DOI:10.13811/j.cnki.eer.2024.10.004

教育领域生成式人工智能应用的伦理风险管理框架研究

王佑镁， 王欣颖， 柳晨晨

（温州大学 大数据与智慧教育研究中心，浙江 温州 325035）

[摘 要] 生成式人工智能对内容生产与传播产生了革命性影响，教育领域进入了智能化与风险性并存的人机协同时代，教育人工智能应用伦理风险频繁发生，急需通过标准化的风险评估工具进行科学管控。文章从技术与教育两个向度出发，将教育领域生成式人工智能应用的伦理风险归纳为技术本体风险、教育数据风险、机器算法风险和教育应用风险四大类。借鉴 ISO31000 风险管理标准与应用指南，搭建“风险识别—风险分析—风险评估—风险应对”四级流程的伦理风险管理框架。构建风险源检查表，以风险矩阵为主要评估工具，在实践中统计得出各类伦理风险综合等级。教育领域生成式人工智能应用的伦理风险管理框架，将为教育教学有效实施伦理风险管理工作提供借鉴价值。

[关键词] 生成式人工智能；伦理风险；教育人工智能；风险管理；风险评估

[中图分类号] G434 [文献标志码] A

[作者简介] 王佑镁（1974—），男，江西吉安人。教授，博士，主要从事人工智能教育、人工智能伦理风险研究。E-mail：wangyoumei@126.com。

一、问题的提出

2024 年初，Sora 的发布重燃公众对生成式人工智能的关注。作为一种视觉生成式人工智能，Sora 不仅具备理解人类话语能力，还能模拟物理世界规则，在某种程度上改变了教育资源的呈现方式。生成式人工智能既能够以教为主，依据教学目标生成创作型教学素材，辅助教师设计有创新性的教学活动[1]；又能够以学为主，根据学生能力和进度给予合适的指导和反馈，拓展学生的思维和理解能力；还能够以管为主，为教育治理提供支持，提供改进教师教学行为和学生学习行为的反馈信息[2]。可见，生成式人工智能在教育中的应用能够提升不同角色教育效能的创意度与完成度[3]。尽管生成式人工智能为教育的发展带来巨大机遇，但仍会产生数据滥用、内容偏见、过度依赖等伦理风险。UNESCO 发布全球首份《教育和研究中的生成式人工智能指南》，指出生成式人工智能可能造成的伤害，如果缺乏公众参与及政府必要的保障和监管，人工智能就无法完全融入教育[4]。我国发布《生成式人工智能服务管理暂行办法》，强调生成式人工智能应当遵守法律法规，尊重社会公德和伦理道德[5]。因此，需要在伦理风险发生之前着手开展风险管理行动，以提高生成式人工智能在教育领域应用的科学性，实现良性发展。

伴随着技术智能化与自主性的发展以及教育主观性强、复杂性突出的特点，教育领域生成式人工智能应用的伦理风险管理始终是一大难题，亟待管控化解。目前，教育领域的风险评估主要依靠专家判断或经验总结，此类方法较为主观。对于特定的伦理风险，也缺少标准的量化方式。为此，采用何种方式对教育领域中生成式人工智能应用产生的伦理风险进行识别、分析与评估，是本研究主要探讨的议题。

二、教育领域生成式人工智能应用的伦理风险

人工智能伦理指人工智能应遵守的道德规范，使

基金项目：国家社会科学基金 2023 年度教育学重大项目“新一代人工智能对教育的影响研究”（项目编号：VGA230012）

人工智能作出可为人接受的决策。智慧教育时代,人工智能伦理风险的发生极具普遍性,数据泄露、算法歧视、师生关系弱化等问题时常发生[6]。究其原因,不外乎人工智能技术本身存在风险,以及人类使用人工智能技术过程中会产生风险。例如,对教育者而言,生成式人工智能带来的教学新范式易使教师将自身责任让渡于人工智能[7],导致教师地位消解,师生情感异化。学者往往担忧人类与技术共同创作时产生的学术剽窃风险,以至于不断有学者提出禁用人工智能的观点,避免教育成为后剽窃时代的起点。

人工智能伦理风险于科技与人二者博弈的关系中产生[8]。管理生成式人工智能伦理风险,须首先从技术与教育两个向度审视其实质。生成式人工智能伦理风险,即主体与技术、自身、他人、社会之间的伦理关系由于正面或负面的影响产生的不确定事件, 尤指伦理关系失调、机制失控、社会失序等伦理负效应[9]。分析梳理已有研究,本文将生成式人工智能伦理风险总结为如图 1 所示的四大类,即技术本体风险、教育数据风险、机器算法风险和教育应用风险,并细化次要伦理风险。

