“智”助评依托多模态数据采集与分析技术,实现学生画像构建、综合素质动态评价及个性化学习诊断等功能,为教育者提供全面的学生发展评估依据,同时为学生提供针对性的改进建议与服务。以“智”助育致力于通过人工智能技术促进学生德智体美劳全面发展,具体包含智能艺术创作辅助、艺术鉴赏能力培养、个性化体育训练方案制定、劳动教育场景模拟及智能心理疏导与支持等,助力实现“五育融合”的现代教育目标。以“智”助研主要为教师的教学研究与学术发展提供技术支撑,涵盖教师专业成长路径规划、基于实证的智能教研分析、科研实验的智能化平台支持及学术研究的智能辅助工具等,有效提升科研效率与教师专业发展水平。以“智”助管则借助人工智能技术实现教育管理全流程的智能化转型,包括学生信息的智能化管理与分析、校园安全的实时智能监控、家校沟通的精准化与高效化等,推动教育治理模式向数字化、智能化与科学化升级。

2.2 AIGC赋能高等教育应用场景的风险透视

AIGC在助教、助学、助评、助育、助研、助管六大场景中,均潜藏着多维度风险,这些风险既涉及技术应用的局限性,也关乎教育本质与生态平衡(见表1)。

应用场景	风险	归因
以“智”助教	教学内容同质化风险	生成式人工智能生成的教案、课件等教学资源因训练数据的局限性呈现出模式化特征,若教师过度依赖,会导致不同课程、不同教师的教学内容趋同,缺乏独特性和创新性,难以满足学生多样化的学习需求。
	智能答疑的局限性风险	生成式人工智能的答疑功能虽能快速响应学生问题,但对于一些复杂的、需要深度互动和情境理解的问题,可能给出表面化答案,无法引导学生深入探究,甚至可能固化学生的思维模式。
以“智”助学	学习路径固化风险	生成式人工智能根据算法为学生规划的学习路径,可能限制学生的自主选择和探索空间,使学生陷入预设的学习框架中,不利于培养学生的自主规划能力和创新思维。
	过度个性化导致的知识壁垒风险	精准推送的学习资料可能使学生只接触到自己感兴趣或已掌握领域的内容,形成“信息茧房”,阻碍学生对跨学科知识的涉猎和全面发展。
以“智”助评	评价标准单一化风险	生成式人工智能的评价体系往往基于预设的指标和数据模型,可能无法涵盖学生综合素质的多个方面,如道德品质、团队协作能力等难以量化的维度,导致评价结果不够全面客观。
	数据采集偏差风险	在构建学生画像过程中,若采集的数据样本存在偏差或不完整,会影响评价的准确性,可能对学生做出错误的评估,进而影响学生的发展机会。
以“智”助育	大模型幻觉及认知偏差风险	生成式人工智能在辅助艺术创作、提供体育训练方案等过程中,可能因训练数据中的不良价值观或偏见,对学生产生潜移默化的负面影响,偏离“五育融合”的正确方向。
	情感交流缺失风险	智能心理疏导虽然能在一定程度上为学生提供情绪支持,但无法替代人与人之间真实的情感互动和共情理解,可能导致学生情感需求得不到真正满足,影响其心理健康发展。
以“智”助研	科研思维弱化风险	生成式人工智能辅助科研实验设计、提供研究思路等,可能使研究者过度依赖技术生成的结果,减少自主思考和探索的过程,弱化科研人员的创新思维和独立研究能力。
	科研数据过度依赖风险	智能教研分析基于大量的数据支撑,若数据来源不可靠或存在质量问题,会导致分析结果失真,可能误导科研方向,造成科研资源的浪费。
以“智”助管	管理隐私泄露风险	生成式人工智能在进行学生信息管理、校园安全监控等工作时,涉及大量学生和教职工的个人隐私信息,若安全防护措施不到位,可能导致信息泄露,侵犯个人权益。
	管理决策机械化风险	智能管理系统基于数据做出的决策可能缺乏人性化考量,无法灵活应对校园管理中的复杂情况和特殊需求,导致管理效果不佳,甚至引发师生的不满情绪。

3 从“制度–技术–管理–教育”四个维度构建风险预警机制

3.1制度维度：构建完善体系,明确行为准则

制度建设是生成式人工智能在高等教育领域安全应用的基石,需从宏观层面制定覆盖全场景的规则体系,为各应用场景划定清晰的行为边界。

进一步明确生成式人工智能在以“智”助教、助学、助评、助育、助研、助管等所有场景应用的基本原则,包括合法性、安全性、公正性、教育性等核心要求。在此基础上,制定系列配套制度,规范数据的采集、存储、传输、使用等全流程,保障师生数据隐私安全；明确在各场景中,学校、教师、学生、技术提供方等不同主体的责任划分,当出现风险问题时能够依规追责。同时,建立制度动态更新机制,根据生成式人工智能技术的发展和应用中出现的新情况、新问题,定期对相关制度进行修订和完善,确保制度的时效性和适用性,为各场景的风险防控提供持续有效的制度保障。

