人工智能教育应用的伦理问题及其治理对策

AI ethics in AI Educational Applications and their Governance Strategies

于越[1]，赵建华[2*]
[1]西北师范大学教育技术学院
[2]南方科技大学未来教育研究中心
* zhaojh@sustech.edu.cn

【摘要】 当今社会，人工智能的迅猛发展对人类生产生活产生了深远影响，人工智能在教育领域的探索与应用方兴未艾，与此同时，人工智能在教育应用中的伦理问题也愈加突出。人工智能伦理概念的形成，历经科幻文学、问题浮现与机器人伦理、理论探讨、规范制定阶段，国际上在人工智能伦理问题和治理原则已经达成了广泛共识。近年来，人工智能已广泛在教育中应用，出现了数据隐私泄漏风险、算法偏见与歧视、学术诚信危机、能力缺失和心理健康等问题。人工智能教育应用中伦理问题的治理，需要政府、教育组织、学校等多方共同参与。在顶层设计层面，要制定人工智能教育应用的法律规范；在教育组织层面，要制定人工智能教育应用的政策指南和指标体系；在学校层面，要积极制定指导性文件开展人工智能伦理教育。
【关键词】 人工智能教育应用；人工智能伦理；伦理治理

Abstract: *Nowadays, the rapid development of artificial intelligence (AI) has had a profound impact on human production and life. The exploration and application of AI in the field of education are burgeoning, while at the same time, ethical issues in its educational applications have become increasingly prominent. The formation of the concept of AI ethics has gone through stages of science fiction literature, problem emergence and robot ethics, theoretical exploration, and the formulation of norms. Internationally, significant consensus has been reached on AI ethical issues and governance principles.In recent years, AI has been widely applied in education, giving rise to issues such as risks of data privacy breaches, algorithmic bias and discrimination, academic integrity crises, skill deficiencies, and mental health concerns. Addressing the ethical challenges in AI educational applications requires the collective participation of multiple stakeholders, including governments, educational organizations, and schools. At the top-level design stage, it is essential to establish legal regulations for AI applications in education. At the educational organization level, policy guidelines and indicator systems for AI in education should be developed. At the school level, it is crucial to actively formulate guiding documents and implement AI ethics education.*
Keywords: AI Educational Application, AI ethics, AI ethics Governance

1.引言

当今社会，人工智能的迅猛发展对人类生产生活产生了深远影响，尤其是随着 ChatGPT 和 Deepseek 的爆火，人工智能在教育领域的探索与应用方兴未艾，受到社会的高度关注。一方面，人工智能的合理应用在推进因材施教、促进教育公平、提升教育质量、破解教育难题、支持终身学习、构建高质量教育体系等方面展现出显著优势；另一方面，人工智能的不当应用也带来价值观、道德伦理、隐私保护、学术诚信、偏见与歧视等社会问题，还涉及知识产权侵犯、数据安全和法律合规风险。人工智能是指用机器模拟、实现和延伸人类的感知、

思考、行动等智力与行为能力的科学与技术(中国信通院，2021)，主要包括自然语言处理、语言识别、图像识别和处理、情绪检测、人工创作等（UNESCO, 2021）。本研究中的人工智能主要指生成式人工智能，即具有文本、图片、音频、视频等内容生成能力的模型与相关技术（生成式人工智能服务管理暂行办法，2023）。据《纽约时报》报道，2024 年 2 月，一名 14 岁的美国男孩痴迷聊天机器人自杀身亡（Roose, 2024），让人工智能在教育应用中的伦理问题备受关注。本研究将对人工智能教育应用中的伦理问题及其治理现状展开论述。

2.人工智能伦理的定义

2.1. 人工智能的由来

人工智能（Artificial Intelligence，简称 AI）之父，英国著名学者阿兰图灵在 1950 年发表的一篇划时代意义的论文《计算机器与智能》中提到“机器是否能够具有思维”的重大问题。图灵的论文直接引发了人类关于人工智能的思考（Turing, 1950）。1956 年，麦卡锡、西蒙等科学家在美国达特茅斯学院召开的研讨会上正式提出了人工智能这个概念，随后，经历 IBM 深蓝智能计算机打败人类国际象棋冠军、谷歌公司的 AlphaGO 打败人类围棋世界冠军李世石、ChatGPT 及具备强大推理能力的 o1 问世等标志性事件(赵建华，2024)，人工智能对社会各行各业产生了深远的影响。虽然人工智能这一概念目前已经为社会和科学领域所广泛接受，但是无论在工业界还是学术界，不同发展时期、不同领域和行业、不同学科的学者、专家以及社会各界人士对人工智能仍有不同角度和层面的理解。关于人工智能的各种定义可以归结为：人工智能是研究智能的机制和规律、构造智能机器的技术和科学。通俗来讲，人工智能可以被看作是一种让计算机系统具备类似人类智能的技术，它使机器能够像人类一样思考、学习和解决问题，甚至在某些领域超越人类的能力，人工智能是研究如何使机器具有智能的科学。

