

# 教育领域人工智能伦理框架:演进脉络与构建原则

吴 砥<sup>1</sup>, 桂徐君<sup>2</sup>, 白钧溢<sup>3</sup>, 陈 旭<sup>3</sup>

(1.华中师范大学 教育部教育信息化战略研究基地(华中), 湖北 武汉 430079;

2.华中师范大学 国家数字化学习工程技术研究中心, 湖北 武汉 430079;

3.华中师范大学 人工智能教育学部, 湖北 武汉 430079)

**[摘要]** 人工智能在赋能教育的同时,引发了对公平、隐私与透明等伦理风险的深刻关切。伦理框架是开展伦理治理、应对伦理风险的关键介质,其成熟建立有赖于伦理原则的不断演进与完善。回顾发展历程,人工智能伦理原则历经观念设想、议题形成、框架构建与落地探索的阶段演进,呈现将抽象伦理价值落实为标准化且可执行的规范框架的发展趋势。受制于发展周期较短、参与主体多元等因素,教育领域人工智能伦理原则的标准化与可执行演进面临要素过泛与边界不清两大瓶颈。为此,文章提出应将提升“契合性”与“可实施性”确立为教育领域人工智能伦理原则的未来方向,并据此提出伦理框架的构建原则,为风险治理及后续应用的伦理评估提供理论与方法支持。

**[关键词]** 教育数字化; 伦理框架; 演进脉络; 构建原则

**[中图分类号]** G434

**[文献标志码]** A

**[作者简介]** 吴砥(1978—),男,湖北洪湖人。教授,博士,主要从事教育数字化规划与评估、数字素养监测与评价研究。E-mail:wudi@mail.ccnu.edu.cn。陈旭为通信作者,E-mail:chenxu@mail.ccnu.edu.cn。

## 一、引言

在教育领域,人工智能技术的迅猛发展引发了对教学模式、学习方式及育人理念的深层变革<sup>[1]</sup>。从知识内容的生产与传播,到学习过程的数据驱动与行为画像,再到师生互动的情感支持与动力激发,人工智能在教育生态中的深度渗透已成为无法回避的现实趋势<sup>[2]</sup>。然而,人工智能在带来机遇的同时,也对平等、包容、学习主体能动性教育价值造成直接冲击<sup>[3]</sup>。例如,生成式人工智能的应用造成对学生批判性思维及原创力削弱的担忧<sup>[4]</sup>;人工智能教学辅助的引入带来对教师教学主导性削弱的质疑<sup>[5]</sup>;个人信息在模型训练和推理中的使用引发对数据安全的关切<sup>[6]</sup>。这些现象均折射出教育生态面对人工智能所需承担的伦理责任与规范需求。

为应对人工智能赋能教育过程中出现的伦理风险,如何建立适切的伦理框架已成为全球各国推进人

工智能与教育融合发展的焦点问题。一个系统化的伦理框架,不仅能够为教育实践者和技术开发者提供行为边界与价值指引,还将成为协调政策制定、标准确立和技术治理的重要基础。然而,虽已有国际组织与部分国家在人工智能伦理治理方面制定了指导性原则,但针对教育情境的系统性伦理框架仍显匮乏,在价值维度的领域选择、原则的适配性以及落地实施等方面缺乏可操作的方法论体系。基于此,本文将从人工智能伦理原则的演进历程出发,厘清教育领域人工智能伦理原则的总体趋势、现实格局及发展方向,推动人工智能技术在保障教育核心价值的前提下持续创新与健康发展。

## 二、人工智能伦理原则的演进脉络

从技术哲学的“技术—伦理”演化视角来看,伦理思考往往在技术的萌芽阶段即已出现,并随技术自身的成熟和社会化应用的推进不断深化、规范化,直至

形成制度化框架。人工智能伦理原则的演进与这一规律高度契合,其历程可划分为观念设想、议题形成、框架构建、落地探索四个递进阶段,如表1所示。

(一)观念设想(20世纪40年代至20世纪70年代中)

人工智能伦理原则的萌芽可追溯至20世纪40年代。人工智能所引发的伦理思考,最早并非来自现实技术发展,而是科幻小说与哲学设想<sup>[7]</sup>。1942年,阿西莫夫(Asimov I)在短篇小说《转圈圈》中首次系统提出著名的“机器人三定律”,开启了现代人工智能和机器人伦理的讨论。1948年,图灵(Turing A)在名为《智能机器》(Intelligent Machinery)的内部报告中,分析了“思维”“意识”“智能”和“机器”等概念,首次将人工智能与哲学作为课题讨论<sup>[8]</sup>。同年,维纳(Wiener N)探讨了超高速计算机造福与危害人类的双重可能。1960年,维纳进一步预测机器智能有朝一日会超越人类并对人类造成危害。与之类似,数学家欧文(Irving G)在《有关第一台超智能机器的猜想》中预测计算机将超越人类的智慧能力,而“第一台超智能机器将是人类的最后一项发明”<sup>[9]</sup>。这一时期的伦理思考主要源于科幻文学和早期控制论的哲学反思,本质上是一种先验于技术发展的伦理观念设想,虽缺乏实证支持,但深刻影响着后续人工智能研究的伦理议程与制度取向。

(二)议题形成(20世纪70年代末至2003年)

20世纪70年代起,人工智能伦理规范的相关议题逐渐形成。1976年,美国应用伦理学家迈纳(Maner W)正式提出“计算机伦理学”概念,并指出其应当作为独立学科存在<sup>[10]</sup>。20世纪80年代起,计算机与网络技术的发展与应用,使其对就业、隐私等社会领域的影响逐渐显露,计算机技术伦理规范及原则成为学界

与专业机构的热点议题。在学术层面,1985年10月,美国著名杂志《形而上学》发表了《计算机与伦理学》(A Computer Ethics Bibliography)和《什么是计算机伦理学》(What is Computer Ethics)两篇文章。其中,《什么是计算机伦理学》提出计算机伦理学的“四步法”,奠定了计算机伦理原则制定的基本思路<sup>[11]</sup>。在规范化进程上,1992年,美国计算机学会(ACM)发布《计算机伦理与职业行为准则》,提出包含为社会福祉作出贡献、避免伤害他人、诚实守信、公平、尊重知识产权、基于知识产权适当的信任、尊重隐私、保密八大要素的计算系统开发一般原则<sup>[12]</sup>。在国际范围的相关主题交流方面,1995年,首届欧洲计算机伦理学会议在英国召开;1997年,“计算机伦理学—哲学”研讨会在荷兰召开。

