

AI技术应用于高校图书馆科研支持服务的 伦理风险及应对策略*

周鑫^{1,2} 施未晔¹ 郑江杰³ 夏凯丽³

(1. 南京工业大学经济与管理学院, 南京 211816; 2. 江苏省数据工程与知识服务重点实验室, 南京 210023;
3. 江苏省科学技术情报研究所, 南京 210042)

摘要: 随着人工智能 (Artificial Intelligence, AI) 技术在高校图书馆科研支持服务中的深度渗透, 数据管理、智能检索等场景的服务效率显著提升, 但伦理风险也随之凸显。本研究旨在明确AI技术应用于高校图书馆科研支持服务的伦理风险类型与成因, 构建科学可行的应对策略体系。采用文献调研法梳理AI技术与高校图书馆科研支持服务融合现状及伦理研究进展, 通过总结国内外高校图书馆的实践经验与风险暴露情况, 结合德尔菲法对风险维度与应对措施进行专家论证。研究发现, AI伦理风险主要集中于数据隐私、算法公平、学术诚信、责任界定等四大维度, 其成因涉及技术特性、制度设计、主体素养等多重因素。基于此提出“制度规范-主体赋能-监管协同-技术优化”的四维应对策略, 为高校图书馆防范AI技术应用于科研支持服务的伦理风险提供理论支撑与实践参考, 助力高校图书馆智能化发展。

关键词: 高校图书馆; 科研支持服务; 人工智能应用; 伦理风险

中图分类号: G252 **DOI:** 10.3772/j.issn.1673-2286.2025.12.002

引文格式: 周鑫, 施未晔, 郑江杰, 等. AI技术应用于高校图书馆科研支持服务的伦理风险及应对策略[J]. 数字图书馆论坛, 2025, 21(12): 8-15.

在数字化转型和智能化发展的时代浪潮下, 人工智能 (Artificial Intelligence, AI) 技术正在深刻重塑高校图书馆的服务形态与功能定位^[1-4]。目前, 国内外高校图书馆正积极探索AI技术的多元化应用场景, 例如: 在智能文献检索与推荐方面, 基于深度学习的语义检索系统能够精准理解用户的科研需求, 提供个性化的文献推荐服务^[5-6]; 在科研数据管理领域, AI技术实现了对海量科研数据的智能分类、标注与挖掘^[7-8]; 在学术评价与知识发现层面, 机器学习算法被用于识别学术前沿、预测研究趋势、评估学术影响力^[9-10]; 在参考咨询服务中, 智能问答系统和虚拟馆员能够7×24小时为科研人员提供即时支持^[11]。然而, AI技术在高校图书馆的深度应用, 也引发了一系列不容忽视的伦理风险问

题, 成为高校图书馆智能化发展的重要挑战。

当前, 学术界关于AI伦理风险的讨论主要围绕3个维度展开。①技术内生性风险。首先, 数据偏见与算法歧视风险尤为突出, 如果训练数据包含已有社会偏见, 算法会将其固化甚至放大, 导致招聘、信贷、司法等领域产生系统性不公平结果, 加剧社会排斥。其次, 算法的“黑箱”特性导致决策过程缺乏透明性与可解释性, 尤其对于深度学习系统, 其复杂的内部运作机制难以追溯, 削弱了问责基础, 危及程序正义与个人知情权。最后, 价值对齐问题变得日益严峻, 如何确保高度自主的AI系统的目标与人类社会的整体利益、安全及伦理规范保持一致, 已成为前沿安全研究的重点问题。②社会应用性风险。一是对劳动就业结构的颠覆性影响, 自

收稿日期: 2025-11-16

*本研究得到江苏省社会科学基金青年项目“高水平科技自立自强背景下江苏高校图书馆服务韧性测度研究”(编号: 24TQC008)资助。

动化与智能化可能导致大规模结构性失业, 加剧经济不平等, 亟待社会政策与教育体系的适应性调整。二是隐私侵蚀与监控强化, 人脸识别、行为预测等技术的滥用, 使得全景监控成为可能, 对个人隐私与自由构成根本威胁, 并可能异化为社会控制工具。三是信息生态与认知安全风险, 深度伪造技术、个性化算法推荐易催生虚假信息泛滥、舆论操纵与认知茧房, 侵蚀公共对话与社会信任基础。③治理结构性风险。现有法律与伦理框架难以匹配AI的迭代速度, 出现“治理赤字”, 甚至当AI系统造成损害时, 其责任在开发者、运营者、使用者或是算法自身之间难以清晰界定, 形成责任真空。