2.2. 人工智能伦理的由来

人工智能伦理的概念形成是一个多维度、跨学科的过程，其发展历程可以概括为以下几个关键阶段和因素。

2.2.1. 科幻文学的人工智能伦理（19 世纪初-20 世纪中）

人工智能所引发的伦理思考，最早并不是现实的技术进步，而是科幻小说。早期科幻作品已开始探讨机器与人类的伦理关系，1818 年玛丽雪莱创作的《弗兰肯斯坦》、1920 年卡雷尔卡佩克发表的《罗萨姆的万能机器人》以及 1940 年阿西莫夫在科幻小说《我，机器人》中提出了“机器人三定律”。

2.2.2. 从问题浮现到机器人伦理（20 世纪中期至 21 世纪初）

人工智能的概念在 20 世纪 50 年代诞生，早期研究集中在逻辑推理和基础算法。人工智能之父图灵在早年论文《智能机器》中提到了人工智能迟早会威胁到人类的生存。20 世纪 60 年代，美国哲学家休伯特德莱弗斯将人工智能专家称作“炼金术士”，呼吁人工智能专家在产品设计和创造时融入人类社会的哲学思考。这一时期，人工智能伦理也成为一批科幻小说和影视作品的主题思想，比如《银翼杀手》《黑客帝国》。进入 21 世纪，关于机器人伦理学、法律和社会问题在学术和专业方面得到重视和研究，并在 2004 年第一届机器人伦理学研讨会上提出了“机器人伦理学”的概念。2005 年，欧洲机器人研究网络设立“机器人伦理学研究室”。欧盟建立机器人伦理学 Atelier 计划，也就是“欧洲机器人伦理路线图”。

2.2.3. 人工智能伦理的理论探讨（21 世纪初-2014 年）

随着计算能力的提升，AI 逐步从理论走向应用。从 2005 年以后，机器人伦理学逐渐延伸到人工智能伦理，并逐渐受到全球各界专家、学者以及政府和企业的关注。计算机科学、哲学、伦理学等领域的学者开始跨学科研究合作，探讨 AI 的潜在伦理挑战。出现算法的公平性以及数据偏见导致决策歧视，如在招聘、信贷中的算法偏见；自主性与责任问题，比如自动驾驶汽车在事故中的责任划分问题。一些新的理论框架构建，功利主义、义务论等传统伦理学理论被重新解读，以适配 AI 场景。

2.2.4. 人工智能伦理的规范制定（2015 年至今）

在这一时期，谷歌微软等科技巨头发布 AI 伦理原则，强调公平、透明、可解释性等。各国政府陆续出台一系列政策法规。经合组织（OECD）、联合国教科文组织（UNESCO）等推出全球性 AI 伦理框架。根据联合国教科文组织在 2021 年发布的《人工智能伦理建议书》，人工智能伦理（AI Ethics）是确保 AI 技术以尊重人权、多样性、环境可持续性和社会正义的方式发展的实践框架(UNESCO, 2021)。

3.人工智能伦理教育应用中的主要挑战

3.1.数据隐私泄漏风险

人工智能收集学生生物特征、学习行为、情绪状态等敏感数据可能侵犯隐私，且数据使用缺乏透明度。《2024 年数据泄露调查报告》显示，30,458 起数据安全事件，其中 10,626 起已确认数据发生泄露，教育行业（1537 起）成为年度数据泄露事件发生最严重的领域，内部误操作、错误配置、外发泄露、外部攻击者的勒索入侵、漏洞攻击等问题交织发生，是该行业数据泄露的元凶(Verizon, 2024)。在疫情期间广泛用于远程考试监控的 Proctorio 在线监考软件，通过摄像头、麦克风及屏幕录制跟踪学生行为，甚至分析眼球运动和背景环境，学生投诉其收集过多数据且未明确告知用途（Monica Chin, 2021）。一些学校也开始“试水”用人脸识别来保证出勤率和提高听课效率，人脸识别系统进入中国药科大学课堂后学生担忧，此举识别有侵犯个人隐私之嫌，一举一动都被人脸摄像头监视（于珍，2019）。