相对而言,我国人工智能伦理研究起步较晚,早期主要依托译介国外研究成果。20世纪70年代中后期,《自然辩证法杂志》与《摘译(外国自然科学哲学)》译介了国外人工智能哲学系列论文,推动学界初步接触机器自主性、责任划分、社会影响评估等伦理议题。1978年,《1978—1985年自然辩证法发展规划纲要(草案)》在国家学科发展议程中首次明确将“人工智能的哲学问题”列为重点研究方向。1981年,中国社会科学院哲学研究所等三院所联合召开“人工智能哲学问题座谈会”,这是国内首次专门聚焦人工智能哲学的学术会议,讨论了人工智能与其他学科及人与机器的关系<sup>[13]</sup>。同年,中国人工智能学会(CAAI)正式成立,并由哲学家童天湘担任副理事长,显示了对人工智能哲学及伦理议题的重视。20世纪90年代起,我国人工智能伦理研究出现“社会转向”,关注人工智能技术对经济、文化及社会结构可能造成的伦理影响及

表1 人工智能伦理原则的发展阶段

发展阶段	时间	背景要素	标志性事件	核心特征
观念设想	20世纪40年代至70年代中	科幻文学的想象推动对人工智能本质的哲学反思	1942年,阿西莫夫提出“机器人三定律”; 1948年,图灵首次将人工智能与哲学作为课题讨论	科幻场景驱动; 哲学逻辑推演; 缺乏实证研究
议题形成	20世纪70年代末至2003年	计算机技术普及引发就业、隐私等现实议题	1976年,迈纳提出“计算机伦理学”概念; 1985年,摩尔发表《什么是计算机伦理学》文章,明确了计算机伦理原则制定的基本思路; 1992年,美国计算机学会发布《计算机伦理与职业行为准则》	明确学科边界; 聚焦现实议题; 制定初步规范
框架构建	2004年至2021年	人工智能能力指数级增长与社会化应用加速,风险日益凸显	2004年,第一届机器人伦理学国际会议召开; 2010年以来,深度学习等关键技术取得突破	多元主体参与; 领域涉及广泛; 形成价值共识
落地探索	2022年至今	生成式人工智能爆发式发展,治理需求空前迫切	2022年,ChatGPT3.5发布; 2024年,欧盟《人工智能法案》正式生效; 2025年,DeepSeek-R1发布	法律化; 标准化; 工具化

应对,其中1992年出版的《智能革命论》是代表性成果。总体来看,这一时期与人工智能伦理规范相关的议题已开始显现并逐步成形,但在理论框架、议题边界与规范逻辑上仍隶属于宏观的计算机伦理范畴,尚未具备独立的学科性。与国际研究相比,我国对人工智能伦理的探讨虽已起步,但在规范性研究尤其是伦理原则构建方面尚显不足。

### (三) 框架构建(2004年至2021年)

2004年起,人工智能伦理原则研究进入框架构建阶段。这一年,第一届机器人伦理学国际研讨会提出“机器人伦理学”,机器行为模式被纳入伦理视域,标志着人工智能伦理作为独立领域的诞生。次年,欧洲机器人伦理学研究室制作了第一个机器人伦理学路线图,首次明确应考虑尊严、正义、公平、多元化、非歧视、自主权、隐私、利益分享、社会责任以及对生物圈的责任等价值原则<sup>[14]</sup>,奠定了后续框架制定的价值基础。2008年,夏基(Sharkey N)在《Science》发表《机器人的道德前沿》(The Ethical Frontiers of Robotics)一文,呼吁各界尽快协作制定机器人道德规范<sup>[15]</sup>,推动伦理框架构建进入多主体讨论议程。2010年以来,随着深度学习等关键技术取得突破,人工智能的能力呈指数级增长并广泛应用于社会各领域,其伦理问题更加紧迫且具体化,伦理框架构建得到国际组织、各国政府、技术企业的进一步关注,相关框架不断涌现。国际组织层面,如未来生命研究所《阿西洛马人工智能原则》<sup>[16]</sup>、欧盟《可信赖人工智能的伦理准则》<sup>[17]</sup>、联合国教科文组织《人工智能伦理问题建议书》<sup>[18]</sup>;政府层面,如美国《人工智能应用规范指南》<sup>[19]</sup>;技术企业层面,如谷歌《人工智能使用原则》<sup>[20]</sup>、微软《计算未来:人工智能及其社会角色》<sup>[21]</sup>。

与国际对人工智能伦理框架的高度关注相呼应,我国在这一阶段亦加快了相关框架的构建进程。2017年,国务院发布《新一代人工智能发展规划》,全面描绘我国人工智能发展的顶层设计蓝图,明确总体战略目标、重点任务和保障措施,标志我国人工智能进入体系化、国家层面战略推进阶段<sup>[22]</sup>。2019年,为引导人工智能健康发展,国家新一代人工智能治理专业委员会发布《新一代人工智能治理原则——发展负责任的人工智能》,首次从治理角度提出和谐友好、公平公正、包容共享、尊重隐私、安全可控等人工智能需遵循的基本原则,实现从“战略规划”到“价值引领”的转变<sup>[23]</sup>。2020年,科技部等五部门印发《国家新一代人工智能标准体系建设指南》,纳入“人工智能安全/伦理标准”,为落实治理原则提供框架支撑<sup>[24]</sup>。2021年,国家

新一代人工智能治理专业委员会发布《新一代人工智能伦理规范》,进一步细化人工智能研发与应用过程中的伦理要求与边界<sup>[25]</sup>。总体而言,这一阶段国内外不仅完成了人工智能伦理原则由理论议题向框架设计的转变,还在多方参与及跨领域协作中凝聚了相对一致的宏观价值取向,为原则的落地实践、标准化推进及国际规则对接提供了重要保障。