高校图书馆是科研全链条的智能引擎, 以权威资源保障为基、嵌入式情报服务为脉、开放科学平台为翼, 贯通选题导航、数据支撑、成果管理和转化推广, 已成为国家科技创新体系中不可替代的战略支点。面对AI技术快速发展与伦理风险治理滞后之间的矛盾, 高校图书馆亟须建立系统化的伦理风险防控体系。因此, 本研究在厘清高校图书馆科研支持服务中的AI伦理风险定义的基础上, 通过系统性识别AI技术在图书馆科研支持服务中的伦理风险类型与形成机制, 构建多维度的风险评估体系, 并提出兼具理论深度与实践价值的应对策略框架, 为高校图书馆在智能化转型过程中实现技术创新与伦理规范的协调发展提供理论指导与实践参考。

1 研究设计

AI技术应用于高校图书馆科研支持服务的伦理风险是指高校图书馆借助AI技术为科研主体提供全流程服务时, 由AI技术特性、数据与价值偏差、权责界定模糊等因素引发的违背学术伦理、侵害科研主体权益、破坏学术生态或弱化图书馆公益属性的潜在负面风险与现实危害。其研究内容属于新兴交叉领域, 既需要从图书馆学视角分析高校图书馆科研支持服务场景的特性, 又需要从计算机科学视角分析技术风险的本质, 还需要从法学视角分析隐私保护及责任界定等。

单一学科研究者难以全面覆盖研究的所有维度, 易出现认知偏差, 而德尔菲法^[12]凭借跨领域专家协同的特性能够适配这一研究需求, 通过多轮专家反馈逐步达成共识, 最终获得集体判断结论。

德尔菲法是根据系统的程序, 采纳匿名方式征询专家意见的方法, 即参与调研的专家之间不允许横向

交流或相互讨论, 只能与调研人员进行沟通, 以保证每位专家独立发表见解。通过多次发放问卷获取反馈的形式, 汇总参与调研专家的意见, 逐步使各位专家的意见达成共识, 最终对专家的意见进行汇总分析^[13]。本研究选择18名专家参与德尔菲法调研分析, 符合15~20人标准^[14-15], 其中高校图书馆馆员6名、AI伦理研究人员3名(含法学方向专家2名)、信息资源管理学研究人员4名、高校科研管理人员3名、计算机科学研究人员2名。调查时间为2025年9月28日—2025年10月26日, 进行3轮调查, 每轮问卷回收率为100%。

2 研究过程

通过文献调研、访谈等形式, 经过课题小组讨论, 初步构建了包含3个一级指标、14个二级指标的评估指标体系。

(1) 第一轮调查主要对初步构建的评估指标进行评判、补充和修正, 最终形成相对完整的指标体系。调查主要包含3个部分: ①评判伦理风险识别的各二级指标与高校图书馆科研支持服务场景是否贴合(如贴合、不贴合), 并补充未涵盖的伦理风险类型和指出与主题无关的指标; ②评判风险成因分析的各二级指标与风险的关联性(如强相关、中等相关、弱相关、不相关), 并补充关键成因的类型和合并重复成因; ③评判应对策略评估的各二级指标与风险的匹配度(如高度匹配、中等匹配、低匹配、不匹配), 并补充操作性强的应对策略和剔除不可行的策略。

(2) 第二轮调查主要对第一轮筛选后的指标进行量化评分, 形成专家的初步共识。调查主要包含3个部分: ①伦理风险识别二级指标的量化评分, 主要针对“重要性”和“发生概率”进行评分(1~5分制, 1=极低, 2=较低, 3=中等, 4=较高, 5=极高); ②风险成因分析二级指标的量化评分, 主要针对“影响程度”和“可干预性”进行评分(1~5分制, 1=极微弱/极难干预, 2=较微弱/较难干预, 3=中等, 4=较显著/较易干预, 5=极显著/极易干预); ③应对策略评估二级指标的量化评分, 主要针对“可行性”“有效性”“实施成本”进行评分(1~5分制, 1=极不可行/极无效/极低, 2=较不可行/较无效/较低, 3=中等, 4=较可行/较有效/较高, 5=极可行/极有效/极高)。

(3) 第三轮调查主要针对第二轮评分差异较大的指标(变异系数大于0.25), 邀请专家补充论证, 最终形

成共识指标^[16]。

3 指标体系构建

经过第一轮调查,伦理风险识别的各个二级指标均保留,18位专家均给出了“贴合”评价,并新增一个二级指标“知识产权侵权风险”,具体表现为AI生成内容侵犯他人著作权、AI爬虫抓取学术资源等,16位专家支持补充。2位专家建议纳入“数据隐私泄露风险”指标,最终将其独立列为一个二级指标。风险成因分析的各个二级指标均获得“强相关”评价,无新增指标,专家反馈“已覆盖核心影响因素,无需补充”。应对策略评估的各个二级指标均获得“高度匹配”评价,此外17位专家支持补充“伦理审查前置策略”作为新增二级指标,具体措施为AI科研支持服务上线前开展伦理合规审查、建立AI项目伦理备案制度。经过第一轮专家调查论证,最终构建包含3个一级指标、16个二级指标的评估指标体系(见表1)。