(一)技术本体风险

在教育领域中,技术本体风险主要指:第一,生成式人工智能成为直接驾驭当前教育活动的元素。人工智能不再是传统教育语境中工具性的存在,而成为了直接关乎教育“如何实施”的最基本元素,造成人与技术的主体错位。第二,教育主体受到人工智能热潮下新教学原则的支配。“人没有技术,就缺少了对环境的作用与反作用”[10],而技术本体风险错在将“技术成为主体”取代“技术创新主体”。

技术本体风险主要诱发两类风险情境:一是对生成式人工智能毫无保留地信任, 产生技术拜物教风险,养成“技术心流”陋习[6];二是对生成式人工智能无条件依赖,产生技术依赖风险。第一类情境主要指教师与学生使用生成式人工智能辅助决策的过程中,将自身真实信息全盘托出,并且对人工智能输出信息不查伪、不审思,一方面夸大人工智能的实际能力,另一方面忽视人工智能可能存在的信息错误与技术局限。第二类情境主要指教师与学生对生成式人工智能的过度使用,使个体在思维与行为上对技术成瘾,从而无法约束自我, 滋生没有技术就难以实现目标的想法。人们对人工智能工具的无条件依赖,实际是被技术所奴役,迷失了对教育目的与意义的关注[11]。

(二)教育数据风险

教育数据是教学过程、教育管理活动、教学科研活动和校园生活等数据的总和,对教育数据的分析与应用是生成式人工智能辅助决策的重要依据。然而,大数据时代前所未有的数据挖掘、数据预测和数据监控,造成了更为严重的教育数据风险。

教育数据风险主要包括三类: 一是数据泄露风险。由于缺乏数据保护机制与数据保护意识,导致教育敏感数据被外部实体获取。以 ChatGPT-3.5 为例,DeepMind 研究人员曾采用分歧攻击让 ChatGPT 逐渐偏离聊天内容,泄露原始训练数据。二是数据失真风险。教育之大数据,指数据数量之大,也指数据价值之大[12]。但由于前期采集教育数据渠道的单向性,使用人工智能过程中未及时更新维护数据,易使数据失去时效性与准确性。三是数据滥用风险。教育数据使用方对数据做出未经授权的采集、超出授权范围的使用、不正当的修改等行为来实现教育服务之外的目的,将违背教育的初衷。因此,如何达到保障教育隐私和教育数据共享之间的巧妙平衡,将是生成式人工智能发展过程中面临的巨大伦理挑战。

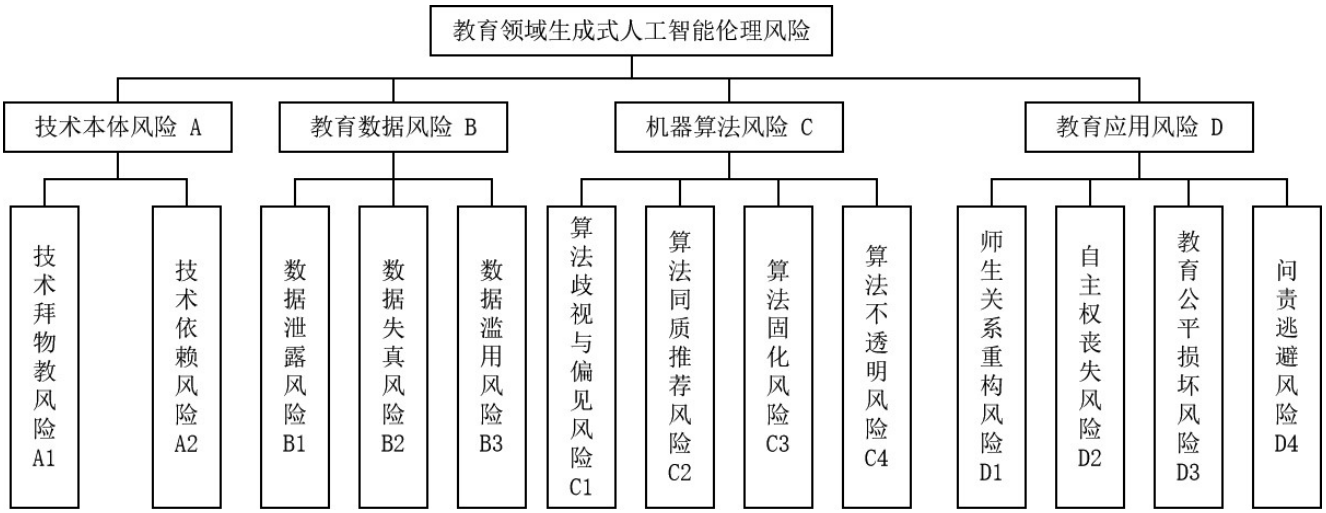

图 1　教育领域生成式人工智能伦理风险分类

(三)机器算法风险

算法是生成式人工智能的底层逻辑,用于驱动内部数据以规范的输入形式转化为特定的输出形式。人工智能机器算法推动教育展开了高效的智能化转型,但也在预设的规则中窄化了教育思维与活动的外延。

