

高校生成式人工智能教学风险分级治理研究 ——基于规范文本编码

罗慧莲^{1*}, 陆彦衡¹

(¹ 南宁学院 人工智能学院, 广西 南宁 530299)

摘要: 生成式人工智能进入高校教学后, 传统以“允许使用”或“禁止使用”为核心的二分治理方式, 已难以回应不同教学任务、不同技术介入程度和不同评价责任中的差异化风险。针对既有研究中风险等级划分依据不足、治理框架与风险评估结果衔接不够紧密等问题, 本文借鉴风险管理的流程思想, 基于国家政策文本、国际组织指南和国内外高校公开人工智能教学规范文本, 构建小型公开规范文本数据集, 并从使用场景、披露要求、责任归属、评价方式、数据安全和违规处理六个维度进行显性编码。在编码结果基础上, 本文将高校生成式人工智能教学应用风险归纳为学习替代风险、评价失真风险、数据安全风险和责任失序风险四类, 并以文本关注强度作为风险可能性的代理指标, 以教学后果影响强度作为风险严重性的判断依据, 运用风险矩阵形成可追溯的风险等级评价。研究发现, 评价失真风险处于重大风险等级, 学习替代风险、数据安全风险和责任失序风险处于重要风险等级。高校应依据风险等级和技术介入方式建立分级治理机制, 形成低风险任务开放使用、中风险任务披露使用、高风险任务过程审查、极高风险任务严格限制的治理路径。

关键词: 生成式人工智能; 高校教学; 风险分级; 风险矩阵; 规范文本编码; 教学治理

DOI: <https://doi.org/10.71411/jyyjx.2026.v1i5.1491>

Risk-Graded Governance of Generative AI in University Teaching: An Analysis Based on Normative Text Coding

Luo Huilian^{1*}, Lu Yanheng¹

(¹ Nanning University, School of Artificial Intelligence, Nanning, Guangxi, 530299, China)

Abstract: After generative artificial intelligence enters university teaching, the traditional binary governance model based on permission or prohibition can hardly respond to differentiated risks across teaching tasks, technological involvement and assessment responsibilities. In response to the insufficient basis for risk grading and the weak connection between risk assessment and governance strategies, this paper draws on the process logic of risk management and builds a small dataset of public normative texts, including national policy documents, international guidelines and public university regulations on AI use in teaching. Six dimensions are encoded, namely use scenarios, disclosure requirements, responsibility allocation, assessment design, data security and misconduct handling. Based on the coding results, the risks are summarized into learning substitution risk, assessment distortion risk, data security risk and resp-

作者简介: 罗慧莲 (1998-), 女, 广西南宁, 硕士, 研究方向: 教育技术

陆彦衡 (1999-), 女, 广西南宁, 硕士, 研究方向: 新闻传播学

通讯作者: 罗慧莲, 通讯邮箱: 1253852387@qq.com

onsibility disorder risk. Textual attention intensity is used as a proxy for risk possibility, and consequence impact intensity is used as the basis for risk severity. A risk matrix is then applied to form a traceable risk rating. The study finds that assessment distortion is at a major risk level, while learning substitution, data security and responsibility disorder are at important risk levels. Universities should establish a risk-graded governance mechanism according to risk level and technological involvement.

Keywords: Generative artificial intelligence; University teaching; Risk grading; Risk matrix; Normative text encoding; Teaching governance

引言

生成式人工智能正在改变高校教学中的知识获取、学习支持和成果生产方式^[1]。它既可以辅助资料检索、概念解释、语言润色、代码调试、课堂互动和个性化反馈,也可能诱发作业代写、文献虚构、实验数据伪造、学习过程被替代、课程评价结果失真以及教学数据泄露等风险^[2]。与一般数字化工具不同,生成式人工智能能够直接生成文本、代码、图像和论证结构,因而对高校原有学术诚信制度、课程评价方式和教学治理秩序形成更深层冲击^[3]。