表1 评估指标体系

一级指标	二级指标
伦理风险识别	数据隐私泄露风险(生物识别数据违规、科研数据外泄等)
	算法偏见与歧视风险(学科/性别/种族歧视等)
	学术诚信异化风险(AI生成内容造假、虚构参考文献等)
	责任界定模糊风险(技术故障、侵权责任划分等)
	技术安全风险(AI爬虫攻击、系统中断等)
	学术独立性弱化风险(过度依赖AI导致研究能力退化)
风险成因分析	知识产权侵权风险(侵犯他人著作权、爬虫抓取学术资源用于商业训练)
	技术特性成因(算法黑箱、技术漏洞等)
	制度设计成因(伦理规范缺失、权责划分不清等)
	主体素养成因(图书馆人员伦理素养不足、研究者AI使用认知欠缺等)
应对策略评估	监管机制成因(多主体协同监管缺位、检测技术不成熟等)
	技术优化策略(算法透明化、数据加密等)
	制度规范策略(伦理准则制定、AI使用声明制度等)
	主体赋能策略(人员伦理培训、研究者引导等)
	伦理审查前置策略(伦理合规审查、AI项目伦理备案制度)
	监管协同策略(多主体监管、风险预警机制等)

4 量化结果分析

严格遵循“多轮反馈-认知微调-共识收敛”的核心逻辑,向专家征求反馈意见,并汇总相关调查结果。采用

肯德尔 W 值和变异系数 C_v 检验专家协调程度^[17]。 W 值越大,表明专家共识度越高^[18]。 W 值 ≥ 0.50 表示高度共识, W 值 ≥ 0.80 表示完全共识,表明专家对各指标的优先级排序形成强一致性共识。根据统计学标准, $C_v \leq 0.25$ 表明专家对单个指标的量化评分分散度低,量化评估结果具有较高的一致性,形成共识。 C_v 值聚焦单个指标的评分精度,肯德尔 W 值侧重指标的相对重要性排序,二者协同验证,形成“点-面”结合的共识检验体系。

4.1 伦理风险识别指标优先级排序及特征分析

(1) 权重排序与层级划分。伦理风险识别指标变异系数汇总表结果如表2所示。基于重要性均值与发生概率均值相乘的综合权重计算结果,伦理风险识别指标呈现清晰的三级层级结构(以下只计算 $C_v \leq 0.25$ 的伦理风险识别指标综合权重)。**①高优先级风险:**数据隐私泄露风险(两轮调查综合权重计算结果分别为19.26和18.74)、学术诚信异化风险(两轮调查综合权重计算结果分别为18.53和19.00),其重要性均值介于4.39~4.50,发生概率均值介于4.22~4.28,且 W 值均为0.78(高度共识),反映此类风险兼具高重要性与高发生概率,是高校图书馆AI伦理风险防控的核心靶点。**②中优先级风险:**知识产权侵权风险(两轮调查综合权重计算结果分别为14.31和14.34)、算法偏见与歧视风险(第二轮调查 $C_v > 0.25$,不计算权重;第三轮调查综合权重计算结果为11.83),重要性均值介于3.44~4.22,发生概率均值介于3.39~3.44, W 值为0.65~0.80,表明此类风险虽发生概率略低于高优先级风险,但对学术生态与知识公平的潜在影响显著,须纳入常规防控体系。**③低优先级风险:**技术安全风险(两轮调查综合权重计算结果分别为10.01和9.70)、责任界定模糊风险(两轮调查综合权重计算结果分别为7.25和7.23),重要性均值介于3.11~3.22,发生概率均值介于2.28~3.11,其中责任界定模糊风险虽发生概率较低,但 W 值达0.90(完全共识),且涉及法律权责划分的核心问题,仍须作为辅助防控指标纳入核心风险清单。

(2) 风险本质与行业特征。数据隐私泄露风险作为权重最高的核心风险,其 C_v 值仅为0.11,反映专家对该风险的紧迫性形成高度共识,本质是AI技术的数据依赖性与科研数据敏感性的叠加效应,表现为生物识别数据违规采集、核心实验数据云端泄露等场景,触及科研诚信与个人信息保护的双重底线。学术诚信异

