

跨学科视角下人工智能 伦理教育的本土化实践与发展性研究*

杨常唯

浙江东方职业技术学院, 温州 325025

摘要 随着人工智能技术在教育领域的不断深度融入, 伦理教育成为培养数字公民过程中不容忽视的重要环节。论文基于对伦理教育本土化现状及面临挑战的分析, 围绕本土化课程体系构建、创新课程动态评价、动态发展性机制建设、本土化实践及其路径以及本土化改革实践成效等伦理教育的本土化实践与探索, 从跨学科角度, 探讨了人工智能伦理教育的差异化实现途径与发展性运行机制。提出了算法时代的技术发展要保持人类的自主性、目的性、伦理性及创造性的主张, 为构建中国特色人工智能伦理教育体系提供了理论与实践两方面的参考。

关键字 跨学科视角; 人工智能; 伦理教育; 本土化实践; 发展性研究

A Study on the Localization and Development of Artificial Intelligence Ethics Education from a Cross-Disciplinary Perspective*

Changwei Yang

Zhejiang Dongfang Polytechnic,
Wenzhou, 325025, China;
y591246235@163.com

Abstract—As artificial intelligence technology becomes increasingly integrated into the education sector, ethical education has emerged as a crucial component that cannot be overlooked in fostering digital citizenship. Based on an analysis of the current state and challenges facing the localization of ethical education, this paper examines relevant issues. Focusing on the development of localized curriculum systems, innovative dynamic course evaluation mechanisms, and the establishment of dynamic developmental frameworks, The localization practices and their implementation pathways, along with the outcomes of localized reform efforts, represent significant endeavors in exploring ethical education within a localized context. From an interdisciplinary perspective, this study explores differentiated approaches and developmental operational mechanisms for ethical education in artificial intelligence. It advocates that technological development in the algorithmic era should maintain human autonomy, purposefulness, ethics, and creativity. It provides theoretical and practical references for establishing an artificial intelligence ethics education system with China's distinctive characteristics.

Keywords—Interdisciplinary perspectives, Artificial intelligence, Ethics education, Localization Practice, Developmental research

1 引言

目前, 在人工智能融入教育领域并不断推广应用背景下, 伦理教育滞后会给社会发展带来很大隐患。其“工具”使用需规避的“风险”矛盾主要如下:

技术普适性与文化特殊性之间存在着工具理性与

价值理性的冲突; 学科壁垒阻碍系统性伦理教育的实施; 城乡差异导致伦理教育资源分布失衡。

“跨学科+本土化”AI 伦理教育的发展性研究, 通过整合技术教育、伦理学、文化学等跨界理论, 推动中国特色人工智能伦理教育及应用实践, 力求完成以下转变:

课程体系从“碎片化”向“系统化”变革; 伦理教育从“工具理性”向“价值理性”转型; 教师角色

***基金资助:** 本文得到 2025 年浙江省高等学校(浙江东方职业技术学院)国内访问工程师项目“AI 赋能校企协同共创电商人才培养新模式——基于珍岛信息技术(上海)股份有限公司调研”(FG2025248)的资助与支持。

从“技术使用者”向“伦理引导者”跃升。奠定人机和和谐共生的数字社会伦理基石。

“嵌入式”教育主张打破学科边界、重构标准，将伦理教育融入跨学科类型课程[1]。AI 伦理教育本土化与现状对比（见表 1）。

表1 AI伦理教育本土化需求与现状对比

维度	现实困境	本土化需求
文化适配	西方伦理框架主导，与传统价值观脱节	融合“中庸之道”“天人合一”等传统智慧
教育实践	部分教师虽然接受过系统 AI 培训，但出现了教学内容流于形式、缺乏实质深度的空心化课程	构建“三主线五模块”课程体系
技术风险	医疗 AI 误诊责任模糊，人脸识别误差率较高	开发伦理沙盘、AR 决策模拟器等治理工具
资源分配	落后地区学校硬件缺失，区域差异显著	借鉴“1+8”经验构建区域协同机制与虚拟实验室