### (四) 落地探索(2022年至今)

2022年以来,人工智能伦理原则加速向落地实践转化。生成式人工智能的爆发式发展加剧了伦理争议与社会影响,凸显伦理治理的紧迫性,并推动人工智能伦理原则深入具体技术实践,相应的伦理标准、行业守则与法律法规相继出台。2022年,OECD发布《人工智能系统分类框架》,提出人工智能系统的类型划分标准,并引入伦理风险维度,旨在为各国在人工智能监管与治理中构建统一的分类及风险识别基础提供参考<sup>[26]</sup>。2023年,美国国家科学技术委员会发布《人工智能风险管理框架》,明确了人工智能风险管理的总体结构,并提出使用定量与定性方法评估人工智能伦理风险的可测量指标<sup>[27]</sup>。同年,全球首份针对人工智能技术的国际性声明《布莱切利宣言》发布<sup>[28]</sup>。2024年,欧盟《人工智能法案》正式生效,标志着全球首个具有全面强制约束力的人工智能立法体系落地,确立了基于风险等级的监管模式,并对高风险人工智能系统提出了合规性要求<sup>[29]</sup>。为推动《人工智能法案》的有效实施,欧盟于2025年发布《人工智能系统定义指南》,旨在对立法所涉人工智能系统范围与技术边界进行细化解释,从而提高技术实践、执法与监管的一致性与可操作性<sup>[30]</sup>。

面对生成式人工智能的快速迭代与伦理议题的不断凸显,我国自2022年以来持续推进人工智能伦理原则的可操作化与制度化建设,实现由引介借鉴到同步发展,再到引领创新的阶段性演进。2022年,中共中央办公厅、国务院办公厅《关于加强科技伦理治理的意见》特别提出应制定并完善科技伦理规范和标准,明确增进人类福祉、尊重生命权利、坚持公平公正、合理控制风险、保持公开透明五项科技伦理原则<sup>[31]</sup>。在此基础上,国家人工智能标准化总体组、全国信标委人工智能分委会《人工智能伦理治理标准化指南(2023版)》进一步梳理细化出以人为本、可持续性、合作、隐私等十类可实施性较强的人工智能伦理原则,并基于标准化视角拆解、具象化各原则的要求,为后续具体标准的研制提供可操作方向<sup>[32]</sup>。同年,科技部等十部门印发《科技伦理审查办法(试行)》,确立了

科技领域伦理审查的基本原则、组织架构、流程规范与监督机制,推动将伦理原则转化为可审查、可落地、可监管的程序<sup>[33]</sup>。法律方面,《中国网络法治发展报告(2024年)》指出,近年来,我国不断加强人工智能立法,初步构建法律、行政法规、部门规章及地方性法规等多层次和数据、算法应用等多维度的人工智能法制框架<sup>[34]</sup>。《国务院2025年度立法工作计划》进一步提出“推进人工智能健康发展立法工作”<sup>[35]</sup>。总体而言,将抽象伦理价值转化为可执行并制度化的规范框架与操作标准,探索价值规范在一致性与兼容性维度上的有效落地,是现阶段及未来一段时间内全球人工智能伦理原则的发展方向。

### 三、教育领域人工智能伦理原则的发展路向

相较于一般人工智能伦理原则,教育领域人工智能伦理原则研究起步相对滞后,但随着生成式人工智能的兴起与应用,相关研究呈现快速升温与加速发展态势。然而,从人工智能伦理原则实践落地的总体发展方向来看,当前教育领域人工智能伦理原则仍存在要素过泛、边界不清等瓶颈问题。因此,提升“契合性”与“可实施性”,应成为未来教育领域人工智能伦理原则发展的重要方向。

#### (一)演进路线回顾

不同于一般人工智能伦理原则历经逾八十年的发展演进,教育领域人工智能伦理原则研究的系统化进程起始于2018年第十九届教育人工智能国际会议召开后。联合国教科文组织国际人工智能研究中心主任霍姆斯(Holmes W)在会议上强调,“当前人工智能在教育领域的研究、开发与部署,很大程度上是在一种‘伦理真空’的状态下进行,尚未形成专门识别并应对其特定伦理挑战的系统性方针、政策或法规”,并呼吁尽快研究并构建能够指导教育领域人工智能开发和应用的伦理原则<sup>[36]</sup>。2019年8月,联合国教科文组织发布首个关注教育领域人工智能伦理原则的指导性文件《北京共识——人工智能与教育》,奠定了全球对教育人工智能伦理的基本认知框架<sup>[37]</sup>。2021年11月,联合国教科文组织通过《人工智能伦理问题建议书》,明确教育领域人工智能的核心价值观与原则,为国际社会提供了统一的伦理指引<sup>[38]</sup>。总体而言,在2019至2022年间,教育领域人工智能伦理原则的发展仍相对缓慢。一方面,这一阶段研究主要集中于宏观层

面的价值理念与通用原则,虽有少数框架尝试结合教育领域特有属性构建针对性原则<sup>①</sup>,但数量与影响力相对有限。另一方面,该阶段研究动力主要源于国际组织与部分国家政府的政策推动,学界与社会关注度、参与度不足,相关研究成果规模较小。

2022年底,生成式人工智能大模型技术的爆发,是教育领域人工智能伦理原则受到广泛关注的起点。以ChatGPT和DeepSeek为代表的大模型技术产品,使人工智能从以往传统的后台辅助工具转变为直接参与内容生产和交互的主体。在教育领域,其以便捷、高效和即时生成内容的特性,迅速引发了学术诚信、隐私保护、内容安全、教育公平等一系列新的伦理挑战。然而,原有面向数据、算法与宏观治理的伦理原则,难以充分应对课堂、作业、考试等微观情境中的复杂问题。2023年初,教育领域进入“应急响应”状态。例如,美国纽约市公立学校于当年1月禁止学生和教师使用ChatGPT,以阻止其对学术诚信造成冲击。此类措施虽短期有效,但也暴露了缺乏长期规范框架的困境。2023年中期,教育领域人工智能伦理原则出现向细化与规范化方向探索的趋势。2023年7月,联合国教科文组织主张教育应发挥对技术的主观引导作用,在建立安全审查规范基础上规划未来发展路线<sup>[38]</sup>。同年9月,联合国教科文组织发布《教育与研究领域生成式人工智能指南》,呼吁各国政府将伦理原则落实至法律法规与教师培训等领域,规范生成式人工智能在教育中的使用<sup>[39]</sup>。2024年下半年以来,教育领域人工智能伦理原则的规范化与可落地发展趋势愈发明显。2024年8月,欧盟《人工智能法案》正式生效,将教育列为人工智能的八大高风险领域之一<sup>[29]</sup>。2025年7月18日,欧盟委员会发布了通用人工智能模型提供商义务范围指南,明确企业在法案生效后的合规义务<sup>[40]</sup>。与此同时,我国教育部基础教育教学指导委员会于2025年5月发布《中小学生生成式人工智能使用指南(2025年版)》,提出应建立生成式人工智能的伦理审查机制,筑牢生成式人工智能在教育领域应用的安全底线<sup>[41]</sup>。总体来看,自2022年底以来,教育领域人工智能伦理原则在发展方向上与一般人工智能伦理原则具有一致性,均指向伦理原则的实践落地。但其发展路线具有鲜明的教育领域特征,经历了从临时应对到情景化探索,再到制度化与法规化建设的发展过程,体现出事件驱动、情境导向、国际引领与本土适