3.2.算法偏见与歧视

生成式人工智能技术的开发、训练和更新取决于海量的数据，训练数据中的偏见导致系统推荐资源或评估学生时存在歧视，出现数字贫困、性别歧视，加剧教育不公平问题。训练数据和算法技术的来源大多源自欧美发达国家的科技公司，同时训练的数据大多也是基于英文，教科文组织将这种行为称为“数字贫困”（Digital Poverty），认为这将导致发展中国家的价值观、文化习俗、行为规范受到西方文化的侵袭，由此产生知识的同质化限制多元化和创造性思维(UNESCO, 2023)。2020 年英国政府使用算法调整 A-level 考试成绩，导致弱势地区学生成绩被系统性压低，引发大规模抗议（A-Levels, 2020）。

3.3.学术诚信危机

生成式人工智能能够通过与学习者在多轮对话情境下生成具有个性化的知识，但更可能加剧作弊和学术不端的诚信危机。教育数据公司麦可思发布的 2024 年中国高校师生生成式 AI 应用情况研究显示，国内高校师生几乎都曾在学习和工作中使用过生成式 AI，从未使用过生成式 AI 的高校师生比例仅为 1%。其中，近六成高校师生频繁使用生成式 AI(赵丽 & 陈颖， 2024)。在国际上，已经有诸多国家与地区指出智能技术将可能加剧课业、考试的作弊和抄袭风险，将限制该技术的使用。

3.4.能力缺失和心理健康

使用人工智能进行学习，可能导致个体不再需要通过和教师与同伴在社会互动和实践中接受教育，从而弱化教育促进学生社会交往和集体学习的重要性，限制学生必要的社会能力获取，如表达技巧、社会协作与问题解决能力。教科文组织在《报告》中提出，这将造成学习者缺乏对真实世界的观察，通过实践获取直接经验的能力，以及与他人交流、合作和处理现实世界中问题的能力(UNESCO, 2023)。同时，人工智能监控学生情绪或学习状态可能加剧学生焦虑，甚至通过算法推送内容操控行为，可能出现引发更大的心理健康问题，出现更多类似于 14 岁的美国男孩痴迷聊天机器人自杀身亡的事件（Roose, 2024）。

4.人工智能教育应用伦理的治理对策

通用大模型时代，生成式人工智能的教育应用不断“制造”出多种新型伦理交往和道德生活场景，成为教育人工智能伦理治理的重心。只有伦理先行，才能为人工智能与教育融合创新提供保障，并且要通过敏捷治理的理念、伦理教育的手段和伦理审查方案的制定三者相辅相成，提供强有力的支撑，需要政府部门、教育组织、学校、社会团体等多方共同参与。

4.1.顶层设计：制定人工智能教育应用的法律规范

首先，在国家层面应该尽快出台人工智能教育应用的政策和执行标准，明确教育行业的发展方向和规范化要求，具体内容应该包括技术开发、数据保护、教学内容和课程设置。通过比较美国、欧洲、亚洲及联合国教科文组织等人工智能教育应用政策，发现总体上有很多共同之处，在应用中注重谨慎、态度比较乐观、坚持决策自主(吴河江&吴砥，2024)。

4.2.教育组织：制定人工智能教育应用的政策指南和指标体系

从 2019 年到 2024 年，联合国教科文组织出台了《关于人工智能与教育的北京共识》《人工智能伦理问题建议书》《生成式人工智能在教育和研究中的应用指南》《学生人工智能能力框架》《教师人工智能能力框架》等近 10 份人工智能教育应用指南和能力框架。联合国教科文组织形成了“赋能”“比较”“协同”“参与”四种教育数字治理策略，建立了治理主体多元、治理方式刚柔相济和治理过程全周期的治理机制(朱莉， 2024)。各国各组织教育人工智能伦理原则参照已有原则达成了很多“共识”的构建取向，如隐私、公平、安全、问责等（白钧溢&于伟，2023）。在目标层面上指导政策的实施和评估，在实施层面上鼓励各方的协同参与，在方案层面上构筑数据的隐私屏障，可以打破目前技术应用中存在的伦理困境。

4.3.学校行动：积极制定指导性文件开展人工智能伦理教育

人工智能伦理教育的开展需要学校与社会、家庭密切配合。美国麻省理工学院（MIT）在青少年人工智能伦理教育领域的探索，涉及社会、家庭和学校三个方面，包含了内嵌式和专题式两种类型，主要的伦理教育形式有与中小学合作的线下 STEAM 选修课程、线上线下夏令营相关活动以及在社会上招募学生参与的混合式工作坊(李艳等，2024)。罗素大学集团内的成员高校对生成式人工智能教育伦理风险的总结为数据陷阱冲击教育管理系统、算法支配损害教师主体权威、智能依赖造成学生学习畸化三个方面，并且率先积极应对、及时发布生成式人工智能应用指南，从教学评估监管、教师教学配套、学生使用培训三个方面的具体举措（张惠彬&许蕾，2024）。哈佛大学、耶鲁大学、剑桥大学、东京大学、南洋理工大学、悉尼大学等 16 所 2024 年 QS 世界大学排名前 100 的世界一流高校在生成式人工智能应用方面的指导性文件，提炼出技术应用、道德伦理、课程评价、师生指导四个主题（楚肖燕等，2024）。