化风险的权重仅次于数据隐私泄露风险, 凸显AI技术降低学术不端门槛的现实困境, 核心表现为AI生成内容造假、虚构参考文献等问题, 本质是技术赋能下学

表2 伦理风险识别指标变异系数汇总

二级指标	调研维度	第二轮调查	第三轮调查	波动幅度
数据隐私泄露风险	重要性均值	4.50	4.44	-0.06
	重要性标准差	0.51	0.50	-0.01
	重要性 C_v	0.11	0.11	0.00
	发生概率均值	4.28	4.22	-0.06
	发生概率标准差	0.46	0.45	-0.01
	发生概率 C_v	0.11	0.11	0.00
学术诚信异化风险	重要性均值	4.39	4.44	+0.05
	重要性标准差	0.49	0.48	-0.01
	重要性 C_v	0.11	0.11	0.00
	发生概率均值	4.22	4.28	+0.06
	发生概率标准差	0.42	0.43	+0.01
知识产权侵权风险	发生概率 C_v	0.10	0.10	0.00
	重要性均值	4.22	4.17	-0.05
	重要性标准差	0.62	0.61	-0.01
	重要性 C_v	0.15	0.15	0.00
	发生概率均值	3.39	3.44	+0.05
算法偏见与歧视风险	发生概率标准差	0.49	0.48	-0.01
	发生概率 C_v	0.14	0.14	0.00
	重要性均值	3.06	3.44	+0.38
	重要性标准差	0.87	0.50	-0.37
	重要性 C_v	0.28	0.14	-0.14
	发生概率均值	2.17	3.44	+1.27
技术安全风险	发生概率标准差	0.45	0.50	+0.05
	发生概率 C_v	0.21	0.14	-0.07
	重要性均值	3.22	3.17	-0.05
	重要性标准差	0.42	0.43	+0.01
	重要性 C_v	0.13	0.13	0.00
	发生概率均值	3.11	3.06	-0.05
责任界定模糊风险	发生概率标准差	0.47	0.46	-0.01
	发生概率 C_v	0.15	0.15	0.00
	重要性均值	3.11	3.17	+0.06
	重要性标准差	0.47	0.48	+0.01
	重要性 C_v	0.15	0.15	0.00
	发生概率均值	2.33	2.28	-0.05
学术独立性弱化风险	发生概率标准差	0.48	0.47	-0.01
	发生概率 C_v	0.21	0.21	0.00
	重要性均值	2.22	2.17	-0.05
	重要性标准差	0.42	0.43	+0.01
	重要性 C_v	0.19	0.20	+0.01
	发生概率均值	2.06	2.11	+0.05
	发生概率标准差	0.38	0.39	+0.01
	发生概率 C_v	0.18	0.18	0.00

术原创性标准与AI工具使用规范的适配滞后。算法偏见与歧视风险经第三轮调查后达成高度共识 (W 值为0.65), 重要性与发生概率均值均提升至3.44, 反映专家对学术资源获取公平性的关注, 其本质是训练数据偏差与算法黑箱特性导致的系统性歧视, 易引发学科发展失衡、少数群体权益受损等衍生问题。

4.2 风险成因分析指标影响机理与可干预性分析

(1) 综合权重与优先级排序。风险成因分析指标变异系数汇总结果如表3所示。基于影响程度均值与可干预性均值相乘的综合权重计算, 风险成因分析指标排序为: 制度设计成因(两轮调查综合权重计算结果分别为15.02和15.05)、主体素养成因(两轮调查综合权重计算结果分别为13.69和13.76)、监管机制成因(两轮调查综合权重计算结果分别为10.01和10.05)、技术特性成因(两轮调查综合权重计算结果分别为8.88和

表3 风险成因分析指标变异系数汇总

二级指标	调研维度	第二轮调查	第三轮调查	波动幅度	
制度设计成因	影响程度均值	4.22	4.17	-0.05	
	影响程度标准差	0.42	0.43	+0.01	
	影响程度 C_v	0.10	0.10	0.00	
	可干预性均值	3.56	3.61	+0.05	
	可干预性标准差	0.50	0.49	-0.01	
	可干预性 C_v	0.14	0.14	0.00	
	主体素养成因	影响程度均值	3.33	3.39	+0.06
主体素养成因	影响程度标准差	0.48	0.47	-0.01	
	影响程度 C_v	0.14	0.14	0.00	
	可干预性均值	4.11	4.06	-0.05	
	可干预性标准差	0.32	0.33	+0.01	
	可干预性 C_v	0.08	0.08	0.00	
	监管机制成因	影响程度均值	3.22	3.17	-0.05
	监管机制成因	影响程度标准差	0.42	0.43	+0.01
影响程度 C_v		0.13	0.13	0.00	
可干预性均值		3.11	3.17	+0.06	
可干预性标准差		0.47	0.46	-0.01	
可干预性 C_v		0.15	0.15	0.00	
技术特性成因		影响程度均值	4.00	4.06	+0.06
技术特性成因		影响程度标准差	0.56	0.55	-0.01
	影响程度 C_v	0.14	0.14	0.00	
	可干预性均值	2.22	2.17	-0.05	
	可干预性标准差	0.42	0.43	+0.01	
	可干预性 C_v	0.19	0.20	+0.01	