从政策取向看,生成式人工智能治理并不是简单压制技术使用,而是要求在促进创新应用与防范安全风险之间形成平衡。《生成式人工智能服务管理暂行办法》强调促进生成式人工智能健康发展和规范应用,同时要求防范虚假有害信息、个人信息权益侵害、数据安全和偏见歧视等风险^[4]。UNESCO《教育和研究中的生成式人工智能指南》强调,教育系统应建立以人为中心的政策框架,使技术真正赋能教师、学习者和研究者^[5]。由此可见,高校面对生成式人工智能时,不宜停留在“全面禁用”或“完全放任”的二元判断,而应依据教学任务属性、技术介入程度和风险后果建立差异化治理机制^[6]。

现有研究多从学术诚信、教师角色转型、教育公平、数据安全和伦理治理等角度讨论生成式人工智能教学应用风险,相关成果为高校认识风险提供了重要基础^[7]。然而,既有研究仍存在三方面不足,第一,风险类型划分较为笼统,往往将不同教学任务中的风险混合讨论,缺少对风险源和教学场景的细化识别;第二,风险分级依据不够清晰,部分研究直接提出分级治理框架,却未说明不同风险等级如何形成;第三,治理措施与风险评估结果之间衔接不足,容易出现“先提出风险、再提出策略”,但二者之间缺少可追溯推导链条的问题^[8]。

基于此,本文主要围绕三个问题展开研究。一是高校生成式人工智能教学应用中主要存在哪些风险,二是如何依据公开规范文本对相关风险的可能性与严重性形成可复核的等级判断,三是高校如何结合风险等级与技术介入方式建立分级治理机制。为避免风险矩阵成为主观赋值工具,本文呈现公开规范文本编码结果,以文本编码频次作为风险可能性的代理指标,以教学后果影响强度作为风险严重性的判断依据,并据此形成由文本编码、风险识别、风险分析、风险评估和风险应对构成的研究路径。

1 数据来源与研究设计

1.1 数据来源

本文采用公开规范文本分析方法^[9]。研究材料包括三类,第一类是原则性文本,包括国家层面的生成式人工智能管理政策、国际组织发布的教育与研究应用指南以及教师生成式人工智能应用指引,用于提炼安全、透明、责任和人的主体性原则;第二类是操作性文本,包括国内外高校或教学机构公开发布的人工智能使用规范、课程政策和教学指引,用于观察高校在教学场景中如何规定人工智能使用边界;第三类是背景性数据,包括关于学生生成式人工智能使用情

况的公开调查报告，仅用于说明生成式人工智能进入高校评价任务的现实背景，不直接纳入本文编码计算^[10]。

为保证文本可核验和规则可识别，本文选取 12 份与高校教学、评价或学生使用人工智能直接相关的公开操作性文本。样本选择遵循三项原则，一是文本公开可得，便于复核；二是内容与高校教学、课程评价、学生作业或教师教学应用直接相关；三是具有明确规则表达，而非一般新闻报道或观点评论。由于样本规模有限，本文不将编码结果推广为全球或中国高校治理实践的总体状态，而是将其作为构建风险分级治理框架的经验基础。

1.2 编码方法

本文采用显性规则识别法进行编码^[11]。若文本中明确出现某类治理规则，则记为 1；若未明确出现，则记为 0。编码维度包括使用场景、披露要求、责任归属、评价方式、数据安全和违规处理六项。其中，“使用场景”指文本是否区分不同课程、作业、考试或评价任务中的人工智能使用边界；“披露要求”指是否要求学生或教师说明使用工具、使用环节和生成内容作用；“责任归属”指是否明确学生、教师或学校对最终成果、教学评价和数据处理承担责任；“评价方式”指是否涉及过程性评价、口头答辩、草稿记录、评价任务分类或评价方式调整；“数据安全”指是否限制输入个人信息、学生作品、课程内部资料、题库或未公开材料；“违规处理”指是否将未授权使用、未披露使用或替代性使用纳入学术不端或纪律处理。