人工智能正在深刻重塑社会生活，数据隐私、数据安全、算法偏见、责任归属及医患关系异化等伦理问题频发，伦理风险日益凸显，亟须构筑伦理教育防线[2]，提升对人工智能伦理风险教育的认知，促进伦理价值认同[3]。

2022 年 2 月，联合国教科文组织发布了《K-12 人工智能课程》报告，强调了伦理教育在课程建设中的基础地位，将算法公平、数据隐私纳入核心素养模块。

在人工智能教育研究多元化发展进程中，国际普遍关注的 AI 伦理教育焦点问题主要可以归纳为以下五项：第一，数据安全与信息保护；第二，透明与公平正义；第三，师生主体性发展；第四，诚信与违法犯罪；第五，实践应用无效化[4]。

目前，国内人工智能伦理教育正处于政策驱动向实践落地转型阶段。国务院发布的《新一代人工智能发展规划》《中国教育现代化 2035》等文件提出“支持可信、安全和负责任地使用人工智能技术，研究建立人工智能技术的伦理道德和治理框架”[5]，要求“开展人工智能伦理教育”，得到各地积极响应。

首先，广东省率先推出师生人工智能素养框架。广州市开发全国首套省级审定人工智能教材，课程嵌入文明、诚信、安全等伦理板块，以学生公约的形式规范 AI 使用。实现城乡“1+8”资源协同共建虚拟实验室，以解决乡村硬件缺失问题。其次，上海通过数据流向绘制等场景化课程，提升伦理认知。再次，南

京与广州将非遗技艺、民乐等与机器学习相结合，不断激活文化基因。

已初具特色的中国 AI 伦理教育本土化实践需要在持续落实中不断完善。现有研究的学科分布与局限分析（见表 2）。

表2 现有研究的学科分布与局限分析

研究领域	代表成果	主要贡献	本土化局限
教育技术学	联合国教科文 AI 课程框架	建立全球伦理教育标准	忽视文化多样性
文化人类学	广州“玩转木雕”项目	链接非遗与 AI 技术	未形成理论体系
教育公平研究	虚拟实验室城乡共享	缩小数字差距	覆盖率不足的局限
教师教育学	教师 AI 素养雷达图	精准诊断培训需求	未解决跨学科瓶颈

2 本土化现状及面临的挑战

2.1 跨学科整合及现存问题

教育对象主体性异化极易引发“信念弱化，价值信仰失序；思维惰化，创新精神遮蔽；感知虚化，人文情感稀释；责任意识淡化，道德行为失范”等伦理风险[6]。

现存问题如下：

课程体系割裂：现有实践多依赖校本课程，缺乏跨学科整合设计，易导致内容重复或断层。

师资能力错配：教师缺乏跨学科培训，难以驾驭技术—伦理相互融合的教学。

评价机制缺失：学校伦理判断能力评价指标有待构建与完善，教学效果难以量化。

2.2 AI 伦理教育不足及应对

跨学科整合成为破解伦理教育碎片化的关键路径，而本土化实践需要适应区域差异与文化适配性的需求。

（1）亟待破解的现实问题

首先，本土化整合框架缺失，需要建立兼顾传统文化与现代伦理的课程内核。

其次，区域适配方案不足，缺乏针对西部资源薄弱校的轻量化教学模式。

再次，发展性机制空白，需构建响应技术演进的动态内容更新路径。

（2）创新性应对解决策略

进行技术伦理一体化设计。微软亚洲研究院价值观罗盘 (Value Compass) 研究项目, 解决了价值观念定义、评测和对齐问题[7]。

推进动态治理与法律革新。2021 年 11 月, 联合国第 41 届大会通过了首份 AI 伦理问题全球性协议《人工智能伦理问题建议书》[8]。

开发分级培训体系。重视不同类别教师能力的完善与提升, 理工 (伦理培训)、人文社科 (技术培训); 构建“虚拟实验平台”, 支持伦理情境模拟; 关注区域差异, 建立“多元评价指标”体系。