机器算法风险表现为四个方面:一是算法歧视与偏见风险。儿童通过学习知识与经验形成三观,同样,人工智能对海量数据进行算法训练得以形成大模型。然而,原始数据中对不同文化与种群根深蒂固的歧视与偏见同样被算法继承,并受算法本身所强化,不断加深决策中的刻板印象[13]。二是算法同质推荐风险。个性化的“推荐算法”以协同过滤为根本机制,将用户选择归为某一标签,算法推荐时过滤其他标签内容,仅推荐用户可能偏好的资源。长此以往,面对低年龄段儿童,算法可能会推荐大量同质且低质内容,为儿童织起信息茧房,影响儿童认知正常发展。国家四部门出台《互联网信息服务算法推荐管理规定》,强调要加强对信息服务中信息茧房现象的监管[14]。三是算法固化风险,即由算法固有功能缺陷衍生出的风险。当使用者提出超过算法设计范围的需求时,可能导致最终结果偏离实际需求或产生错误结果。如果不对固化算法及时改进与更新,可能导致使用受到限制或无法作出准确的决策。四是算法不透明风险,人们常用“黑箱”隐喻算法的不透明性,即无法从用户视角直接观测到算法的运算逻辑。由于企业或国家的保密、个人技术素养的不足[15],逐渐使生成式人工智能的任何决策都缺乏解释,最终形成难以捉摸的黑箱社会形态。

(四)教育应用风险

教育应用风险特指除生成式人工智能自身技术局限、算法逻辑等问题外,对教育主体的社会属性和个体属性造成的风险。进一步可以划分为师生关系重构风险、自主权丧失风险、教育公平损坏风险、问责逃避风险。生成式人工智能逐渐成为师生之间知识获取与情感表达的纽带,但是通过语音识别、眼球追踪等设备捕捉师生生理和行为数据[16],是对师生的过度监管与评估。人工智能以“看管者”的角色进入校园,无疑是对师生自主权的“隐形剥削”和“软性压迫”,从而导致教师地位的消解和师生情感的异化。除此之外,生成式人工智能的普及也将引发新的数字鸿沟。伴随马太效应,具有技术优势的地区与个人将依靠技术的加持获取更多优势。但是,在尚且缺乏技术问责机制与反馈处理通道的环境下,若是将师生视作人工智能数据训练的“工具人”,将教育的营养由机器反复咀嚼以喂养师生,人们的思想和行为必会受到限制,教育主体的地位也将受到挑战。

三、教育领域生成式人工智能应用的伦理风险识别

国际标准化组织 ISO 发布 ISO31000 风险管理标准,提供详细的风险管理指南,帮助组织制定相应的策略和措施来评估、应对风险。在 ISO31000 标准中,风险评估过程包括风险识别与风险分析。风险识别是发现、承认和描述风险的过程,包括对风险源、风险事件的识别,风险识别的准确性将直接影响到风险管理工作的质量和结果。风险分析是理解风险性质和确定风险等级的过程,风险分析的结果是项目决策与行动的输入条件,重点在于分析风险发生的可能性和对结果造成影响的严重程度,从而进行风险等级的划分。

厘清教育领域生成式人工智能应用的伦理风险后,应随之辨识导致伦理风险发生的来源。技术、数据、算法和教育之间的衔接,构成了完整统一的人工智能教育生态系统。风险因素对教育生态系统产生的影响常常难以准确根治,因此,在风险发生之前识别风险源,有助于应对措施的精准施加。根据教育领域生成式人工智能多样的伦理风险,需要编制条理清晰、具有普适性的风险源检查表,以便不同教育场景下伦理风险的识别。风险的发生是一个渐进的过程,因此,需要从生成式人工智能设计与使用的全流程排查风险因素。生成式人工智能设计与使用的全流程包括设计时所依策略、所用输入、所产输出和使用时所依策略、所用输入、所产输出。设计阶段主要指开发商和程序员在内的开发人员设计生成式人工智能的过程,使用阶段主要指学生、教师和教育管理者在内的教育人员使用生成式人工智能的过程。在分析生成式人工智能设计与使用的全流程后,得到的风险因素如图 2 所示。

