 DOI: https://doi.org/10.12345/bdai.v7i2.37668

Research on the Current Status and Enhancement Pathways of College Students' AI Security Literacy

Dandan Wu Wu Zhu Xiaochun Li

Hangzhou Medical College, Hangzhou, Zhejiang, 311300, China

Abstract

With the widespread application of domestic generative AI (AIGC) tools such as DeepSeek, Wenxin Yiyang, Tongyi Qianwen, and Doubao in campus settings, potential risks like data breaches and academic ethical violations have become increasingly prominent as students utilize AI for learning, research, and practical activities. To systematically investigate the current state and challenges in cultivating AI security literacy among college students, this study constructs a "Technology Acceptance-Risk Perception-Behavioral Adjustment" literacy development model based on survey findings. From a student-centered perspective, it proposes enhancement pathways including establishing a collaborative campus cultivation system, strengthening university-industry resource integration, and developing immersive scenario-based simulation tools. These findings provide theoretical references and practical guidance for universities to implement targeted AI security education.

Keywords

College students; Generative artificial intelligence; Safety literacy; Artificial intelligence education; Medical schools; Empirical research

高校学生人工智能安全素养现状调查与提升路径研究

吴丹丹 朱武 李晓春

杭州医学院，中国·浙江 杭州 311300

摘 要

随着 DeepSeek、文心一言、通义千问、豆包等国产生成式人工智能（AIGC）工具在校园场景的广泛应用，大学生借助人工智能辅助学习、科研与实践的过程中，数据泄露、学术伦理失范等潜在风险愈发突出。为系统探究当前高校学生人工智能安全素养的现实状况与培育困境，本研究基于调研的问题，构建“技术接受—风险感知—行为调适”素养生成模型，并从学生视角出发，提出构建校园协同培育体系、强化校企资源整合、开发沉浸式情景模拟教育工具等提升路径，为高校开展针对性的人工智能安全教育提供理论参考与实践借鉴。

关键词

大学生；生成式人工智能；安全素养；人工智能教育；医学院校；实证研究

1 引言

1.1 研究背景

在数字智能技术快速迭代的背景下，生成式人工智能已深度融入高校教学、科研与学生日常学习过程，成为辅助各类文稿写作和代码编写、论文优化、课件制作等任务的重要工具，为学生提升学习与科研效率提供了显著技术红利。

【基金项目】2025 年浙江省大学生创新创业训练计划项目“大学生人工智能安全素养的动态演化机理与协同培育机制——基于质性研究的分析”（项目编号：S202513023031）。

【作者简介】吴丹丹（2004—），女，中国浙江湖州人，在读本科，从事大学生安全素养研究。

然而，技术应用的普及也伴随着各类安全风险的滋生：国外出现律师使用某大模型撰写庭审辩护词时引用虚构案例遭受法律制裁的事件，国内高校中亦存在学生将未公开发表的实验数据、研究隐私数据直接上传至云端大模型，导致核心信息泄露的隐患。习近平总书记强调，要确保人工智能“安全、可靠、可控”。大学生作为人工智能的重度使用群体，其技术应用能力与安全素养、伦理意识的协同发展，不仅关系到个人学术诚信与信息安全，更对人工智能技术在教育领域的健康发展具有重要意义。因此，系统探究大学生人工智能安全素养的现状与培育困境，提出科学有效的提升路径，成为当前高校教育面临的重要课题。

1.2 研究意义

1.2.1 理论意义

本研究聚焦生成式人工智能时代高校学生安全素养培育的新问题，通过实证调研揭示学生安全认知与行为之间的

内在矛盾，构建“技术接受—风险感知—行为调适”的素养生成模型，丰富人工智能教育与安全素养培育的理论体系，为相关领域的后续研究提供新的分析框架。

1.2.2 实践意义

研究基于某医学院校的实证数据，识别当前高校人工智能安全教育存在的突出问题，提出针对性的提升策略，可为高校优化人工智能安全教育内容、创新教育形式、完善管理机制提供实践指导，助力大学生在享受人工智能技术便利的同时，筑牢安全防线，实现技术应用与安全素养的协同发展。

1.3 文献综述

现有研究多聚焦于传统网络安全教育，主要围绕防病毒、防电信诈骗等外部攻击防御展开，而针对生成式人工智能带来的内生性风险的研究相对匮乏。部分学者关注到人工智能在教育领域的伦理问题，但多以理论探讨为主，缺乏基于学生实际使用场景的实证分析。此外，现有研究尚未形成系统的人工智能安全素养生成机制模型，对安全教育的针对性与实效性提升缺乏有效支撑。基于此，本研究通过问卷调查与深度访谈相结合的方式，开展实证研究，填补现有研究的实践空白。

2 研究方法

2.1 调研对象

本研究选取某医学院校学生作为调研对象，采用分层随机抽样方法，覆盖医学类、医学技术类（含药学、医学检验技术、卫生检验与检疫）、工学类三大类专业，样本构成比例分别为 45%、35%、20%，确保样本的代表性。