在编码结果基础上，本文进一步采用主题归纳方法，将六项治理规则归并为四类教学风险。具体而言，使用场景、披露要求和评价方式主要指向学习过程是否被替代；使用场景、披露要求、评价方式和违规处理主要指向评价结果是否失真；数据安全维度直接指向教学数据外泄与使用失控；责任归属和违规处理共同指向风险发生后的责任识别与处置闭环。通过上述处理，本文将公开规范文本中的治理规则转化为风险识别和风险评估的可追溯依据。

表 1 高校生成式人工智能教学规范文本编码结果

序号	文本来源	场景	披露	责任	评价	数据	违规
1	ECNU-BNU 学生指南	1	1	1	0	1	1
2	SJTU 教师指南	1	1	1	1	1	0
3	Fudan 应用指引	1	1	1	1	1	0
4	PKU 边界指南	1	0	1	1	1	0
5	Harvard 指南	1	1	1	0	1	1
6	Oxford 评价政策	1	1	1	0	0	1
7	Cambridge 评价指南	1	1	1	1	1	1
8	UCL 分类指南	1	1	1	1	1	1
9	JHU 学生指南	1	1	1	0	0	1
10	JHU 教师指南	1	1	1	1	1	0
11	UPenn 政策示例	1	1	1	0	0	1
12	Stanford 课程政策	1	0	1	0	0	0

注：1 表示文本明确出现该项治理规则，0 表示文本未明确出现。“评价”指文本是否涉及过程评价、评价任务分类、口头答辩、草稿记录或评价方式调整；“数据”指文本是否涉及隐私保护、敏感数据、学生作品或未公开材料的输入限制。资料来源为作者根据 12 份国内外高校公开人工智能教学规范文本整理。

由表1可知,12份文本均涉及使用场景区分和责任归属,10份文本明确提出使用披露或声明要求,7份文本涉及数据安全,6份文本涉及评价方式调整,7份文本明确将违规使用、未披露或未授权使用纳入学术不端或纪律处理范围。这表明高校生成式人工智能教学治理已经从单纯讨论“能否使用”转向规定“何种场景、何种程度、何种责任下使用”,但过程评价、数据安全负面清单和违规处理程序仍有待细化。

2 高校生成式人工智能教学应用的风险识别

风险识别是发现、承认和描述风险的过程,其准确性直接影响后续风险分析和治理效果。依据表1中的编码结果及相关治理规则,本文将高校生成式人工智能教学应用风险归纳为学习替代风险、评价失真风险、数据安全风险和责任失序风险四类。四类风险并非彼此孤立,而是在教学任务、技术工具、评价制度和责任规则共同作用下相互叠加。

2.1 学习替代风险

学习替代风险是指学生在知识理解、资料阅读、问题分析、文本写作、代码编制和实验解释等环节中过度依赖生成式人工智能,使本应由学生完成的实质性学习劳动被技术替代。生成式人工智能的便捷性容易使学生将“完成任务”误认为“达成学习”,从而削弱主动探究、批判性阅读和独立表达能力。在低风险学习活动中,人工智能可作为资料检索和概念解释工具;但在课程论文、研究报告和实验报告等任务中,若人工智能介入提纲生成、主体写作和结论形成等核心环节,则会使学习过程被压缩为结果提交。

2.2 评价失真风险

评价失真风险是指生成式人工智能介入正式评价任务后,教师难以根据学生提交成果准确判断其真实能力、学习投入和学术贡献。高校课程评价具有认证、选拔和反馈功能,若人工智能直接生成作业主体、论文主体、实验分析、代码答案或考试回答,则评价对象将从学生能力转变为“学生加工具”的混合产物。评价失真不仅影响个体成绩公平,也可能破坏课程质量保障和学位授予制度的可信度。