排查设计与使用阶段风险因素后,需细化风险因素至具体风险类型,以辨识伦理风险源。例如,技术本体风险主要存在于教育人员的使用阶段,不同类型的生成式人工智能会对师生主体性造成不同的影响。教育数据风险与开发人员的设计和教育人员的使用均存在关联,例如,开发过程中缺少数据保护措施会引发数据泄露风险,教育人员做好对教育数据的权限管理有利于防范数据滥用风险。机器算法风险存在于整个设计阶段,策略中所用算法模型功能的片面化、输入使用的训练数据本身的偏见、标榜个性化的同质内容分别会导致算法固化、算法歧视与偏见和算法同质推荐风险。教育应用风险受到使用与设计的共同影响,技术人员的过度监管易使师生丧失自主权,同时,教育人

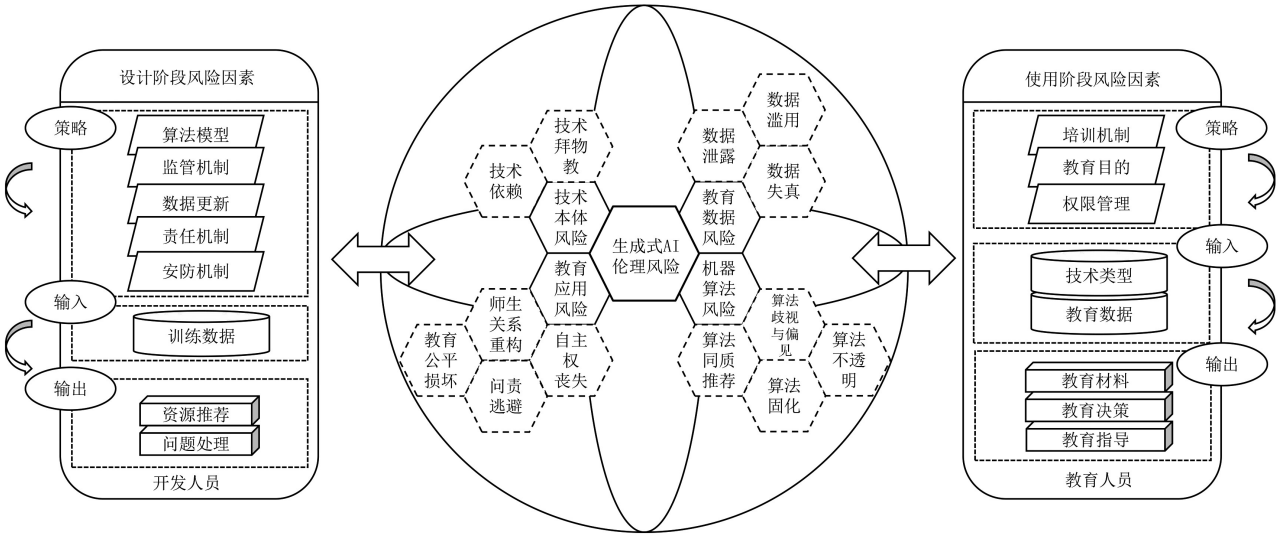

图 2　生成式人工智能设计与使用阶段的风险因素

员缺乏对教育目的的正确认识也将损害教育公平。教育领域生成式人工智能伦理风险源检查表,见表 1。

表 1　教育领域生成式人工智能伦理风险源检查表

伦理风险分类	风险源
技术本体风险	夸大技术效果;忽视技术局限;行为受到支配;无节制的技术使用
教育数据风险	泄露隐私数据;采集数据单向;缺乏数据保护措施; 数据老化未更新; 数据操纵与篡改;数据超范围使用;数据未授权使用
机器算法风险	强化偏见;存在刻板印象;资源内容单一;推荐资源质量低;算法功能片面化;问题处理僵化;算法黑箱;决策机制不透明;结果缺乏解释性
教育应用风险	教师地位消解;师生情感异化;学生情感遮蔽;过度监管;行为受限;思维创新受限;扩大数字鸿沟; 教育目的不明; 责任主体模糊;缺乏问责机制

四、教育领域生成式人工智能应用的伦理风险分析

(一)风险分析方法选定

风险分析是在识别风险后, 对风险造成的危害和风险发生的频率进行评估,从而得到具体的风险值,综合计算得出风险等级。长久以来,风险管理研究飞速发展,但主要存在于金融与工程行业。目前教育领域的风险分析研究,主要采用文献分析法与德尔菲法,主观性较强,缺乏实证数据支持。国家现行标准 GB/T 27921–2023《风险管理 风险评估技术》涵盖大量项目风险分析方法,如 HACCP 分析法、蒙特卡罗模拟分析法、风险矩阵等。《风险管理 风险评估技术》中指出,在风险评估的不同阶段有不同的适用方法[17],见表 2。