8.81)，且所有指标 W 值 ≥ 0.62 （高度共识），表明专家对风险成因的核心影响因素形成一致认知。

(2) 成因本质与作用机理。制度设计成因的综合权重最高，影响程度均值分别为4.22和4.17 ($C_v=0.10$)，可干预性均值分别为3.56和3.61 ($C_v=0.14$)，是风险的根源性成因，其本质是高校图书馆AI伦理治理的制度真空，表现为专项伦理规范缺乏、权责划分机制不清晰、AI使用审核流程缺失等，导致AI应用无据可依，放大了技术与主体层面的潜在风险。主体素养成因的可干预性均值分别为4.11和4.06 ($C_v=0.08$)，为所有成因中最高，综合权重排名第二，属于传导性成因，核心是图书馆员AI伦理素养与技术操作能力不足、研究者风险认知与规范意识欠缺，形成“技术风险-主体行为-风险爆发”的传导链条，是风险治理中性价比最高的干预靶点。监管机制成因的综合权重排名第三 (W 值为0.85)，两轮调查中影响程度均值分别为3.22和3.17，可干预性均值分别为3.11和3.17，属于保障性成因，其本质是跨部门协同监管缺位与风险监测技术不成熟，表现为图书馆、科研管理、信息安全等部门权责分割，缺乏专业化的AI伦理风险监测与应急处置机制。技术特性成因的影响程度均值分别为4.00和4.06 ($C_v=0.14$)，但可干预性均值分别为2.22 ($C_v=0.19$)和2.17 ($C_v=0.20$)，综合权重最低，属于基础性成因，核心是AI技术的黑箱性、迭代快、数据依赖性等固有特性，图书馆作为技术使用者缺乏主动干预能力，须通过技术选型把关与供应商协同降低风险。

(3) 成因关联性分析。制度设计成因与主体素养成因的综合权重占比超过60%，且二者 W 值均为0.62，表明专家普遍认为制度缺失与主体素养不足是引发风险的核心矛盾，二者相互强化，即制度空白导致主体行为缺乏规范引导，主体素养不足使现有制度难以有效落地，形成风险治理的双重瓶颈。

4.3 应对策略评估指标适配性与性价比分析

(1) 综合权重与优先级排序。应对策略评估指标变异系数汇总结果如表4所示。基于可行性均值与有效性均值乘积与实施成本均值之比的综合权重计算，最优应对策略排序为制度规范策略、主体赋能策略、监管协同策略、伦理审查前置策略、技术优化策略，所有策略的 W 值 ≥ 0.75 （高度共识），验证了策略选型的科

表4 应对策略评估指标变异系数汇总

二级指标	调研维度	第二轮调查	第三轮调查	波动幅度
制度规范策略	可行性均值	3.89	3.83	-0.06
	可行性标准差	0.54	0.53	-0.01
	可行性 C_v	0.14	0.14	0.00
	有效性均值	4.22	4.28	+0.06
	有效性标准差	0.42	0.43	+0.01
	有效性 C_v	0.10	0.10	0.00
	实施成本均值	2.11	2.17	+0.06
	实施成本标准差	0.32	0.31	-0.01
	实施成本 C_v	0.15	0.14	-0.01
主体赋能策略	可行性均值	4.00	4.06	+0.06
	可行性标准差	0.50	0.49	-0.01
	可行性 C_v	0.13	0.12	-0.01
	有效性均值	4.17	4.11	-0.06
	有效性标准差	0.38	0.39	+0.01
	有效性 C_v	0.09	0.09	0.00
	实施成本均值	2.22	2.17	-0.05
	实施成本标准差	0.42	0.43	+0.01
	实施成本 C_v	0.19	0.20	+0.01
监管协同策略	可行性均值	3.22	3.17	-0.05
	可行性标准差	0.42	0.43	+0.01
	可行性 C_v	0.13	0.13	0.00
	有效性均值	3.17	3.22	+0.05
	有效性标准差	0.49	0.48	-0.01
	有效性 C_v	0.15	0.15	0.00
	实施成本均值	3.06	3.00	-0.06
	实施成本标准差	0.24	0.25	+0.01
	实施成本 C_v	0.08	0.08	0.00
伦理审查前置策略	可行性均值	3.11	3.17	+0.06
	可行性标准差	0.47	0.46	-0.01
	可行性 C_v	0.15	0.15	0.00
	有效性均值	3.17	3.11	-0.06
	有效性标准差	0.49	0.50	+0.01
	有效性 C_v	0.15	0.16	+0.01
	实施成本均值	3.06	3.11	+0.05
	实施成本标准差	0.24	0.23	-0.01
	实施成本 C_v	0.08	0.07	-0.01
技术优化策略	可行性均值	2.67	2.61	-0.06
	可行性标准差	0.48	0.49	+0.01
	可行性 C_v	0.18	0.19	+0.01
	有效性均值	3.11	3.17	+0.06
	有效性标准差	0.47	0.46	-0.01
	有效性 C_v	0.15	0.15	0.00
	实施成本均值	4.00	3.94	-0.06
	实施成本标准差	0.50	0.51	+0.01
	实施成本 C_v	0.13	0.13	0.00