3 伦理教育的本土化实践

平衡技术发展与社会伦理关系, 倡导责任意识与批判思维, 为培育既具有技术能力, 又具备伦理素养的数字时代责任公民, 不断探索以下支撑框架体系:

(1) 教育执行系统

优化组合: 技术专家+伦理学者组成的跨学科培训团队, 可采用算法工程师与哲学教授联合开设“技术哲学”等课程; 资源开发: 建立深度伪造技术鉴别伦理案例库。

(2) 社企协同系统

社企共建: 科技企业开放伦理审查流程, 对自动驾驶路测数据等进行脱敏展示; 社区实践: 以“AI 伦理角”, 展示平衡安防监控与隐私保护实施方案。

(3) 政策保障系统

法规衔接: 将《中华人民共和国个人信息保护法》条款转化为教学情境; 资金机制: 设立 AI 伦理教育专项基金, 支持校园伦理实验室建设。

3.1 本土化课程体系构建

(1) “三条主线”课程结构

第一, 文化基因与技术应用融合。将“技术向善”理念注入算法, 引领伦理教育融入 AI, 构建以能力创新为导向的新型伦理教育机制[9], 基于智能场景创新课程思政教育生态环境。开发《AI 课程思政》读本, 用传统智慧解析算法伦理。

第二, 将社会实景带入伦理分析。选取社区团购、算法定价等“本土化场景”, 以实际问题为驱动, 设计完成“算法调研”“数据治理”等实际伦理分析项目。

第三, 技术伦理对接社会实践。基于医学健康等相关课程解析 CRISPR 基因编辑技术伦理案例, 在法律课程中, 组建自动驾驶责任归属模拟法庭等, 开展跨学科典型伦理主题探究 (见表 3)

表3 跨学科典型伦理主题探究列表

主题	学科整合	产出形式
外卖骑手系统优化	数学 (路径算法)+ 社会学 (劳工权益)	骑手保障提案
医疗 AI 诊断责任界定	医学 (诊断逻辑)+ 法学 (医疗事故法规)	责任归属流程图

(2) “五模块”任务结构

其一, 信息技术+法律数据隐私与所有权模块。

其二, 数学+社会学算法公平性验证模块。

其三, 哲学+心理学自主决策责任归属模块。

其四, 伦理学+设计人机关系边界探索模块。

其五, 传统文化+现代伦理文化价值观嵌入模块。

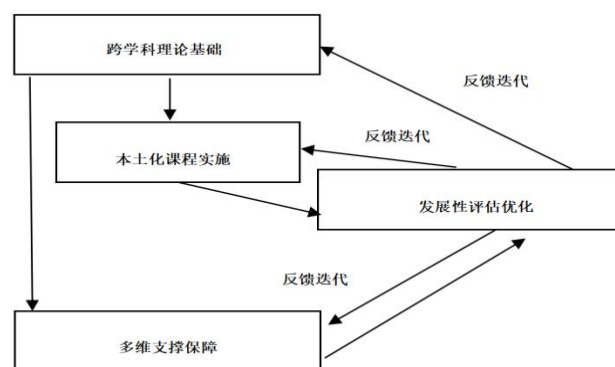
开发融数据追踪技术、区域差异社会分析等“一体化”跨学科各模块项目任务。

(3) 设计开发 AI 伦理教学资源

不断开发完善包含伦理决策沙盘、虚拟仿真实验等课程工具包。建立伦理“活页课程”机制, 力求每学期本土案例有更新。

“基础—实践—保障—优化”四维动态资源 (见图 1)

图1 “基础—实践—保障—优化”四维动态资源体系



3.2 创新课程动态评价

AI 伦理教育评估是动态演变的过程, 其流程也随着教育理念与人工智能技术的协同发展而趋于完善。[10]。遵守以人为本的价值导向、安全可控和隐私保护的准入条件, 贯彻可持续发展的应用准则、秉承公平全纳和相称无害的教育宗旨、落实透明可解释的问责监督保障机制[11]。