表 2　风险评估方法在风险评估各子过程的适用性[17]

风险分析方法	风险评估过程				
	风险识别	风险分析			风险评价
		后果	可能性	风险等级	
德尔菲法	SA	NA	NA	NA	NA
调查法	SA	NA	NA	NA	NA
风险指数法	A	SA	SA	A	SA
HACCP 分析法	SA	SA	NA	NA	SA
HAZOP 分析法	SA	A	NA	NA	NA
风险矩阵	SA	SA	SA	SA	A

注:“A”为适用,“SA”为非常适用,“NA”为不适用。

在选择风险分析方法时, 应基于具体环境和用途,并以利益相关者需要的形式提供信息。总体而言,风险分析方法的选择宜考虑下列要素:(1) 评估的目的;(2)利益相关者的需求;(3)运行环境和场景;(4)法律与监管要求;(5) 既定的决策准则及其形式;(6)现有信息和可获得信息;(7) 情况的复杂程度;(8)可用或可获得的专业性知识[18]。

基于已识别的风险类型,本研究将采用定性和定量相结合的风险矩阵。生成式人工智能的技术形态呈现多样化的特点,其应用的灵活高效引发更多层次的伦理问题。风险矩阵一方面可以直观地体现风险发生的可能性与影响程度, 提供伦理风险等级的定量结果。另一方面,风险矩阵受资源、能力、环境的不确定性影响较小,相对于其他的风险等级确定方法而言复杂性较低。综合考虑,采用风险矩阵分析教育领域生成式人工智能应用的伦理风险,能够保障风险评估过程较高的稳定性和可操作性。

(二)风险矩阵

风险矩阵(Risk Matrix)最初由美国空军电子系

统中心在 1995 年 4 月提出[19]。风险矩阵在项目风险管理方面有着广泛的应用,一般在识别项目管理过程中可能产生的风险后,用于对风险的可能性与后果严重性进行评估[20]。

可能性(Probability)是风险发生的概率,记为 P。风险可能性依据可能性准则编写,数值呈阶梯式跳跃,一般分为五个等级:(1)一级:风险发生概率为 0~10%,几乎不会发生;(2)二级:风险发生概率为 11%~40%,有一定概率会发生;(3)三级:风险发生概率为 41%~60%,偶尔会发生;(4)四级:风险发生概率为 61%~90%,经常多次发生;(5)五级:风险发生概率为 91%~100%,会频繁发生。

严重性(Impact)是风险发生后对项目产生负向影响的强度,记为 I,可划分为五个等级:(1)可忽略:风险发生后仅造成微不足道的负向影响;(2)可接受:风险将产生可接受范围内的负向影响;(3) 需合理控制:风险将造成明显的负向影响,接近项目可承受的边缘;(4)需严格控制:风险将对项目造成重大威胁;(5)不可接受:风险将对项目造成根本性破坏。

风险等级(Risk Rating)由可能性与严重性共同决定,是二者的乘积,用 R 表示:R=P×I。风险矩阵主要描述不同风险的等级,如图 3 所示。风险矩阵中风险等级准则如下:风险等级在 1~2 之间为可忽略风险,表示风险水平非常低,无需专门控制;风险等级在 3~5 之间为可接受风险,表示风险水平比较低,已有适当控制措施,可以接受;风险等级在 6~10 之间为可容忍风险,表示风险水平适中,已有充分的控制措施,可以容忍;风险等级在 12~16 之间为重要性风险,表示风险水平较高,但在有效的控制下可以被接受;风险等级在 17~25 之间为灾难性风险,表示风险水平非常高,无法接受,需立即采取行动应对风险。

风险等级

可能性等级						
	5	5（可接受风险）	10（可容忍风险）	15（重要性风险）	20（灾难性风险）	25（灾难性风险）
	4	4（可接受风险）	8（可容忍风险）	12（重要性风险）	16（重要性风险）	20（灾难性风险）
	3	3（可接受风险）	6（可容忍风险）	9（可容忍风险）	12（重要性风险）	15（重要性风险）
	2	2（可忽略风险）	4（可接受风险）	6（可容忍风险）	8（可容忍风险）	10（可容忍风险）
	1	1（可忽略风险）	2（可忽略风险）	3（可接受风险）	4（可接受风险）	5（可接受风险）
		1	2	3	4	5
		严重性等级				