2.3 数据安全风险

数据安全风险是指师生在使用外部生成式人工智能平台时,将学生身份信息、成绩数据、考试题库、课程内部资料、未公开科研数据、涉密项目材料或未经授权的学生作业输入平台,从而导致个人信息泄露、教学资源外流和科研数据失控。高校教学数据具有对象明确、关联性强和用途敏感等特点,一旦缺乏输入边界和脱敏要求,生成式人工智能工具的便捷使用可能转化为数据泄露通道。

2.4 责任失序风险

责任失序风险是指在生成式人工智能教学应用中,学生、教师、学校和平台之间的责任边界不清,导致风险发生后难以识别责任主体、认定违规行为和实施有效追责^[12]。现有讨论容易将人工智能使用风险简化为学生诚信问题,但教师使用人工智能备课、命题、生成案例、辅助批改和开展评价时,同样涉及内容准确性、知识产权、评价公平和数据保护责任。若课程规则、披露要求、审核程序和违规处理机制不明确,风险治理就难以形成闭环。

表 2 风险类型与编码维度对应关系

风险类型	对应编码维度	归纳依据
学习替代风险	使用场景、披露要求、评价方式	文本是否区分 AI 使用场景，是否要求说明 AI 介入程度，是否要求过程材料或评价方式调整。
评价失真风险	使用场景、披露要求、评价方式、违规处理	文本是否对正式评价任务作出限制，是否要求披露，是否将未授权或替代性使用纳入违规处理。
数据安全风险	数据安全	文本是否限制输入学生信息、成绩数据、课程资料、题库、未公开科研数据和未经授权学生作品。
责任失序风险	责任归属、违规处理	文本是否明确学生、教师、学校责任及违规认定、处理和申诉程序。

3 高校生成式人工智能教学风险分析

3.1 风险分析方法选定

风险分析是在识别风险后，对风险发生可能性和风险后果严重性进行判断，并据此形成风险等级。ISO 31000 (Risk management—Guidelines) 风险管理标准强调风险识别、风险分析、风险评估和风险应对的连续过程^[13]。需要说明的是，本文并非对高校生成式人工智能教学风险进行严格意义上的标准化风险审计，而是借鉴其流程思想，结合公开规范文本编码结果，构建一种适用于高校教学治理研究的风险分析框架。

在风险管理研究中，德尔菲法、层次分析法、风险指数法和风险矩阵法均可用于风险排序^[14]。考虑到高校生成式人工智能教学应用具有场景复杂、历史统计数据不足、不同高校规范文本可获得性较强等特点，本文选用风险矩阵作为启发式分析工具。与单纯经验判断相比，风险矩阵能够将风险可能性和影响严重性进行结构化呈现；与复杂统计模型相比，风险矩阵对大样本行为数据的依赖较低，适合在规范文本编码基础上开展初步评估^[15]。

为避免风险矩阵中的 P 值和 I 值完全来自作者主观赋值，本文将 P 值与公开规范文本编码频次相连接，将 I 值与教学后果影响维度相连接。换言之，本文的风险矩阵并不声称测量现实发生概率或精算损失，而是以可复核的编码结果和明确的后果评分规则，形成风险等级的规范化推导。

3.2 可能性与严重性的计算规则

本文将风险发生可能性记为 P，将风险影响严重性记为 I，将风险综合等级记为 R。风险综合等级由风险发生可能性与影响严重性共同决定，计算公式为

$$R = P * I \quad (1)$$

其中，P 和 I 均采用五级赋值，取值范围为 1 至 5。为降低风险可能性赋值的主观性，本文以相关编码维度在 12 份公开规范文本中的出现频次作为风险可能性的代理指标。具体而言，若某一风险类型对应 m 个编码维度，则先计算该风险的文本关注强度 C，计算公式为

$$C = \frac{\sum_{j=1}^m F_j}{12m} * 5 \quad (2)$$

式中, F_j 表示第 j 个相关编码维度在 12 份公开规范文本中的出现次数, m 表示该风险类型对应的编码维度数量, 12 表示本文纳入编码分析的规范文本总数。随后对 C 值进行四舍五入处理, 得到风险发生可能性等级 P 。若换算结果低于 1, 则按 1 计。风险可能性 P 值计算规则如表 3 所示, 通过上述处理, 风险发生可能性不再完全依赖作者主观判断, 而是与公开规范文本中的治理关注程度相联系, 从而提高风险等级判断的可复核性。