2.2 调研工具

2.2.1 问卷调查

参考相关研究成果，设计《高校学生人工智能使用与安全素养调查问卷》，内容涵盖认知维度、行为维度、需求维度三个模块。问卷经专家评审修订后，采用电子问卷形式发放，共发放 300 份，回收有效问卷 264 份，有效回收率为 88%。

2.2.2 深度访谈

选取 12 名不同年级、专业的学生进行一对一深度访谈，访谈时长为 30-45 分钟 / 人，访谈内容围绕人工智能使用场景、安全风险经历、安全教育体验、政策需求等展开。访谈结束后，对访谈录音进行转录，整理形成近 3 万字的访谈实录，为研究提供质性分析资料。

2.3 数据处理

采用统计软件对问卷数据进行描述性统计、相关性分析等量化分析；对访谈实录采用编码分析方法，提取关键信息与核心观点，与问卷数据进行交叉验证，确保研究结论的科学性与可靠性。

3 研究结果与分析

通过对问卷数据与访谈资料的综合分析，当前医学院校学生人工智能安全素养培育存在三大突出矛盾，具体如下：

3.1 知行分离：安全认知与行为实践的严重脱节

受医学伦理教育的长期影响，医学生对患者隐私保护的认知水平普遍较高，但在人工智能工具使用过程中，认知与行为呈现显著倒挂现象。

量化数据显示，在认知维度调查中，95.2% 的受访者明确表示 “绝对不能泄露患者隐私数据”，对隐私保护的重要性形成高度共识。然而在行为维度调查中，当被问及“为辅助科研统计或病历润色，是否将包含患者姓名、住院号或未脱敏临床数据上传至人工智能工具” 时，38.5% 的学生选择“偶尔会”或“经常会”，存在明显的违规操作行为。

进一步的动机分析表明，导致这一现象的首要原因是“科研 / 临床压力大，优先追求效率”，其次是“认为人工智能仅为工具，不会留存数据”。访谈中，某医学信息工程专业大三学生提到：“专业常需通过代码分析医院信息化应用场景，某次实验课作业中，我的 Python 语言代码持续报错且临近截止时间，便将包含数十个样本信息的代码直接发送至文心一言调试。虽知晓医学伦理规范，但当时仅希望尽快完成程序调试，且主观认为国产生成式人工智能模型安全性较高。”这一表述直观反映了认知与行为脱节的现实场景。

3.2 治理碎片化：临床实践与学术规范的认知撕裂

调查发现，学校不同管理部门及临床教学环节对人工智能使用的规范要求存在差异，缺乏统一的指导标准，导致学生在医院实习与学校科研两大场景中面临认知冲突，难以准确把握合规边界。

问卷数据显示，68.2% 的学生认为当前关于人工智能使用的规定“界限模糊”；在“遇到人工智能使用困惑时向谁求助”的问题中，仅有 14.5% 的学生选择“学校 / 医院官方渠道”，反映出官方指导渠道的有效性不足。更值得关注的是，37.8% 的学生表示会通过非官方渠道寻找“功能更强大”的人工智能工具辅助诊断或文献综述，这一行为显著增加了医疗数据外泄、学术不端的风险。

某医学影像学专业大五学生在访谈中提及：“实际使用中存在明显的场景割裂感。在附属医院实习时，带教老师会演示如何运用人工智能辅助诊断系统快速阅片，甚至鼓励使用人工智能生成三维重建图，称其为行业发展趋势；但返回学校开展科研、撰写论文时，教学秘书会发布通知严查人工智能生成率，即便使用人工智能润色英文摘要也可能被判定为学术不端。目前使用人工智能工具如同规避监管，难以明确合规红线。” 这一表述清晰呈现了治理碎片化带来的认知困惑。

3.3 评估滞后：理论知识与实战能力的发展失衡

传统人工智能安全教育与评估方式以理论记忆为主，缺乏对实战应用能力的培养与考察，导致学生虽掌握基础理论知识，但在面对真实场景中的人工智能安全风险时，应对能力不足。

本研究在问卷中设置理论测试与实战测试两组题目进行对比分析：理论测试环节，在回答“什么是医疗数据脱敏”的选择题时，98.5% 的学生能够选出正确答案，表明学生对基础概念的掌握程度较高；实战测试环节，问卷植入高仿真的人工智能生成虚假医学会议邀请函与伪造临床试验数据图表，结果显示仅有23.5%的学生选择“核实会议官网”或“检查数据源头”进行验证，54.3% 的学生仅凭直观判断便轻信人工智能生成的专业术语与图表，风险识别能力薄弱。