图 3　风险矩阵示例

五、教育领域生成式人工智能应用的伦理风险评估

风险可能性与严重性的计算一般多以统计数据作为基础,需要在大量数据之上进行模拟推算。因此,研究将风险发生的可能性表示为社会公众对该风险的关注度,社会讨论热度越高,则风险发生的频率越高。对于公众而言,主要进行探讨的平台是各大社交媒体,可选择“百度”平台为代表进行数据检索,评估各风险发生的可能性;对于专家学者而言,选择“中国知网”论文数据库为代表进行数据检索。风险发生的严重性表示为社会公众对该风险的认知态度,主要对涵盖公众认知调查的相关文献进行整合,基于现有文献数据,设定伦理风险严重性的合理范围。认知调查数据包括专家评估的伦理风险指标数据与公众伦理风险认知的问卷调查数据。本研究参考李梦薇等提出的计算方法[21],在搜索引擎与知识资源数据库中检索伦理风险相关文章数量,根据风险认知调查文章归纳严重性平均值,得到伦理风险的综合等级。

(一)评估方法

1. 可能性

根据图 1 的伦理风险分类,将伦理风险中的四大类(技术本体风险、教育数据风险、机器算法风险、教育应用风险)依次编号为 A、B、C、D;四大类风险下的具体细化项为次要风险,通过序列编码(如 A1、A2、A3 等,B1、B2、B3 等)进行编号;将检索平台百度和知网编号为 i=1,2。在社交媒体与知识资源数据库中,使用相应的关键词进行数据检索,重点统计发文时间自 2022 年 11 月 30 日(ChatGPT 发布日)以来的文章。每一风险在各平台选取数量为 K 的文章进行数据处理,为保证数据量的有效和可操作,要求 K 为大于等于 200 的定值,若数据量不足 K,则按实际数据量计算。以技术本体风险(A)为例,风险 A 的检索数据总量为 K_A,百度平台中检索出的文章数量为 K_{Ai} (i=1)。风险 A 的文章总数为:$\sum_{i=1,2}K_{Ai}$。风险 A 发生的综合可能性为:$P_A=\frac{\sum_{i=1,2}K_{Ai}}{K_A}$。计算得出风险可能性,将可能性等比调整至 1~5 之间的等级范围,以便在风险矩阵中更好地进行比较。

2. 严重性

风险严重性的主要指标为社会公众对于不同风险认知态度的量化分析。具体而言,通过文献调查法,统计现有人工智能伦理认知调查相关文献数据,将结果等比转化为 1~5 之间的等级范围,以反映风险发生影响的严重性。以技术本体风险(A)为例,风险 A 综合严重性为数据中等比转换相关风险严重性后的平均值。

(二)实践分析

图 4 为教育领域生成式人工智能的伦理风险管理框架,分为“风险识别—风险分析—风险评估—风

险应对”四级管理流程。风险识别层,根据现有的生成式人工智能教育伦理研究,归纳总结风险发生阶段和诱发风险因素,构建风险源检查表。风险分析层,利用风险矩阵,统计以风险热度为依据的风险可能性和以社会认知为主的风险严重性。风险评估层,对比伦理风险等级准则和风险综合等级评价表,得到各伦理风险等级。风险应对层,在生成式人工智能教育过程中持续监测并强化管理过程,逐级缓解伦理风险。

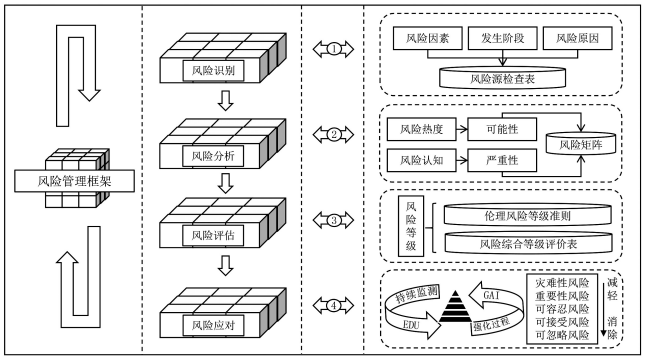

图 4　教育领域生成式人工智能的伦理风险四级管理框架

在“百度”中,检索技术本体风险相关文章,检索关键词为“生成式人工智能”“ChatGPT”“支配”“教师”“学生”;在“知网”中,采用的检索式为 SU=[(ChatGPT+生成式人工智能+GAI+AIGC)AND(教育+教师+学生+主体+支配)]。检索关键词均需符合对该风险的描述。每类风险需检索阅读 200 篇文章（存在部分风险文章不满 200 篇的情况），并判断检索文章与风险的相关程度。其中，在百度平台中，得到各风险文章数量为 A“48”、A1“84”、A2“66”、B“126”、B1“154”、B2“57”、B3“124”、C“86”、C1“164”、C2“164”、C3“31”、C4“147”、D“85”、D1“68”、D2“59”、D3“121”、D4“113”。随后,根据公式得出各伦理风险发生的综合可能性，得到 P_A=2、P_B=3、P_C=2、P_D=3。而后,根据“知网”中伦理风险认知调查文章归纳公众对风险的认知，等比计算伦理风险的综合严重性,得到 I_A=3、I_B=5、I_C=5、I_D=4。最后计算各伦理风险等级,得到风险综合等级评价,见表 3。