学性与实操性。

(2) 策略适配性与实施逻辑。制度规范策略的综合权重最高(两轮调查结果分别为7.78和7.55), 有效性均值分别为4.22和4.28 ($C_v=0.10$), 可行性均值分别为3.89和3.83 ($C_v=0.14$), 实施成本均值分别为2.11 ($C_v=0.15$)和2.17 ($C_v=0.14$), 是风险治理的核心策略。该策略直接靶向制度设计成因, 通过制定AI科研支持服务伦理准则、声明制度、权责划分办法等规范性文件, 构建AI应用的制度框架, 实现对所有核心风险的系统性覆盖。主体赋能策略的综合权重仅次于制度规范策略(两轮调查结果分别为7.51和7.69), 可行性均值分别为4.00 ($C_v=0.13$)和4.06 ($C_v=0.12$), 实施成本均值分别为2.22 ($C_v=0.19$)和2.17 ($C_v=0.20$), 性价比突出。该策略聚焦主体素养成因, 通过开展图书馆员AI伦理与技术培训、编制研究者AI使用指南、建立咨询服务机制等方式, 快速提升主体风险认知与规范操作能力, 有效阻断风险传导链条。监管协同策略与伦理审查前置策略的综合权重介于3.22~3.33, 属于中期配套策略: 监管协同策略通过构建跨部门监管小组、部署风险监测工具, 弥补监管机制成因的短板; 伦理审查前置策略针对高风险AI应用(如核心数据管理工具)实施上线前伦理评估, 形成“事前防控-事中监测-事后处置”的闭环治理。技术优化策略的综合权重最低(两轮调查结果分别为2.08和2.10), 实施成本均值分别为4.00和3.94 ($C_v=0.13$), 可行性均值分别为2.67 ($C_v=0.18$)和2.61 ($C_v=0.19$), 属于长期补充策略。该策略聚焦技术特性成因, 通过数据加密、算法透明化、操作留痕等技术手段提升风险防控的技术支撑能力, 但受限于成本与技术, 需要在制度与主体层面治理见效后逐步推进。

(3) “策略-成因-风险”的适配性闭环。数据显示, 制度规范策略与主体赋能策略的综合权重占比超过60%, 且二者分别与制度设计成因、主体素养成因高度适配, 对应覆盖数据隐私泄露、学术诚信异化等高优先级风险。监管协同策略与伦理审查前置策略则针对性解决监管机制成因, 覆盖技术安全风险、算法偏见与歧视风险等较低级风险。技术优化策略作为补充, 适配技术特性成因, 为风险治理提供长期技术支撑, 形成“策略-成因-风险”的精准适配闭环。

5 应对策略分析

AI伦理风险主要集中于数据隐私、算法公平、学

术诚信、责任界定等四大维度, 其成因涉及技术特性、制度设计、主体素养等多重因素。本研究提出“制度规范-主体赋能-监管协同-技术优化”的四维应对策略, 为高校图书馆规范AI科研支持服务应用、防范伦理风险提供理论支撑与实践参考, 助力科研支持服务高质量与伦理合规性协同发展。

(1) 制度规范为核心抓手, 尽快构建伦理治理刚性框架。针对制度设计缺失这一根源性成因, 以标准化文件构建AI应用的行为边界, 应对数据隐私与学术诚信等核心风险。①制定专项伦理准则。明确伦理准则的三大核心条款, 即数据采集遵循“最小必要+知情同意”原则、AI生成内容标注工具参与度、算法应用预留人工干预接口。②建立权责划分机制。联合科研处、法务处出台AI科研支持服务伦理权责划分办法, 界定三方责任, 即图书馆承担工具选型伦理审核、数据安全主体责任, 科研用户承担规范使用AI、确保成果原创责任, 技术供应商承担算法透明化、漏洞修复责任, 明确风险发生后的追责流程。③推行使用声明制度。设计AI科研工具使用伦理声明书, 要求科研用户在使用AI文献检索、内容生成等工具前签署, 声明内容包括不滥用AI生成数据、不规避引用标注、接受伦理审查等, 声明书与科研项目立项、论文提交挂钩。