① 如世界经济论坛《人工智能一代:建立儿童和人工智能的全球标准》及凯西·亚当斯(Cathy Adams)等人的《K-12教育的人工智能伦理原则》。

配并行的特征。

## (二) 瓶颈问题剖析

教育领域人工智能伦理原则走向实践落地,既是技术应用深化的必然结果,也具有完善教育治理体系、健全伦理保障机制的现实诉求。然而,受制于发展历程较短、学术积淀不足等因素,当前教育领域人工智能伦理原则存在要素过泛与边界不清两大突出问题。二者互为表里,不仅削弱了伦理原则在教育实践中的指导效能,更阻碍了教育领域相关政策与技术规范的科学制定与系统完善,是伦理原则落地的瓶颈阻碍。

要素过泛表现为伦理原则条目繁多且缺乏统一性,这与其演进路线密切相关。一方面,由于教育领域人工智能伦理原则的学术研究与框架制定起步于通用人工智能伦理原则的框架构建后期,此时通用人工智能伦理框架在原则条目上已臻于成熟,使得教育领域人工智能伦理框架在原则选择与结构设计上普遍依循涉及覆盖面广泛、原则全面的通用人工智能伦理框架<sup>[42]</sup>。另一方面,教育领域人工智能伦理原则并非由单一部门或组织制定,而是政府监管机构、国际组织、行业协会、高校科研院所及教育技术公司等多方共同推动,且在人工智能与教育持续融合的背景下,各类政策、指南及白皮书被频繁颁布或更新,以回应技术应用中的新需求与新挑战。然而,不同时间、主体提出的伦理原则往往并未整合归纳,而是直接并列堆积,加之目前缺乏权威机构牵头对已有伦理原则进行整合、去重与权重排序,导致不同框架中的原则条目“自成一派”。伦理原则的庞杂既导致伦理框架部分内容出现功能交叉甚至冲突,引发原则间的相互掣肘,削弱了规范体系的内部一致性与执行效率,还可能在原则堆叠中淹没本土教育的核心价值与现实诉求,从而限制其在具体教育场景中的针对性与引导效能。

边界不清表现为伦理原则停留在宏观价值层面,缺乏可操作说明与实施指南。教育领域现有伦理框架多聚焦于诸如“公平”“透明”“隐私保护”“责任”等价值性目标,且对具体情境下如何实现这些目标缺乏细化说明。例如,“公平”原则往往笼统描述为“保障教育中人工智能应用的机会平等”,并未指明其在诸如个性化推荐算法设计、学习资源分配或无障碍教学环境建设等实践中的具体衡量指标和标准。教育领域人工智能伦理原则的宏观性特征也使其在落地过程中面临两方面的困境:其一,由于缺乏可量化的指标,伦理原则在实际教学与技术部署中难以转化为清晰、可执行的操作规范。其二,不同技术企业、学校及监管部门在理解与执行同一原则时,往往因缺乏统一的细化说

明而出现差异,不仅使原则实践的效果依赖于各主体的自我解读,还可能导致“伦理购物”(Ethics Shopping)与“伦理漂蓝”(Ethics Washing)问题的出现<sup>[43]</sup>。正是在这一意义上,韦恩·霍姆斯等人认为,概念空泛与内涵模糊是教育领域人工智能伦理原则的“根基性问题”<sup>[44]</sup>。

## (三) 发展方向研判

针对当前教育领域人工智能伦理原则存在的要素过泛与边界不清的瓶颈问题,亟须将提升“契合性”与“可实施性”确立为其未来发展的关键方向。

“契合性”要求教育领域人工智能伦理原则契合教育的育人目标,契合我国的基本国情,深度融合我国教育政策导向与文化价值内核。与一般人工智能伦理原则相比,教育领域的人工智能伦理原则有其自身的价值导向与实践逻辑。这种特殊性不仅源自教育的育人使命与社会公益属性,也深受全球教育合作理念与未来人才培养目标的影响。据此,未来我国教育领域人工智能伦理框架应聚焦教育公平性、隐私安全性、算法透明性、发展导向性四项原则。其中,教育公平性是我国教育政策的基石。《中国教育现代化2035》将“促进公平”列为首要战略任务<sup>[45]</sup>,党的二十大报告进一步明确将教育公平作为教育的重要目标,提出“坚持以人民为中心发展教育,加快建设高质量教育体系,发展素质教育,促进教育公平”<sup>[46]</sup>。教育领域人工智能应确保技术普适和资源普惠,避免算法偏见加剧教育机会不平等,保障弱势群体平等受益权,在多元文化背景下实现包容性设计。隐私安全性是教育领域使用智能技术时的重要关切。教育数据涵盖学生敏感信息,易引发隐私与法律风险。相关原则强调严格遵守法律法规,严控数据收集与使用范围,杜绝数据商品化,通过可审计机制保障学生与家长知情权<sup>[47]</sup>。算法透明性是建立教育信任的关键。教育场景要求人工智能决策逻辑具备可解释性,使教师理解算法依据,避免“黑箱”操作损害专业性,确保利益相关者能够监督与问责。发展导向性强调技术应服务于人的全面可持续发展,是教育变革与技术应用过程中的核心理念。相关原则要求人工智能辅助而非替代教师角色,尊重教师和学生的自主性,通过全球合作缩小数字鸿沟,使技术成为促进社会进步与教育创新的可持续动力<sup>[48]</sup>。

“可实施性”要求教育领域人工智能伦理原则能够转化为具备可计算性、可验证性的操作化框架,通过明确可观测指标、标准化数据采集、厘清算法逻辑,使伦理要求得以嵌入教育人工智能的全生命周期。为此,教育人工智能伦理原则亟须具备“可测评化”特