目前教育领域应用生成式人工智能的过程中,各种类型的伦理风险已经显现，大多数风险都需要充分、适当且有效的措施进行风险管理。观察风险综合等级评价表可知,主要伦理风险中教育数据风险的风险等级最高,其次是教育应用风险、技术本体风险与机器算法风险。对比风险矩阵可知,次要伦理风险中数据泄露风险等级最高,为灾难性风险;重要性风险为算法同质推荐风险。因此,针对最需管理的教育数据泄露风险与教育应用风险,需要加强生成式人工智能针对教育敏感数据的安全防护措施,规范数据授权范围,谨防数据泄露[22];在人工智能应用的过程中,还需要清晰界定责任主体,完善人工智能的法律责任规制研究,发展负责任的人工智能,防止教育主体的错位。同时,从风险源视角可以看出,教育领域中最需加强生成式人工智能使用阶段的风险管理,这需要在明确教育目的后制定生成式人工智能使用原则与规范,解决使用中的主要矛盾。

六、结 束 语

生成式人工智能是影响教育创新发展的关键性技术，实施伦理风险评估既是稳定发展技术的必要条件,也是实现高质量教学的有力支撑,更是发展负责任人工智能的根本保障。本文讨论了教育领域生成式人工智能伦理风险的识别与分析问题，从设计与使用阶段探究各风险发生的成因,基于伦理风险分类和风险矩阵,评估伦理风险发生的可能性与严重性，为管理教育中

表 3　　风险综合等级评价表

主要伦理风险	P	I	R	说明	次要伦理风险	P	I	R	说明
技术本体风险(A)	2	3	6	可容忍风险	技术拜物教风险(A1)	2	3	6	可容忍风险
					技术依赖风险(A2)	2	3	6	可容忍风险
教育数据风险(B)	3	5	15	重要性风险	数据泄露风险(B1)	4	5	20	灾难性风险
					数据失真风险(B2)	1	2	2	可忽略风险
					数据滥用风险(B3)	2	4	8	可容忍风险
机器算法风险(C)	2	3	6	可容忍风险	算法歧视与偏见风险(C1)	3	1	3	可接受风险
					算法同质推荐风险(C2)	4	4	16	重要性风险
					算法固化风险(C3)	1	1	1	可忽略风险
					算法不透明风险(C4)	3	3	9	可容忍风险
教育应用风险(D)	3	4	12	重要性风险	师生关系重构风险(D1)	2	3	9	可容忍风险
					自主权丧失风险(D2)	1	4	4	可接受风险
					教育公平损害风险(D3)	2	4	8	可容忍风险
					问责逃避风险(D4)	3	3	9	可容忍风险

的生成式人工智能伦理风险提供了一定的参考价值。

后续的研究需要进一步考虑针对具体事件的伦理风险管理，并根据风险等级逐级提出风险应对建议。伴随新技术、新应用的出现，生成式人工智能在教育领域的发展过程中会不断出现新的风险。因此，还需要源源不断吸纳新的风险因素，开发合乎人工智能伦理的风险预警系统[23]，强化风险管理流程，助力生成式人工智能的可持续发展。

[参考文献]

[1] 卢宇，余京蕾，陈鹏鹤，等. 生成式人工智能的教育应用与展望——以 ChatGPT 系统为例[J]. 中国远程教育，2023，43(4)：24-31，51.

[2] 杨海燕，李涛. ChatGPT 教学应用：场景、局限与突破策略[J]. 中国教育信息化，2023，29(6)：26-34.

[3] 王佑镁，王旦，梁炜怡，等. “阿拉丁神灯”还是“潘多拉魔盒”：ChatGPT 教育应用的潜能与风险[J]. 现代远程教育研究，2023，35(2)：48-56.

[4] MIAO F C，HOLMES W. Guidance for generative AI in education and research [EB/OL]. (2023-09-07)[2024-02-04]. https://unesdoc.unesco.org/ark:/48223/pf0000386693.

[5] 国家互联网信息办公室等七部门.生成式人工智能服务管理暂行办法[EB/0L]. (2023-07-10)[2024-02-18]. https://www.gov.cn/zhengce/zhengceku/202307/content_6891752.htm.

[6] 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.

[7] 刘磊，刘瑞. 人工智能时代的教师角色转变：困境与突围——基于海德格尔技术哲学视角[J]. 开放教育研究，2020，26(3)：44-50.