表 3 风险可能性 P 值计算表

风险类型	相关编码维度及频次	C 值	P 值	说明
学习替代风险	场景 12、披露 10、评价 6	3.89	4	学习过程替代同时受到场景边界、使用披露和评价方式调整影响。
评价失真风险	场景 12、披露 10、评价 6、违规 7	3.65	4	评价失真与正式评价场景、披露制度、评价方式和违规处理均相关。
数据安全风险	数据 7	2.92	3	数据安全规则在 7 份文本中出现, 说明关注较明确但覆盖仍不足。
责任失序风险	责任 12、违规 7	3.96	4	责任归属普遍出现, 但违规处理规则并未完全同步形成闭环。

I 值以风险后果影响强度为依据。如表 4 所示, 设置五个后果维度: 是否直接替代学习过程, 是否影响正式评价结果, 是否涉及个人信息、教学数据安全或学生权益, 是否触发学术诚信或纪律责任, 是否影响学校治理公信力。每一维度若影响明显记 1 分, 影响不明显记 0 分; 总分为 1-5 分, 对应 I 值。该规则虽然仍具有规范判断成分, 但评分维度公开透明, 能够说明不同风险严重性差异的来源^[16]。

表 4 风险严重性 I 值评分表

风险类型	学习过程	评价结果	数据/权益	诚信责任	治理公信	I 值
学习替代风险	1	1	0	1	1	4
评价失真风险	1	1	1	1	1	5
数据安全风险	0	1	1	1	1	4
责任失序风险	0	1	0	1	1	4

注: 1 表示该风险在相应后果维度上影响明显, 0 表示影响不明显。“数据/权益”既包括个人信息和教学数据安全, 也包括因评价失真造成的学生评价权益受损。

为呈现本文的整体研究路径和分级治理逻辑, 本文将公开规范文本编码、风险识别、风险评估和治理应对纳入同一分析框架。如图 1 所示, 高校生成式人工智能教学风险分级治理以规范文本编码为起点, 经由风险识别和风险矩阵评估, 最终形成差异化治理策略, 并通过持续修正实现动态治理。