征,即从规范性表述转向可转译为技术指标、算法逻辑与操作流程的落地体系。可计算人工智能伦理(Computable AI Ethics)旨在将伦理原则转化为可通过算法实现、数据度量与系统验证的操作性框架,提供“如何实现”的路径。基于此,教育人工智能伦理原则的“可测评”可以区分为可量化、可测度、可执行和可验证四个核心要素:可量化(Quantifiable)要求将伦理目标转化为具有明确表达形式和可计算参数的指标体系,为基础测度监控提供支撑;可测度(Measurable)要求建立客观评估伦理目标达成度的量化方法与标准,使原则遵循状况可被追踪比较;可执行(Executable)要求将伦理要求嵌入人工智能运行机制与过程逻辑,使其成为系统决策的内在驱动力;可验证(Verifiable)要求构建系统化验证框架和证据链条,使伦理符合性能够被系统检验、动态调整与历史回溯。

#### 四、教育领域人工智能伦理框架的构建原则

基于教育领域人工智能伦理原则提升“契合性”与“可实施性”的发展方向,伦理框架的构建应遵循以下原则:一是精准识别并定位核心伦理维度,确保维度名称直接映射顶层伦理原则;二是明确定义各维度的教育情境边界与保护对象,避免计算口径模糊;三是设计具备可观测性、可采集性、可比较性的计算指标,覆盖维度内涵中的关键价值点;四是明确数据来源、采集频率、算法公式、精度与阈值标准,确保计算过程透明且可重复验证。

##### (一)教育公平性原则

教育公平性的核心目标在于系统性破除由社会经济地位、文化背景、语言能力、地域条件或身心状态差异所造成的结构性机会壁垒,确保每一位学习者不仅能够无差别接入人工智能教育系统,更可在其中获得与其个性化需求相匹配的发展资源与支持服务。这一原则强调从“形式公平”走向“实质公平”,不仅关注资源在数量上的均等分配,更注重其在质量与适用性上的合理配置。为实现该目标,需要从机会平等与包容多元两个核心维度构建教育公平性的评价框架(见表2),前者强调起点公平,旨在破除技术与资源壁

垒,保障人工智能教育产品在接入与内容上的广泛可及性与普惠性;后者强调过程与结果的公平,要求系统能够基于不同学生语言、文化及能力方面的差异,提供个性化支持。

教育公平性观测点的选取应遵循“接入保障—包容响应”的双层逻辑,将抽象的价值理念转化为可量化、可改进的技术与机制指标。在机会平等维度,观测点如硬件适配性、网络适应性和资源可及性直指物理与资源层面的准入公平。由于我国数字化基础设施在区域、城乡等方面存在显著差异<sup>[49]</sup>,必须通过终端兼容性、离线可用性、多语种资源覆盖率等指标,确保人工智能教育应用在低性能设备与不稳定网络环境中仍能提供核心功能,避免因技术门槛排除弱势群体学生。在包容多元维度,特殊群体学习支持度与学习路径定制化程度等观测点则用于评估系统对异质需求的响应能力。教育公平的本质并非绝对均等,而是通过个性化支持回应多元需求。因此,需借助特殊群体学习支持度、学习路径定制化程度等观测指标,衡量人工智能系统是否真正具备包容性设计能力。

##### (二)隐私安全性原则

隐私安全性旨在确保对学习者的个人信息、学业记录与行为数据的处理合法、正当与必要。在此过程中,不仅要规避数据再识别与泄露风险,还需通过安全防护措施与合规治理安排,对系统运行和外部攻击进行有效管控。教育数据应用场景高度多样化,从学生信息采集、数据标注、模型训练到结果输出与迭代优化,每一环节都存在数据泄露、滥用或篡改的风险。若隐私安全缺位,不仅会直接侵害学生权益,更将严重侵蚀师生对人工智能教育应用的信任基础。因此,隐私安全性应从隐私保护、数据安全与应用合规三个关键维度来构建其原则框架(见表3)。隐私保护聚焦数据脱敏和漏洞管理,为学习者信息安全提供基础屏障,数据安全旨在通过传输加密、入侵检测和访问控制等技术确保系统具备足够的防御与应对能力,应用合规主要指数据处理需符合相关法律法规与标准,并确保处理流程可审计、可追溯与可问责。

隐私安全性观测点的设计应以“匿名度”“稳健

表2

教育公平性评估要素

内容点	价值点	内涵释义	观测点
机会平等	技术普适、资源普惠	教育领域人工智能须适配不同硬件与网络条件,学习资源应涵盖多元文化、多语种环境以及多样化学科,以保障各类学生获得充分而公平的学习机会	硬件适配性
			网络适应性
			资源可及性
包容多元	差异包容、个性支持	教育领域人工智能应基于学生语言、文化及能力等差异,为其提供适配的教学内容与路径	特殊群体学习支持度
			学习路径定制化程度

性”和“合规性”为核心,系统覆盖全生命周期,确保教育领域人工智能在伦理与法律框架内可靠运行。在匿名度方面,观测点聚焦数据采集与预处理阶段可依据差分隐私参数等量化指标,评估系统对个人信息的去标识化的强度,从源头遏制再识别与隐私泄露风险。在稳健性方面,观测体系贯穿数据传输、存储与处理环节,从技术层面评估系统抗攻击能力,具体包括传输与静态加密强度、入侵检测响应时间等指标,确保数据机密性、完整性与可用性不受威胁。在合规性方面,观测点应侧重于制度治理与审计追溯,依托数据处理日志完整性、合规认证状态等证据,验证其是否符合《中华人民共和国个人信息保护法》等法律法规及相关标准。

### (三)算法透明性原则

算法透明性旨在确保教育领域人工智能在决策过程中运作机制、推理依据及用户交互过程具备可理解、可追溯与可预测的特性。教育领域人工智能在实现个性化推荐、智能测评或学习分析时,往往涉及多层算法流程与复杂模型结构。若算法与模型均处于黑箱状态,将使用户难以理解推荐或评分的依据,从而可能质疑系统的可信度与合理性,甚至排斥技术应用。因此,算法透明性应从适度开放、决策可信和交互可溯三个维度进行系统构建(见表4)。其中,适度开放强调对教师、管理者等利益相关者提供算法审查的条件,以支持模型审计与验证;决策可信要求系统以使用者可理解的方式呈现决策依据,提升决策可知性