5.结语

人工智能在教育领域的应用为教学创新和个性化学习带来了前所未有的机遇，但同时也伴随着数据隐私、算法偏见、学术诚信等伦理挑战。确保人工智能教育的健康发展，需要多方协同治理：政府应完善法律法规以提供制度保障，教育组织需制定伦理准则和评估体系，学校则应加强师生的人工智能伦理教育。唯有在技术创新与伦理约束之间取得平衡，才能让人工智能真正成为促进教育公平、提升学习效能的可持续力量。未来，随着技术的演进，伦理治理框架也需动态调整，以应对新兴问题，最终实现“以人为本”的智能教育生态。

参考文献

白钧溢 & 于伟. (2023). 超越“共识”：教育人工智能伦理原则构建的发展方向. 中国电化教育, 6, 9 – 17, 24.

楚肖燕, 沈书生, 王敏娟, 王会军, 李晓文, & 翟雪松. (2024). 世界一流高校探索生成式人工智能应用规范的经验及对我国的启示——基于 LDA 主题模型分析的文本挖掘. 现代远距离教育, 3, 38 – 47.

李艳, 朱雨萌, & 樊小雨. (2024). 青少年人工智能伦理教育的探索及启示——以 MIT 为例. 现代远距离教育, 1, 3 – 13. https://doi.org/10.13927/j.cnki.yuan.20240022.001

生成式人工智能服务管理暂行办法, Pub. L. No. 国家互联网信息办公室 中华人民共和国国家发展和改革委员会 中华人民共和国教育部 中华人民共和国科学技术部 中华人民共和国工业和信息化部 中华人民共和国公安部 国家广播电视总局令第 15 号 (2023). https://www.gov.cn/zhengce/zhengceku/202307/content_6891752.htm

吴河江 & 吴砥. (2024). 生成式人工智能教育应用:发展历史、国际态势与未来展望. 比较教育研究, 46(6), 13 – 23. https://doi.org/10.20013/j.cnki.ICE.2024.06.02

于珍. (2019). 中国教育报：辅助教学还是泄露隐私，AI 进校园边界在哪里？_教育家_澎湃新闻-The Paper. https://www.thepaper.cn/newsDetail_forward_4758874

张惠彬 & 许蕾. (2024). 生成式人工智能在教育领域的伦理风险与治理路径——基于罗素大学集团的实践考察. 现代教育技术, 34(6), 25 – 34.

赵建华. (2024). 人工智能时代的教育转型与重塑. 电化教育研究, 45(12), 37 – 43, 97. https://doi.org/10.13811/j.cnki.eer.2024.12.005

赵丽, & 陈颖. (2024). 媒体：从“一键生成作文”到数据泄露隐患，学生用 AI 边界何在？_教育家_澎湃新闻-The Paper. https://www.thepaper.cn/newsDetail_forward_30601968

中国信通院. (2021). 《人工智能治理白皮书》. https://www.sohu.com/a/www.sohu.com/a/421423992_735021

朱莉. (2024). 联合国教科文组织教育数字治理的机制与策略——基于 44 份文本的 Nvivo 质性分析. 比较教育学报, 5, 31 – 44.

A-levels: Anger over 《unfair》 results this year. (2020, 八月 12). https://www.bbc.com/news/education-53759832

Monica Chin. (2021). University will stop using Proctorio remote testing after student outcry | The Verge. https://www.theverge.com/2021/1/28/22254631/university-of-illinois-urbana-champaign-proctorio-online-test-proctoring-privacy

Roose, K. (2024, 十月 23). Can A.I. Be Blamed for a Teen’s Suicide? The New York Times. https://www.nytimes.com/2024/10/23/technology/characterai-lawsuit-teen-suicide.html

UNESCO. (2021). 人工智能与教育：政策制定者指南. https://unesdoc.unesco.org/ark:/48223/pf0000378648

UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence. https://unesdoc.unesco.org/ark:/48223/pf0000381137

UNESCO. (2023). Guidance for generative AI in education and research — UNESCO Digital Library. https://unesdoc.unesco.org/ark:/48223/pf0000386693.locale=en

Verizon. (2024). 4900万份客户数据遭泄漏；教育行业既成为年度数据泄露事件发生最严重的领域？_科学技术_人工智能_信息. https://www.sohu.com/a/www.sohu.com/a/780623261_121728280