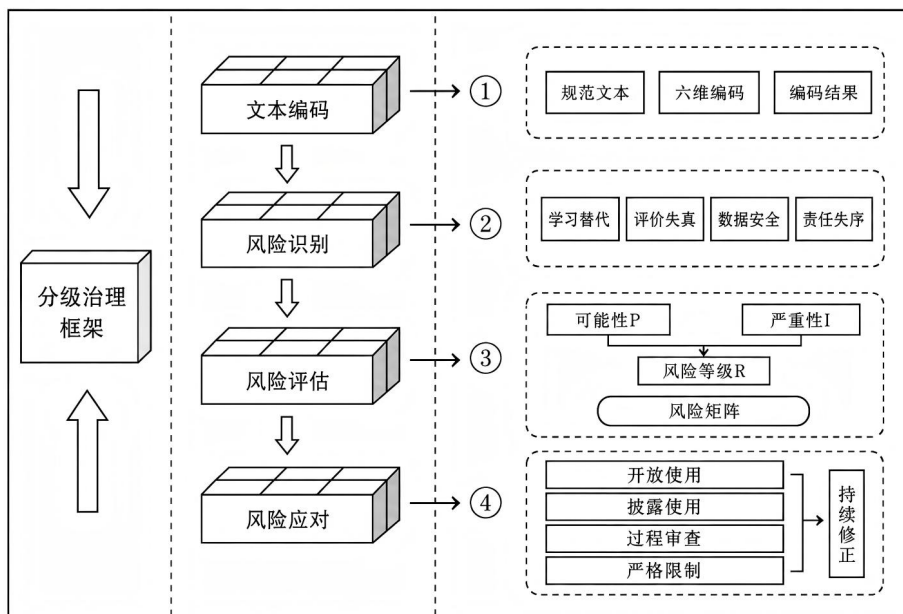


图 1 高校生成式人工智能教学风险分级治理分析框架

4 高校生成式人工智能教学风险评估

依据 P 值和 I 值计算结果，高校生成式人工智能教学应用风险综合等级如表 5 所示。风险等级划分如下：R 为 1-2 时为可忽略风险；R 为 3-5 时为可接受风险；R 为 6-10 时为可容忍风险；R 为 12-16 时为重要风险；R 为 17-25 时为重大风险。上述等级用于说明不同风险在高校教学治理中的优先顺序，并不等同于严格统计意义上的损失测量。

表 5 高校生成式人工智能教学风险综合等级评价表

主要风险	P	I	R	风险等级	治理含义
学习替代风险	4	4	16	重要风险	需要强制披露、学习过程记录和教师过程审查。
评价失真风险	4	5	20	重大风险	直接影响正式评价真实性，应限制或禁止替代性使用。
数据安全风险	3	4	12	重要风险	应建立教学数据负面清单和脱敏处理规则。
责任失序风险	4	4	16	重要风险	应明确学生、教师、学校和平台责任链条。

由表 5 可知，评价失真风险的综合等级最高，属于重大风险。这说明生成式人工智能对高校教学治理的最大冲击并非一般性工具使用，而是正式评价任务中学生真实能力、学习投入和学术贡献难以识别的问题。学习替代风险、数据安全风险和责任失序风险均属于重要风险，表明高校需要在学习过程、数据输入边界和责任闭环机制上建立较高强度的制度安排。

与仅凭作者判断直接给出 P 值和 I 值不同，本文的风险等级由表 1 的编码频次、表 2 的风险维度对应关系、表 3 的可能性换算和表 4 的后果评分共同推导而来。尽管这种推导仍受到样本规模和文本公开性的限制，但其计算过程可以被复核和修正，避免了风险矩阵中常见的量化形式化、依据主观化问题^[17]。

5 高校生成式人工智能教学风险分级治理路径

风险评估的目的并非停留在排序，而是为风险应对提供依据。为避免治理框架与风险等级脱节，风险等级建立治理强度转换规则如表 6 所示，可忽略风险和可接受风险以开放使用和—

般核验为主；可容忍风险要求必要披露和教师提示；重要风险要求强制披露、过程材料和审核记录；重大风险则应严格限制或禁止替代性使用。

表 6 风险等级与治理策略对应关系

R 值区间	风险等级	治理策略	适用说明
1—2	可忽略风险	开放使用，要求核验	适用于资料检索、概念解释等非正式评价任务。
3—5	可接受风险	允许使用，建议披露	适用于课堂讨论、语言润色等低影响任务。
6—10	可容忍风险	有限使用，必要披露	适用于一般课程作业和辅助性学习活动。
12—16	重要风险	强制披露，过程审查	适用于课程论文、研究报告、实验报告等高风险任务。
17—25	重大风险	严格限制或禁止替代性使用	适用于考试、毕业论文、核心能力测试等极高风险任务。

风险评估的意义在于为具体治理措施提供依据。为避免风险等级与治理策略脱节，本文进一步将四类主要风险、综合等级、对应教学任务和治理措施进行衔接。如表 7 所示，不同风险类型应对应不同治理强度，其中评价失真风险应重点采取严格限制措施，学习替代风险、数据安全风险和责任失序风险则应通过过程审查、数据负面清单和责任链条完善加以治理。