(2) 主体赋能为关键传导, 主动阻断风险行为链条。针对主体素养不足这一传导性成因, 采用“分层培训+精准指导”策略提升图书馆员与科研用户的风险认知。①图书馆员分层赋能体系。构建“基础层-进阶层-专家层”培训体系: 基础层(全体馆员)开展“AI伦理通识+数据安全操作”培训(如个人信息加密方法); 进阶层(技术服务馆员)开展“算法偏见识别+伦理审核流程”培训; 专家层(学科服务馆员)参与跨校伦理研讨, 具备定制化风险咨询能力。②科研用户精准指导机制。编制面向科研用户的AI工具使用指南, 按工具类型、风险点、操作规范分类说明, 在高校图书馆的官方网站设置“AI伦理咨询专栏”, 提供在线答疑与案例警示, 并针对新生、硕/博士研究生等群体开展“专题讲座+情景模拟”培训。

(3) 监管协同为保障支撑, 尽快构建全流程防控闭环体系。针对监管机制缺位的保障性成因, 通过“跨部门协同+技术监测”来实现“事前-事中-事后”全流程管控, 防范算法偏见与歧视风险和技术安全风险。①组建跨部门监管小组。由图书馆牵头, 联合科研处、信息安全中心、法务处成立AI伦理监管小组, 明确分

工,如科研处负责学术诚信审查、信息安全中心负责数据安全监测、法务处负责法律风险评估。②部署动态监测工具。引入数据流转追踪系统,使核心科研数据(如实验原始数据)的采集、存储、使用全流程留痕,对异常操作(如批量导出敏感数据)自动预警。针对AI推荐算法,部署偏见检测工具,定期检测学科资源推荐的均衡性。③建立应急处置机制。制定AI伦理风险的应急处置预案,明确对3类风险的处置流程:数据泄露事件须在24小时内启动溯源、通知受影响用户并上报主管部门;学术诚信争议须组织监管小组与专家评审,判定责任归属;算法偏见事件须暂停工具使用,联合供应商优化模型后重新上线。

(4)技术优化为长期补充,强化风险防控技术支撑能力。针对技术固有特性的基础性成因,结合图书馆技术能力实际,以“低成本适配+供应商协同”策略提升技术防控水平,作为核心策略的补充。①数据安全技术适配。对科研数据实施分级加密处理,核心数据(如涉密实验数据)本地加密存储,普通数据云端加密传输。引入匿名化处理工具,对用户检索行为、借阅数据等进行去标识化处理,避免个人隐私关联追溯。②算法透明化协同。在采购AI工具时,将算法可解释性作为核心指标,要求供应商提供算法原理说明书与偏见测试报告。对已采购工具,联合计算机学院开展二次优化,如为智能检索系统增加学科权重调节功能,由馆员人工校准偏见倾向。③操作留痕技术强化。为所有AI科研工具加装操作日志模块,记录用户的工具使用时间、功能选择、数据输入输出等信息,日志保存期限不少于3年,为学术诚信核查与责任界定提供技术依据。

6 结论与展望

AI技术的迅猛发展为高校图书馆科研支持服务迭代注入了强劲动力,但伦理风险的隐匿性与传导性也对学术生态安全构成潜在挑战,亟须构建风险可控、效能提升的伦理治理体系。本研究立足高校图书馆AI科研支持服务场景,以德尔菲法为核心研究工具,通过跨领域专家多轮共识收敛,揭示了“制度设计缺失-主体素养不足”的核心成因链条,并构建了“制度规范-主体赋能-监管协同-技术优化”的四维应对体系,形成“风险识别-成因解析-策略构建”的研究闭环。但由于德尔菲法依赖的专家群体在学科背景上多样性不足,研究结论可能具有一定的局限性,将在后续的一系列研究中予以突破。

综上所述,高校图书馆AI伦理风险防控不是技术应用的枷锁,而是保障科研支持服务高质量发展的基石。本研究构建治理框架与实践路径,旨在为AI技术与高校图书馆科研支持服务的深度融合筑牢伦理防线,助力高校科研生态的健康演进与创新活力的持续迸发。