(1) 创新多维共评方式

多维共评占比分别为: 知识掌握 (30%)、伦理判断 (40%)、实践能力 (20%)、社会影响 (10%)。

(2) 建立伦理行为档案

记录学生在真实场景下的自主选择。利用自然语言处理技术，分析学生伦理教育现状并生成报告。

(3) 创新 AI 教学工具

开发辅助评价系统,集成 AR 伦理决策模拟器呈现校园道德困境场景,运用伦理阈值测试沙盒进行量化评估。

(4) 获社会影响力积分

建立社会影响力积分体系,将伦理学习转化为社会干预行动。推动企业修改用户协议,明确积分可用于兑换及参评优秀等级。

发展性评估优化: 动态闭环机制 (见图 2)

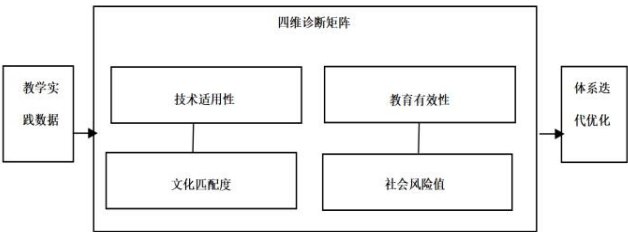


图 2 发展性评估优化: 动态闭环机制

3. 3 动态发展性机制建设

(1) 政校协同治理机制

借鉴广州“1+8”人工智能跨学科教育经验,激活课堂教学新活力,运用 AI 技术工具解决问题、创新实

践经验[12], 升级构建高校模式。高校联合建设联邦学习+区块链存证虚拟实验室。

社区与高校合作完成“谣言算法”等现实任务,创新区校共治模式。根据多方共守“契约”, 签订“AI 代写边界”等协议, 共建伦理教育工作坊。编写《高校 AI 伦理教学活页指要》, 纳入“数据治理”“社区团购算法”等本土案例, 完善本土化标准建设, 跨学科 AI 伦理课程发展性框架 (见表 4)。

表4 跨学科AI伦理课程发展性框架

学科	核心模块	跨学科融合案例	评价方式
自然科学	提升技术社会影响	数据考察项目	AI 驾驶权责等讨论
人文社科	深度理解算法逻辑	人脸识别误差实验+公平性等报告	软件系统优化提案
综合学科	提高伦理建设能力	伦理数据调研	社会影响力积分评估

(2) “追踪—更新—迭代”机制

技术追踪模块: 监控由人工智能新技术引起的伦理纠纷。

案例更新机制: 适当增补更新具有时代感的教学案例。

资源迭代路径: 根据教学反馈结果, 定期优化内容。营造蕴含伦理教育的教学生态环境体系, 建立多边合作协同参与的综合治理机制[13]AI 伦理教育跨学科发展性融合矩阵 (见图 3)