与用户信任;交互可溯则支持使用者通过交互式操作,回溯系统的决策路径,进而对其反馈机制的鲁棒性与合理性进行验证。

算法透明性观测点的选取应基于以下核心原则:一是以用户为中心,观测点的设计紧紧围绕不同利益相关者(教师、学生、管理者)的认知需求与操作权限,确保提供的透明度与其责任和能力相匹配,避免产生信息过载或无效透明。二是技术可行与伦理合规平衡,观测点必须在技术上是可实现的,如通过“可解释人工智能技术”(Explainable AI, XAI)<sup>[50]</sup>实现局部解释,同时严格遵循伦理与法律边界。三是行为导向,观测点的价值最终应体现在能否促进行为改变上,即用户是否能利用提供的透明信息进行更有效的教学决策、提出更精准的反馈或增强对技术的批判性使用能力,从而将“透明”转化为实际的“赋能”效果。

### (四)发展导向性原则

发展导向性强调技术应用立足于学习者的长远与全面发展,充分发挥其对主体性、高阶能力及正确价值观的积极引领与赋能作用,是衡量教育领域人工智能价值深度的核心原则。当前,教育领域人工智能应用易陷入工具理性陷阱,过度关注短期分数与精准推送,可能导致学生被动依赖、思维僵化以及教学同质化问题出现,背离教育的育人初心。因此,发展导向性原则要求教育领域人工智能的设计与应用必须超越辅助角色,转向成为激发学习者内在动机、培育核心素养、塑造健全人格的“赋能者”。这一理念具体体

表3 隐私安全性评估要素

内容点	价值点	内涵释义	观测点
隐私保护	数据脱敏、漏洞管理	教育领域人工智能应降低个体数据被识别与滥用的风险,为学习者信息安全提供保障	匿名度
数据安全	传输加密、入侵检测、访问控制	教育领域人工智能应对网络攻击、数据篡改与未授权访问等行为具备足够的防御与应对能力	稳健性
应用合规	合法合规、标准适配、制度完备	教育领域人工智能的数据处理须符合相关法律法规与标准,并确保处理流程可审计、可追溯与可问责	合规性

表4 算法透明性评估要素

内容点	价值点	内涵释义	观测点
适度开放	模型可视、审计验证	教育领域人工智能的算法模型应对教师、管理者等利益相关者适度开放,允许其在合规范围内审查系统逻辑与性能	可访问性
决策可信	决策可知、信任提升	教育领域人工智能应让使用者了解模型如何对输入数据进行推理与决策,提升信任与使用意愿	可解释性
交互可溯	路径可溯、反馈可验	教育领域人工智能应支持使用者通过交互式操作,追溯系统的决策路径,并对其反馈机制的鲁棒性与合理性进行验证	可探索性

表5

发展导向性评估要素

内容点	价值点	内涵释义	观测点
人机协同	高效施教、有效学习	教育领域人工智能应与师生协同,辅助教学而非替代,明确人机分工以提高教学效率与质量	教学目标达成度 学习目标达成度
动机激发	自主赋能、内在驱动	教育领域人工智能应用须培养学习者自主性,激发内在动力	自主学习行为占比
价值引领	观念认同、行为践行	教育领域人工智能须在系统设计与交互中融入核心价值观、创新精神、协作意识等价值要素,并引导学习者在观念上认同、在行为上践行	认知与行为一致性

现在人机协同、动机激发与价值引领三个相互关联的维度上(见表5)。人机协同强调人工智能与师生形成优势互补,通过合理分工提升教学效率与质量,实现高效施教与有效学习的统一;动机激发注重培养学习者自主性,激发学习者的内在动力,促进其从被动接受向主动探索转变;价值引领则着眼于引导学生在观念上认同、在行为上践行核心价值观、创新精神和协作意识等价值要素,为其长远发展奠定基础。

发展导向性观测点的选取应基于系统性与可测性原则,聚焦“教学目标达成度”“学习目标达成度”“自主学习行为占比”及“认知与行为一致性”,以系统评估教育领域人工智能是否真正促进人的全面发展。其中,教学目标达成度与学习目标达成度旨在衡量人机协同下的教学效率与学习效果,保障人机合理分工;自主学习行为占比旨在评估系统激发学生内在动力、培养其自主规划与调节学习行为的成效,是实现“学会学习”的关键指标;认知与行为一致性旨在考查系统是否有效引导学生认同并践行核心价值观、创新

精神、协作意识等价值要素,彰显教育领域人工智能在促进学生全面发展方面的作用。

## 五、结束语

当前,人工智能技术正快速在教育领域迭代和扩张,一方面赋能了差异化教学和个性化学习,有力促进了大规模因材施教,另一方面也凸显出算法歧视、隐私泄露与价值导向失衡等伦理风险<sup>[1]</sup>。技术的迅猛演进不断打破既有教育格局,使“守正创新”成为迈向智慧教育的关键议题,也对教育领域人工智能的伦理框架提出了更高要求。本文在全面系统梳理人工智能伦理原则发展脉络与未来方向的基础上,结合教育领域的特定需求、发展路线、瓶颈问题与发展方向,提出教育领域人工智能伦理框架的构建原则,不仅为伦理风险防控提供价值与技术层面的双重路径,也为智能时代未来教育政策制定、技术标准建设与伦理评估奠定路向基准,最终实现人工智能赋能教育的向善发展与人文关怀。

## [参考文献]

- [1] 杨宗凯,王俊,吴砥,等. ChatGPT/生成式人工智能对教育的影响探析及应对策略[J]. 华东师范大学学报(教育科学版),2023,41(7):26-35.
- [2] 张绒. 生成式人工智能技术对教育领域的影响——关于 ChatGPT 的专访[J]. 电化教育研究,2023,44(2):5-14.
- [3] 吴砥,郭庆,郑旭东. 智能技术进步如何促进学生发展[J]. 教育研究,2024,45(1):121-132.
- [4] 黄荣怀,张国良,刘梦彧. 面向智慧教育的技术伦理取向与风险规约[J]. 现代教育技术,2024,34(2):13-22.
- [5] 祝智庭,韩中美,黄昌勤. 教育人工智能(eAI):人本人工智能的新范式[J]. 电化教育研究,2021,42(1):5-15.
- [6] 吴砥,李环,陈旭. 人工智能通用大模型教育应用影响探析[J]. 开放教育研究,2023,29(2):19-25,45.
- [7] 白钧溢. 全球人工智能伦理治理的文化逻辑与实践理路[J]. 科学技术哲学研究,2025,42(4):103-110.
- [8] TURING A. Intelligent machinery(1948)[M]//COPELAND B J. The essential Turing. Oxford: Oxford Univeristy Press,2004.
- [9] GOOD I J. Speculations concerning the first ultraintelligent machine[J]. Advances in computers, 1965, 6: 31-88.
- [10] MANER W. Starter kit in computer ethics[M]. West Virginia: Helvetia Press, 1980.
- [11] MOOR J H. What is computer ethics?[J]. Metaphilosophy, 1985,16(4):266-275.
- [12] ANDERSON R E, JOHNSON D G, GOTTERBARN D, et al. Using the new ACM code of ethics in decision making[J]. Communications of the ACM, 1993,36(2):98-107.
- [13] 孙焕林. 人工智能哲学问题座谈会在京举行[J]. 哲学研究,1981(8):80-67.
- [14] VERUGGIO G. The EURON roboethics roadmap [C]/2006 6th IEEE-RAS International Conference on Humanoid Robots. Genova:

- IEEE, 2006: 612-617.
- [15] SHARKEY N. The ethical frontiers of robotics[J]. *Science*, 2008, 322(5909): 1800-1801.
- [16] FLI. Asilomar AI principles[EB/OL]. (2017-08-11)[2025-03-27]. <https://futureoflife.org/open-letter/ai-principles/>.
- [17] European Commission. Ethics guidelines for trustworthy AI[EB/OL]. (2019-04-08)[2025-08-13]. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.
- [18] UNESCO. 人工智能伦理问题建议书[EB/OL]. (2021-11-23)[2025-08-13]. [https://unesdoc.unesco.org/ark:/48223/pf0000381137\\_chi](https://unesdoc.unesco.org/ark:/48223/pf0000381137_chi).
- [19] The White House. Guidance for regulation of artificial intelligence applications [EB/OL]. (2020-11-17)[2025-08-13]. <https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf>.
- [20] Google. AI at Google: our principles [EB/OL]. (2018-06-07)[2025-08-14]. <https://blog.google/technology/ai/ai-principles/>.
- [21] 微软亚洲研究院. 计算未来: 人工智能及其社会角色[M]. 北京: 北京大学出版社, 2018.
- [22] 中华人民共和国中央人民政府. 国务院关于印发新一代人工智能发展规划的通知 [EB/OL]. (2017-07-20)[2025-08-31]. [https://www.gov.cn/zhengce/content/2017-07/20/content\\_5211996.htm](https://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm).
- [23] 中华人民共和国科学技术部. 发展负责任的人工智能; 新一代人工智能治理原则发布[EB/OL]. (2019-06-17)[2025-08-31]. [https://www.most.gov.cn/kjbgz/201906/t20190617\\_147107.html](https://www.most.gov.cn/kjbgz/201906/t20190617_147107.html).
- [24] 中华人民共和国中央人民政府. 国家标准化管理委员会 中央网信办 国家发展改革委 科技部 工业和信息化部关于印发《国家新一代人工智能标准体系建设指南》的通知 [EB/OL]. (2020-07-27)[2025-08-31]. [https://www.gov.cn/zhengce/zhengceku/2020-08/09/content\\_5533454.htm](https://www.gov.cn/zhengce/zhengceku/2020-08/09/content_5533454.htm).
- [25] 中华人民共和国科学技术部. 《新一代人工智能伦理规范》发布[EB/OL]. (2021-09-26)[2025-08-31]. [https://www.most.gov.cn/kjbgz/202109/t20210926\\_177063.html?ref=salesforce-research](https://www.most.gov.cn/kjbgz/202109/t20210926_177063.html?ref=salesforce-research).
- [26] OECD. OECD Framework for the Classification of AI Systems [EB/OL]. (2022-02-22) [2025-08-31]. [https://www.oecd.org/en/publications/oecd-framework-for-the-classification-of-ai-systems\\_cb6d9eca-en.html](https://www.oecd.org/en/publications/oecd-framework-for-the-classification-of-ai-systems_cb6d9eca-en.html).
- [27] National Institute of Standards and Technology. Artificial intelligence risk management framework (AI RMF 1.0) [EB/OL]. (2023-01)[2025-09-08]. <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>.
- [28] GOV.UK. The Bletchley declaration [EB/OL]. (2023-11-02) [2025-03-27]. <https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023>.
- [29] European Parliament. EU AI act: first regulation on artificial intelligence [EB/OL]. (2025-02-19) [2025-03-27]. <https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>.
- [30] European Commission. Commission publishes guidelines on AI system definition to facilitate the first AI act's rules application [EB/OL]. (2025-02-18) [2025-09-09]. <https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-ai-system-definition-facilitate-first-ai-acts-rules-application>.
- [31] 中华人民共和国中央人民政府. 中共中央办公厅 国务院办公厅印发《关于加强科技伦理治理的意见》[EB/OL]. (2022-03-20) [2025-08-13]. [http://www.gov.cn/zhengce/2022-03/20/content\\_5680105.htm](http://www.gov.cn/zhengce/2022-03/20/content_5680105.htm).
- [32] 中国人工智能产业发展联盟. 人工智能伦理治理标准化指南[EB/OL]. [2025-08-13]. <https://www.aipubservice.com/airesource/fs/人工智能伦理治理标准化指南.pdf>.
- [33] 中华人民共和国科学技术部. 关于印发《科技伦理审查办法(试行)》的通知 [EB/OL]. (2023-09-07)[2025-11-08]. [https://www.most.gov.cn/xxgk/xinxifenlei/fdzdkgknr/fgzc/gfxwj/gfxwj2023/202310/t20231008\\_188309.html](https://www.most.gov.cn/xxgk/xinxifenlei/fdzdkgknr/fgzc/gfxwj/gfxwj2023/202310/t20231008_188309.html).
- [34] 中华人民共和国国家互联网信息办公室. 中国网络法治发展报告(2024年) [EB/OL]. (2025-06-06)[2025-09-08]. [https://www.cac.gov.cn/2025-06/06/c\\_1750919392401385.htm](https://www.cac.gov.cn/2025-06/06/c_1750919392401385.htm).
- [35] 中华人民共和国中央人民政府. 国务院办公厅关于印发《国务院2025年度立法工作计划》的通知 [EB/OL]. (2025-05-14) [2025-09-08]. [https://www.gov.cn/zhengce/content/202505/content\\_7023697.htm](https://www.gov.cn/zhengce/content/202505/content_7023697.htm).
- [36] HOLMES W, BEKTIK D, GENNARO M D, et al. Ethics in AI/ED: who cares? [C]//International Conference on Artificial Intelligence in Education. Cham: Springer, 2019.
- [37] UNESCO. Beijing consensus on artificial intelligence and education [EB/OL]. (2019-05-18)[2025-08-31]. <https://unesdoc.unesco.org/ark:/48223/pf0000368303>.