表 7 风险等级与高校教学治理措施衔接表

风险	R 值	风险等级	对应教学任务	治理措施
学习替代风险	16	重要风险	课程论文、研究报告、实验报告、代码作业	要求学生提交工具使用说明、资料筛选记录、阶段性草稿、修改记录和反思报告。
评价失真风险	20	重大风险	考试、毕业论文、核心能力测试、正式评价任务	禁止人工智能替代性生成核心答案、主体论证、数据分析和结论；必要时采用口头答辩和现场写作。
数据安全风险	12	重要风险	智能批改、案例生成、论文指导、科研训练	建立教学数据负面清单，禁止输入身份信息、成绩数据、题库、未公开科研数据和未经授权学生作品。
责任失序风险	16	重要风险	课程大纲、作业说明、论文指导、违规处理	明确学生披露责任、教师审核责任和学校制度责任，形成违规认定、申诉与处理程序。

5.1 低风险任务的开放使用

对于资料检索、概念解释、语言润色、课堂讨论和学习反馈等低风险任务，高校可以允许学生和教师使用生成式人工智能。此类任务的治理重点不是限制使用，而是要求信息核验、来源确认和批判性判断。学生可以使用人工智能辅助理解概念、梳理阅读材料、优化语言表达，但不得直接复制未经核验的生成内容。

5.2 中风险任务的披露使用

对于提纲生成、一般课程作业、代码调试和案例分析等中风险任务，高校应允许有限使用，但必须要求学生说明使用工具、使用环节和生成内容在最终成果中的作用。教师应在课程大纲和作业说明中提前规定人工智能使用边界，避免学生在规则不明确的情况下自行判断。

5.3 高风险任务的过程审查

对于实验报告、课程论文、研究报告和项目作品等高风险任务，高校应将治理重点从最终文本检测转向过程审查^[18]。学生应提交选题说明、资料筛选记录、阶段性草稿、修改记录、数据来源说明和反思报告。教师应通过课堂展示、面谈和口头答辩判断学生是否真正理解并掌握自己的成果。

5.4 极高风险任务的严格限制

对于毕业论文、闭卷考试、现场写作和核心能力测试等极高风险任务，应严格限制或禁止生成式人工智能的替代性使用。毕业论文可以允许有限的语言润色、格式整理和资料管理辅助，但选题设计、核心论证、数据分析和结论形成必须由学生独立完成；闭卷考试、现场写作和核心能力测试应禁止使用人工智能。

5.5 教学数据安全负面清单

高校应明确不得向外部人工智能平台上传学生个人身份信息、成绩数据、考试题库、课程内部资料、未公开科研数据、涉密项目材料和未经授权的学生作业。对确需使用人工智能处理的数据，应进行脱敏处理，并优先使用校内合规平台。数据安全治理不能只停留在原则提醒，而应转化为可操作的负面清单、审批流程和责任边界。

6 讨论与研究局限

本文试图解决高校生成式人工智能教学风险分级治理研究中的方法论缺口，即风险等级不能仅由作者经验判断，而应具有可追溯的文本依据和明确的换算规则。通过恢复公开规范文本编码表，并将编码频次与风险矩阵相连接，本文在框架建构研究中引入了较强的经验材料支撑。与此同时，本文并未将 12 份规范文本视为充分的大样本数据，也未声称 P 值等同于现实发生概率。因此，本文的风险评估更适合作为高校教学治理规则设计的规范化分析，而非严格意义上的行为统计研究。

本文仍存在三方面局限。第一，样本数量有限，且文本主要来自公开可得的高校规范，难以覆盖所有高校的实际治理实践。第二，编码过程以显性规则识别为主，虽然有利于复核，但可能忽略隐含规则和具体执行效果。第三，I 值基于后果维度评分形成，仍需要后续通过专家访谈、师生问卷、课程案例或学习行为数据进一步验证。未来研究可扩大高校规范文本样本，采用双人编码和一致性检验，并结合学生使用行为调查、教师访谈和课程评价数据，对风险等级和治理策略进行进一步校准。