参考文献

- [1] 王戈非. 基于大语言模型的高校图书馆智能咨询服务发展策略[J]. 图书馆工作与研究, 2025, (7): 80-87.
- [2] 李雪, 林晓欣, 吕采威, 等. 数智启航 向新求质: 人工智能赋能图书馆新一轮高质量发展: “第十八届图书馆管理与服务创新论坛”综述[J]. 大学图书馆学报, 2025, 43 (6): 123-128.
- [3] 刘睿琳, 刘文科, 张新雨, 等. 全球视野下AIGC赋能高校图书馆服务创新现状与未来路径研究[J]. 图书馆杂志, 2025, 44 (11): 48-63.
- [4] 秦奋, 宋妙茹, 高健, 等. 数智技术赋能高校图书馆信息服务转型的实践路径与展望[J]. 图书馆工作与研究, 2025 (7): 40-52.
- [5] 贺轩. 生成式人工智能视域下高校图书馆智慧阅读推广服务研究[J]. 图书馆工作与研究, 2025 (10): 104-112.
- [6] 牛悦, 郝保权, 赵志艳. 基于大模型的高校图书馆个性化资源推荐系统构建与实践研究[J]. 新世纪图书馆, 2025 (7): 66-73.
- [7] 伏安娜, 汪东伟, 程蕴涵, 等. 中国高校图书馆研究数据管理服务十年实践与思考[J]. 图书情报工作, 2025, 69 (4): 23-33.
- [8] 王晓鹏. 数智赋能科研数据管理特点与启示: 以英国高校图书馆为例[J]. 新世纪图书馆, 2024 (11): 87-93.
- [9] 巫芯宇. AIGC赋能未来学习中心建设: 态势分析、实践瓶颈与进路探索[J]. 图书馆, 2025 (10): 108-114.
- [10] 王营盈. 基于大语言模型的高校图书馆嵌入式知识服务研究[J]. 图书馆工作与研究, 2025 (10): 87-95.
- [11] 富国瑞, 王平利, 王一展, 等. 基于大语言模型的高校图书馆智能参考咨询服务构建与应用研究: 以山东大学图书馆为例[J]. 图书馆杂志, 2025, 44 (12): 31-40, 47.
- [12] HUMPHREY-MURTO S, WOOD T J, GONSALVES C, et al. The Delphi method[J]. Academic Medicine, 2020, 95 (1): 168.
- [13] BEIDERBECK D, FREVEL N, VON DER GRACHT H A, et al. Preparing, conducting, and analyzing Delphi surveys: cross-disciplinary practices, new directions, and advancements[J]. MethodsX, 2021, 8: 101401.
- [14] NWODOH C O, OKORONKWO I L, ANARADO A N, et al. A modified Delphi consensus on generic indicators for a low- and

- middle-income country's quality nursing care measurement[J]. Nursing Open, 2022, 9 (5) : 2397-2408.
- [15] MEI W B, HSU C Y, OU S J. Research on the evaluation index system of the construction of communities suitable for aging by the fuzzy Delphi method[J]. Environments, 2020, 7 (10) : 92.
- [16] SANTONEN T, KAIVO-OJA J. The crowdsourcing Delphi: a method for combing expert judgements and wisdom of crowds[M]//UDEN L, TING I H, FELDMANN B. Knowledge management in organisations. Communications in computer and information science. Cham: Springer, 2022, 1593: 233-244.
- [17] 赵玉遂, 许燕, 吴青青, 等. 应用德尔菲法构建网络健康信息质量评价指标体系[J]. 预防医学, 2018, 30 (2) : 121-124.
- [18] 杨志勇. 基于德尔菲法的艾滋病防治专项支出绩效考评指标体系构建[J]. 中国农村卫生事业管理, 2014, 34 (6) : 644-647.

作者简介

周鑫, 男, 博士, 讲师, 研究方向: 信息分析与科学计量、知识产权与创新管理。

施未晔, 女, 本科生, 研究方向: 信息分析与科学计量。

郝江杰, 男, 硕士, 助理研究员, 通信作者, 研究方向: 科技情报, E-mail: 290996186@qq.com。

夏凯丽, 女, 硕士, 助理研究员, 研究方向: 科技情报。

Ethical Risks and Coping Strategies of AI Technology Applied to Scientific Research Support Services in University Libraries

ZHOU Xin^{1,2} SHI WeiYe¹ JIA JiangJie³ XIA KaiLi³

(1. School of Economics and Management, Nanjing Tech University, Nanjing 211816, P. R. China;

2. Jiangsu Provincial Key Laboratory of Data Engineering and Knowledge Service, Nanjing 210023, P. R. China;

3. Jiangsu Institute of Science and Technology Information, Nanjing 210042, P. R. China)

Abstract: With the deep penetration of artificial intelligence (AI) technology in scientific research support services in university libraries, the service efficiency of data management, intelligent retrieval, and other scenarios has significantly improved, but ethical risks have also become prominent. The aim of this study is to clarify the ethical risk types and causes of AI technology applied to scientific research support services in university libraries, and to construct a scientifically feasible coping strategy system. We use literature research method to sort out the current situation and ethical research progress of the integration of AI technology and scientific research support services in university libraries. By summarizing the practical experience and risk exposure of university libraries at home and abroad, and combining Delphi method, we conduct expert argumentation on risk dimensions and response measures. Research has found that the ethical risks of AI are mainly concentrated in four dimensions: data privacy, algorithm fairness, academic integrity, and responsibility definition. The causes involve multiple factors such as technological characteristics, institutional design, and subject literacy. The study proposes a four-dimensional coping strategy of "institutional norms-subject empowerment-regulatory collaboration-technology optimization" to provide theoretical support and practical reference for university libraries to prevent ethical risks in the application of AI technology in scientific research support services, and to assist in the intelligent development of university libraries.

Keywords: University Library; Research Support Service; AI Application; Ethical Risk

(责任编辑: 王玮)