- [38] UNESCO. Generative AI and the future of education[EB/OL]. [2025-08-31]. <https://unesdoc.unesco.org/ark:/48223/pf0000385877>.
- [39] UNESCO. Guidance for generative AI in education and research [EB/OL]. (2023-09-07)[2025-08-31]. <https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research>.
- [40] European Commission. Guidelines on the scope of the obligations for providers of general-purpose AI models under the AI act [EB/OL]. (2025-07-18)[2025-09-11]. <https://digital-strategy.ec.europa.eu/en/library/guidelines-scope-obligations-providers-general-purpose-ai-models-under-ai-act>.
- [41] 中国教育学会. 中小生成式人工智能使用指南(2025年版)[EB/OL]. (2025-05-13)[2025-08-31]. <https://www.cse.edu.cn/index/detail.html?category=31&id=4242>.
- [42] 白钧溢,于伟. 超越“共识”:教育人工智能伦理原则构建的发展方向[J]. 中国电化教育,2023(6):9-17,24.
- [43] FLORIDI L. Translating principles into practices of digital ethics: five risks of being unethical [J]. Philosophy & technology, 2019, 32(2):185-193.
- [44] HOLMES W, PORAYSKA-POMSTA K, HOLSTEIN K, et al. Ethics of AI in education: towards a community-wide framework[J]. International journal of artificial intelligence in education, 2022, 32(3):504-526.
- [45] 中华人民共和国中央人民政府. 中共中央、国务院印发《中国教育现代化 2035》[EB/OL]. (2019-02-23)[2025-08-26]. [https://www.gov.cn/zhengce/2019-02/23/content\\_5367987.htm](https://www.gov.cn/zhengce/2019-02/23/content_5367987.htm).
- [46] 习近平. 高举中国特色社会主义伟大旗帜 为全面建设社会主义现代化国家而团结奋斗——在中国共产党第二十次全国代表大会上的报告[N]. 人民日报,2022-10-26(1).
- [47] 中国人大网. 中华人民共和国个人信息保护法 [EB/OL]. (2021-08-20)[2025-08-31]. [http://www.npc.gov.cn/npc/c2/c30834/202108/t20210820\\_313088.html](http://www.npc.gov.cn/npc/c2/c30834/202108/t20210820_313088.html).
- [48] NGUYEN A, NGO H N, HONG Y, et al. Ethical principles for artificial intelligence in education [J]. Education and information technologies, 2023, 28(4):4221-4241.
- [49] 郭绍青,华晓雨. 教育数字化转型助推城乡教育公平的路径研究[J]. 国家教育行政学院学报,2023(4):37-46,95.
- [50] GUNNING D. Explainable artificial intelligence (xai)[J]. Defense advanced research projects agency (DARPA), 2017, 2(2):1.
- [51] 吴砥, 吴河江. 通用大模型教育应用的潜在风险及其规避——基于技术伦理的视角 [J]. 华东师范大学学报(教育科学版), 2024, 42(8):64-75.

## Ethical Framework for Artificial Intelligence in Education: Evolutionary Trajectory and Construction Principles

WU Di<sup>1</sup>, GUI Xujun<sup>2</sup>, BAI Junyi<sup>3</sup>, CHEN Xu<sup>3</sup>

(1.Educational Informatization Strategy Research Base Ministry of Education(Central China), Central China Normal University, Wuhan Hubei 430079; 2.National Engineering Research Center for E-learning, Central China Normal University, Wuhan Hubei 430079; 3.Faculty of Artificial Intelligence in Education, Central China Normal University, Wuhan Hubei 430079)

[Abstract] Artificial intelligence, while empowering education, has also raised profound concerns regarding ethical risks such as fairness, privacy, and transparency. The ethical framework serves as a critical medium for carrying out ethical governance and addressing ethical risks, and its maturity depends on the continuous evolution and refinement of ethical principles. Reviewing the developmental trajectory, AI ethical principles have evolved through distinct stages, from conceptual vision, issue formation, and framework construction to operationalization exploration, demonstrating a trend toward translating abstract ethical values into standardized and actionable normative frameworks. Constrained by factors such as short development cycles and diverse stakeholders, the standardization and executable evolution of AI ethical

(下转第 58 页)

## The Application Potential of Artificial Intelligence Synthetic Data in Education —An Empirical Study Based on the IPA Model

WANG Youmei<sup>1</sup>, LI Xiaohan<sup>2</sup>, GUO Funing<sup>3</sup>, LIU Chenchen<sup>2</sup>

(1.Research Center for Artificial Intelligence in Vocational Education, Zhejiang Dongfang Polytechnic, Wenzhou Zhejiang 325035; 2.Research Center for Big Data and Smart Education, Wenzhou University, Wenzhou Zhejiang 325035; 3.Information service center, Wenzhou Business College, Wenzhou Zhejiang 325035)

**[Abstract]** With the rapid development of generative artificial intelligence, artificial intelligence synthetic data technology has become an important tool for promoting the progress of AI. Although artificial intelligence synthetic data has shown significant application potential in the field of education, related research is still in its infancy, and the discussion of its core value and potential problems is not yet sufficient. This study employs the IPA analytical framework, combined with expert questionnaire survey data, to systematically explore the application potential of artificial intelligence synthetic data in the field of education and empirically analyze its significance and specific manifestations. The results indicate that artificial intelligence synthetic data demonstrates high application value in privacy protection, teaching strategy optimization, and personalized learning. However, further improvements are needed in areas such as predicting health risks for teachers and students and evaluating educational policies. Future research should focus on technology optimization, expansion of application scenarios, deepening of empirical studies, and interdisciplinary cooperation. This will further enhance the application efficiency of artificial intelligence synthetic data in education, providing theoretical foundations and practical references for the sustainable application of artificial intelligence synthetic data in the field of education.

**[Keywords]** Generative AI; Artificial Intelligence Synthetic Data; IPA Analysis Model

---

(上接第 49 页)

principles in education face two major bottlenecks: overly broad elements and unclear boundaries. To address these challenges, this paper proposes that enhancing "alignment" and "implementability" should be established as the future direction for AI ethical principles in education, and based on this perspective, sets forth principles for constructing an ethical framework, thereby providing theoretical and methodological support for risk governance and subsequent ethical evaluation in application.

**[Keywords]** Educational Digitalization; Ethical Framework; Evolutionary Trajectory; Construction Principles