7 结束语

生成式人工智能进入高校教学已成为不可回避的治理议题。高校治理的核心，不是简单判断技术能否使用，而是根据教学任务风险、技术介入程度和责任链条，划定差异化的使用边界。本文基于公开规范文本编码结果，构建由文本编码、风险识别、风险分析、风险评估和风险应对构成的分级治理框架，将高校生成式人工智能教学应用风险归纳为学习替代风险、评价失真风险、数据安全风险和责任失序风险，并运用风险矩阵进行等级评价。研究发现，评价失真风险属于重大风险，学习替代风险、数据安全风险和责任失序风险属于重要风险。高校应在开放使用与严格限制之间建立分层治理机制，使生成式人工智能成为提升教学质量的辅助工具，而不是削弱学习过程、评价公平与学术诚信的替代机制。

参考文献:

- [1] 李世瑾, 胡艺龄, 顾小清. 如何走出人工智能教育风险的困局: 现象、成因及应对[J]. 电化教育研究, 2021, 42(7): 19-25.
- [2] 胡小勇, 黄婕, 林梓柔, 等. 教育人工智能伦理: 内涵框架、认知现状与风险规避[J]. 现代远程教育研究, 2022, 34(2): 21-28+36.
- [3] 王媛, 王日臻, 樊丽. 大语言模型在学术期刊中应用的风险防控与治理策略——基于内容原创性、学术诚信与数据安全的视角[J]. 科技与出版, 2026(04): 119-128.
- [4] 国家互联网信息办公室, 国家发展和改革委员会, 教育部, 等. 生成式人工智能服务管理暂行办法[EB/OL]. (2023-07-10)[2026-05-16]. https://www.gov.cn/zhengce/zhengceku/202307/content_6891752.htm
- [5] UNESCO. Guidance for generative AI in education and research[R]. Paris: UNESCO, 2023.
- [6] 孙平. 高校规范人工智能使用的法治化路径[J]. 法学, 2025(08): 35-54.
- [7] 王佑镁, 王欣颖, 柳晨晨. 教育领域生成式人工智能应用的伦理风险管理框架研究[J]. 电化教育研究, 2024, 45(10): 28-34+42.
- [8] 王佑镁, 王旦, 王海洁, 等. 基于风险性监管的 AIGC 教育应用伦理风险治理研究[J]. 中国电化教育, 2023(11): 83-90.
- [9] 陈星月. 基于三维分析框架的世界一流高校生成式人工智能使用政策文本量化研究[D]. 太原: 山西大学, 2025.
- [10] 教育部教师队伍建设专家指导委员会. 教师生成式人工智能应用指引: 第一版[Z]. 2025.
- [11] 金展旭, 蔡天骥. 生成式人工智能在文本编码中的应用——以裁判文书为例[J]. 智能社会研究, 2025, 4(02): 202-224+253.
- [12] 赵磊磊, 吴小凡, 赵可云. 责任伦理: 教育人工智能风险治理的时代诉求[J]. 电化教育研究, 2022, 43(6): 32-38.
- [13] International Organization for Standardization. ISO 31000:2018 Risk management—Guidelines[S]. Geneva: ISO, 2018.
- [14] 国家市场监督管理总局, 国家标准化管理委员会. GB/T 27921-2023 风险管理 风险评估技术[S]. 北京: 中国标准出版社, 2023.
- [15] GARVEY P R, LANSLOWNE Z F. Risk matrix: an approach for identifying, assessing, and ranking program risks[J]. Air Force Journal of Logistics, 1998, 22(1): 18-21.
- [16] 李梦薇, 徐峰, 晏奇, 等. 服务机器人领域人工智能伦理风险评估方法的设计与实践[J]. 中国科技论坛, 2023(10): 74-84.
- [17] 盘大清, 李白杨, 任尚升. 人工智能生成内容(AIGC)深度伪造信息风险管理能力评估方法研究[J]. 图书与情报, 2025(05): 12-23.
- [18] 洪涛. AIGC 检测的功能反思与规范适用[J/OL]. 科学学研究: 1-13[2026-05-17]. <https://doi.org/10.16192/j.cnki.1003-2053.20260224.002>.