3.2技术维度：打造智能防控网络,实现全面监测预警

技术层面需构建一体化的智能监测与防控系统,为各场景的风险预警提供技术支撑,实现对潜在风险的及时发现和有效处置。

搭建生成式人工智能应用风险监测中枢平台,整合各场景的风险监测数据,通过大数据分析和人工智能算法,对以“智”助教、助学、助评、助育、助研、助管等场景中的风险进行全面感知和实时监控。开发通用的风险识别模型,能够识别各场景中可能存在的共性风险,如大模型幻觉风险、内容质量风险等。建立风险等级评估体系,根据风险的严重程度、影响范围等因素,将识别到的风险划分为不同等级,并对应不同的预警和处置策略。当监测到风险时,系统能够自动发出预警信号,并推送至相关责任部门和人员,同时提供初步的处置建议,提高风险响应速度和处置效率。此外,加强技术防护能力建设,采用加密技术、访问控制技术、安全审计技术等,为各场景的应用提供安全保障,从技术源头降低风险发生的可能性。

3.3管理维度：健全组织架构,强化协同防控

管理层面需建立健全统一协调、分工明确的组织架构,加强各部门之间的协同配合,确保风险预警机制在各场景中有效落地。成立学校层面的生成式人工智能应用风险管理领导小组,由学校领导牵头,成员包括教务处、科研处、学生处、保卫处、信息技术中心等相关部门负责人,负责统筹规划全校生成式人工智能应用的风险防控工作,制定整体防控策略,协调解决跨部门的风险问题。建立常态化的沟通协调机制,通过定期会议、信息共享平台等方式,加强各专项工作小组之间以及与领导小组之间的沟通交流,及时通报风险信息,协调防控措施。

3.4教育维度：提升全员素养,筑牢思想防线

教育层面需面向全校师生开展系统的教育培训,提升其对生成式人工智能应用风险的认知和防范能力,从思想根源上降低风险发生的概率。

制定分层分类的教育培训计划,针对教师、学生、管理人员等不同群体,设计不同的培训内容和方式。对于教师,重点培训生成式人工智能在教学、科研中的合理应用方法,以及识别和防范教学内容同质化、科研不端等风险的能力；对于学生,着重培养其在使用生成式人工智能进行学习时的自主判断能力、信息素养和学术诚信意识,引导其正确看待智能工具,避免过度依赖；对于管理人员,加强对生成式人工智能应用管理政策、风险防控流程的培训,提高其管理和决策水平。

4 总结

首先,本文阐述了生成式人工智能的发展现状,其技术不断迭代且应用规模持续扩大,在高等教育领域展现出巨大潜力,但也存在数据、算法、系统层面的固有风险。其次,针对AIGC赋能高等教育的六大应用场景：助教、助学、助评、助育、助研、助管,深入剖析各场景下的具体风险。并针对这些风险,从制度、技术、管理、教育四个维度构建了风险预警机制。通过以上研究,本文明确了AIGC在高等教育应用中的机遇与风险,所构建的预警机制为高等教育领域规范、安全地应用生成式人工智能提供了切实可行的路径。未来,还需持续关注技术发展与应用新情况,不断完善风险预警机制,以实现技术赋能与风险防控的动态平衡,推动高等教育健康发展。

[基金项目]

乌鲁木齐职业大学2025年校级科研课题,项目名称《生成式人工智能大模型应用风险的识别与预警机制构建研究》,项目编号：2025ZC002。

[参考文献]

[1]戴琼海.大模型技术:变革、挑战与机遇[J].中国科学基金,2023,37(05):713.

[2]李婧.基于知识图谱与强化学习的双路径个性化在线学习资源推荐方法研究[J].信息系统工程,2025,(05):132-135.

[3]曾华,刘文娟.基于人工智能的个性化智能辅导系统设计与实现[J].智能物联技术,2025,57(03):103-106.

[4]赵博,王海福.生成式人工智能赋能高等教育的价值、风险与纾解路径[J].人工智能,2024,(01):100-107.

[5]褚如心.生成式人工智能赋能高等教育发展的应用、风险及破解之道[J].青年学报,2024,(04):60-66.

[6]梁远高.论人工智能大模型训练数据风险的分层规制[J].郑州大学学报(哲学社会科学版),2025,58(03):61-67+144.

[7]徐琦,孙智蒲.大模型智能体幻觉难题:成因、风险与应对[J].中国传媒科技,2025,(05):7-14.

[8]郭园方,余梓彤,刘艾杉,等.多模态大模型安全研究进展[J].中国图象图形学报,2025,30(06):2051-2081.

[9]国家互联网信息办公室发布《生成式人工智能服务已备案信息》公告[EB/OL].

[10]中国互联网络信息中心在京发布第56次《中国互联网络发展状况统计报告》[EB/OL].

作者简介：

姚苗(1993--),女,汉族,江苏泰州人,硕士研究生,讲师,研究方向：人工智能大模型、教育技术。