[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] 陈爱华.高技术的伦理风险及其应对[J]. 伦理学研究，2006(4)：95-99.

[10] MITCHAM C，MACKEY R. Philosophy and technology[M]. New York：Simon and Schuster，1983：293.

[11] 汪怀君. 技术恐惧与技术拜物教——人工智能时代的迷思[J]. 学术界，2021(1)：197-209.

[12] 杨现民，唐斯斯，李冀红. 发展教育大数据：内涵、价值和挑战[J]. 现代远程教育研究，2016(1)：50-61.

[13] 沈苑，汪琼. 人工智能教育应用的偏见风险分析与治理[J]. 电化教育研究，2021，42(8)：12-18.

[14] 国家互联网信息办公室，中华人民共和国工业和信息化部，中华人民共和国公安部，等. 互联网信息服务算法推荐管理规定[EB/OL]. (2022-01-04)[2024-02-18]. https://www.cac.gov.cn/2022-01/04/c_1642894606364259.htm.

[15] BURRELL J. How the machine ´thinks´：understanding opacity in machine learning algorithms[J]. Big data & society，2016，3(1)：1-12.

[16] 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.

[17] 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，2023：55-58.

[18] 李素鹏. 风险矩阵在企业风险管理中的应用[M]. 北京：人民邮电出版社，2013.

[19] GARVEY P R，LANSDOWNE Z F. Risk matrix：an approach for identifying，assessing，and ranking program risks [J]. Air force journal of logistics，1998，22(1)：18-21.

[20] 李树清，颜智，段瑜. 风险矩阵法在危险有害因素分级中的应用[J]. 中国安全科学学报，2010，20(4)：83-87.

[21] 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.

[22] 郝祥军，顾小清，王欣璐. 回避还是规避：风险社会中的教育危机与安全防线[J]. 电化教育研究，2023，44(1)：42-47，90.

[23] 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.

Research on Ethical Risk Management Framework for Generative Artificial Intelligence Application in Education

WANG Youmei，WANG Xinying，LIU Chenchen

(Research Center for Big Data and Smart Education，Wenzhou University，Wenzhou Zhejiang 325035)

[Abstract] Generative artificial intelligence has had a revolutionary impact on content production and

(下转第 42 页)

Analysis of the Influencing Factors and Complex Configuration Paths of Digital Transformation in Vocational Education —A Mixed Empirical Study Based on the "Resource-Dynamic" Model

DONG Tongqiang, CHEN Ronglong, XU Zhenguo
(School of Communication, Qufu Normal University, Rizhao Shandong 276826)

[Abstract] Promoting the deep integration of digital technology and vocational education is not only a fast track to deepen the construction of modern vocational education system, but also a new track to help vocational education in the new era to move from "great potential" to "making significant achievements". Based on resource-based theory and dynamic-capability theory, the study constructed a "resource-dynamic" model. The study took 26 vocational colleges and universities in China as case samples and employed the fuzzy-set Qualitative Comparative Analysis (fsQCA) to identify the linkage effect and adaptive selection strategy of internal resource endowment and external dynamic in promoting digital transformation of vocational education. It is found that: (1) the single antecedent condition such as technological resources, organizational resources, environmental resources, government support, school competition, and enterprise demand cannot independently become the necessary condition for the high-level construction of digitalization in vocational colleges and universities; (2) there exist three driving models to promote the digital transformation of vocational education: the resource utilization-led, the external motivation-led, and the resource-dynamic coupling-led. The study provides theoretical guidance for accelerating the high-level digital construction of vocational education and offers practical insights for various vocational colleges and universities to open up the new phase of high-quality development of vocational education.

[Keywords] Vocational Education; Digital Transformation; Fuzzy-set Qualitative Comparative Analysis; Internal Resource Endowment; External Motivational Forces

---

(上接第 34 页)

dissemination, and education has stepped into an era of human-computer collaboration in which intelligence and risk coexist. The ethical risks of educational artificial intelligence application occur frequently, and there is an urgent need for scientific management through standardized risk assessment tools. Starting from technological and educational dimensions, this paper summarized the ethical risks of generative artificial intelligence application in education into four main categories: technological ontology risk, educational data risk, machine algorithm risk, and educational application risk. Drawing on the ISO31000 risk management standards and application guidelines, this paper built an ethical risk management framework with a four-level process of "risk identification - risk analysis - risk assessment - risk response". A checklist of risk sources was constructed, with the risk matrix as the main assessment tool to statistically conclude the comprehensive level of various ethical risks in practice. This framework can provide reference for the effective implementation of ethical risk management in teaching and education.

[Keywords] Generative Artificial Intelligence; Ethical Risks; AI Trained for Education; Risk Management; Risk Assessment