相关性分析结果显示，“医学伦理知识得分”与“人工智能安全实战得分”的相关系数仅为 0.18，呈现弱相关关系，表明传统医学伦理教育并未有效覆盖人工智能时代所需的识假防骗、风险应对等实战能力。某临床医学专业大四学生在访谈中提到：“入学时参加过网络安全考试，题型均为选择题，以理论记忆为主。但前段时间收到一封伪装成某医学期刊编辑部的邮件，称投稿需要补充数据，邮件内容专业且符合医学逻辑，后续才知晓是人工智能生成的钓鱼邮件，当时险些将原始数据库发送出去。书本上学到的防诈骗知识，在面对这类高仿真人工智能风险时，难以快速做出有效应对。”

4 讨论与建议

基于研究发现的三大矛盾，结合“技术接受—风险感知—行为调适”的素养生成机制，从学生视角出发，提出以下针对性提升路径：

4.1 立足素养生成规律，优化安全教育内容设计

通过对访谈资料的深度分析，本研究发现学生人工智能安全素养的形成需经历三个关键阶段：第一阶段为技术接受阶段，学生受感知有用性驱动，核心关注人工智能工具对学习、科研效率的提升作用；第二阶段为风险感知阶段，这是素养形成的关键转折点，学生通过亲身经历人工智能使用失误或目睹相关风险事件，安全意识被有效激活；第三阶段为行为调适阶段，在风险感知的驱动下，学生主动学习数据脱敏、交叉验证等安全技能，形成稳定的安全使用习惯。

基于这一规律，人工智能安全教育内容不应局限于理论宣讲，需强化真实风险案例的教学价值：选取脱敏处理后的人工智能使用失误案例、数据泄露事件、学术伦理失范案例等，通过案例分析、情景讨论等形式，让学生直观感受风险危害，激活风险感知意识；同时，补充数据脱敏操作规范、人工智能生成内容验证方法、学术合规使用边界等实操性知识，实现从理论认知到行为实践的转化。

4.2 强化协同治理，构建统一规范的培育环境

4.2.1 推进校内治理一体化

建议学校成立人工智能教育与管理指导委员会，统筹教务、学工、信息中心、科研管理等相关部门，制定统一、清晰的人工智能使用指导规范，明确学术研究、临床实践中人工智能使用的合规边界、禁止性条款及白名单工具；借鉴部分高校“创新激励 + 底线约束”的管理模式，鼓励学生在合规前提下开展技术应用创新，同时强化违规行为的监督与惩戒，实现疏堵结合。

4.2.2 深化校企资源协同

鉴于高校人工智能安全案例更新滞后、实战教学资源不足的问题，建议引入百度、阿里、科大讯飞等国内头部人工智能企业的安全专家资源，通过专题讲座、模拟攻击演练等形式，向学生展示最新的人工智能安全风险类型、攻击手段及防御方法；建立校企合作实践基地，为学生提供真实场景下的人工智能安全实操训练机会，提升安全教育的针对性与实战性。

4.3 创新教育形式，开发沉浸式实战训练工具

针对传统教育评估方式僵化、实战能力培养不足的问题，建议开发基于情景模拟的沉浸式教育工具，构建沙箱式试错环境，让学生在虚拟场景中强化安全技能训练：

4.3.1 隐私保卫战模块

模拟临床数据、科研数据上传场景，系统自动检测上传内容是否包含患者姓名、住院号、身份证号等敏感信息。若学生未对数据进行脱敏处理便直接上传，系统即时触发数据泄露预警提示，并展示风险传导路径与潜在危害，通过即时反馈强化学生的数据脱敏意识与操作习惯。

4.3.2 内容甄别模块

提供人工智能生成的学术论文片段、医学文献摘要、会议通知等内容，要求学生通过交叉验证、源头核查等方式，识别其中的虚构事实、虚假引用及侵权内容；工具内置详细的甄别思路解析与正确操作指引，帮助学生掌握人工智能生成内容的验证方法，提升风险识别能力。

通过交互式、沉浸式的训练模式，让学生在安全可控的虚拟环境中试错学习，将安全使用技能内化为稳定的行为习惯。

5 结语

在人工智能技术持续渗透教育领域的背景下，高校应主动回应技术发展带来的安全挑战，构建科学完善的人工智能安全素养培育体系，助力大学生在数字智能时代实现技术应用能力与安全素养的协同发展，为人工智能技术的健康可持续发展奠定坚实基础。

参考文献

[1] 乔雪峰.从工具赋能到智能协同：生成式人工智能驱动的教育模式转型[J].南京社会科学,2025,(01):126-134.DOI:10.15937/j.cnki.issn1001-8263.2025.01.013.

[2] 孙立会,周亮.生成式人工智能赋能教育变革的逻辑——基于新质生产力的视角[J].教育研究,2024,45(10):38-49.

[3] 辛征,郝丽丽,侯传晶,等.AI时代大学生国家安全素养培养策略研究[J].电气电子教学学报,2025,47(01):87-90.

[4] 郑腾.人工智能赋能高校治理的背景阐释、潜在风险与改善路径[J].黑龙江高教研究,2025,43(09):1-7.DOI:10.19903/j.cnki.CN23-1074/G.2025.09.001.