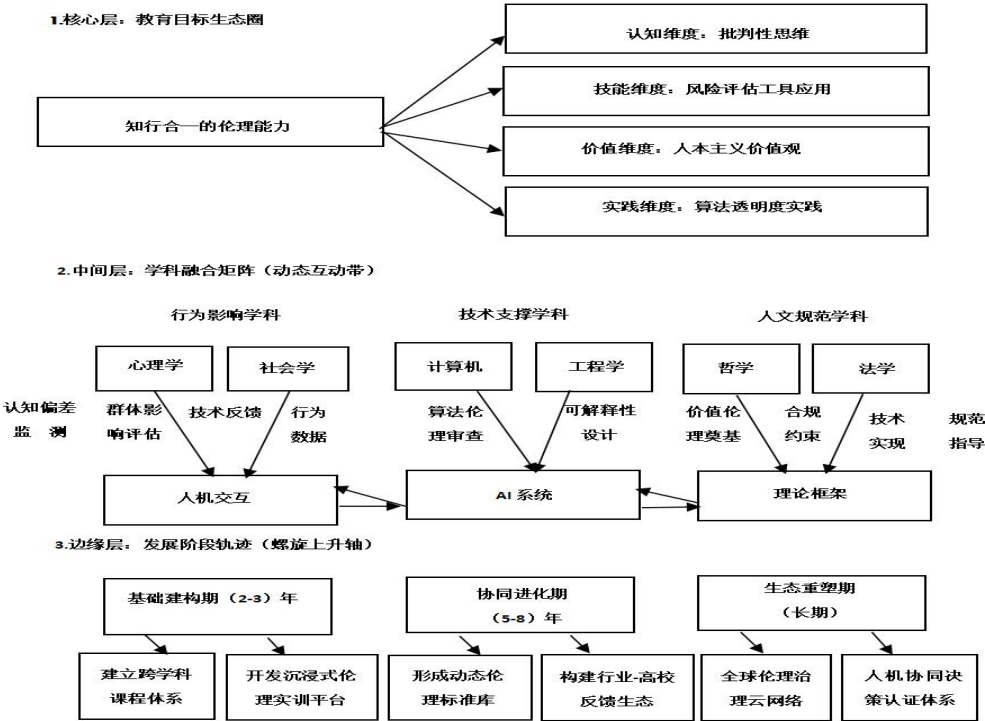


图 3 AI 伦理教育跨学科发展性融合矩阵（动态互动带）

3.4 本土化实践及其路径

（1）问题驱动型

围绕智慧医疗、社会信用等本土场景，导入伦理设计案例。运用虚实融合的“元宇宙场景”，实现计算机科学与各类学科跨界融合。

（2）区域适配型

人工智能赋能跨学科教育研究与实践，创新未来人才培养范式，助力区域教育高质量发展[14]。地区资源充裕的高校设计实体机器人伦理场景实验，为薄弱地区高校提供轻量化虚拟实验平台，降低硬件依赖，开发手机端虚拟实验，把实验室装进口袋。

（3）文化回应型

倡导文化回应式教学，将“中西方先进文化”融入算法公平讨论，对比差异优选中西伦理观精华，跨学科整合伦理实践框架（见表 5）。

表 5 跨学科整合伦理实践路径

整合维度	学科载体	核心教学内容	教学策略
技术认知	信息技术课程	算法原理、数据隐私、系统局限性等	编程实践、系统仿真、漏洞分析
价值判断	道德与法治	公平正义、责任归属、人权保护等	伦理两难讨论、角色扮演、公约制定
行为养成	心理学课程	技术依赖预防、自主决策训练等	情境模拟、自我反思日记、行为合约
社会语境	社会学专题	技术影响、法律法规、文化差异等	社区调研、跨文化比较、政策辩论

（4）提炼跨学科整合最佳实践路径

系统梳理国内外人工智能伦理教育的学科整合模式，为本土化提供理论支持。

（5）“虚实结合”“动态响应”机制

开发“虚实结合”的教学载体。设计具有隐私泄露警示等低成本的伦理实验器具；升级人工智能教学平台的 3D 仿真功能，强化生成式 AI 伦理模块。

建立“动态响应”式技术监测机制。基于预警信号，动态生成：原理性基础级、新风险应对预警级、未来场景预判级的三级响应机制，确保技术教育与风险防控同步。

3.5 本土化改革实践成效

伦理教育属于思想政治教育的范畴，旨在培育良好的价值观[15]。树立责任意识、社会担当和家国情怀，构建跨学科视角下人工智能伦理教育本土化思路，满足当前社会人才培养需求[16]。2025 年（下半年）高校实验班学生（专本各占 50%）伦理教育本土化实践课程改革前后核心能力提升效果对比（见表 6）。

表 6 教学模式实施前后学生核心能力提升对比

能力维度	评估指标	期初表现	期末表现	提升幅度（%）
文化基因与技术应用融合力	设计方案合格率	61.4%	91.6%	+49.2%
社会实景带入伦理分析力	逻辑链条完整性比例	44.9%	86.3%	+92.2%
技术伦理对接社会实践力	创新方案评分（5 分制）	1.57	2.43	+54.8%

4 结论

AI 伦理教育能够不断内化人工智能从业者、教育者的责任意识，有助于提升其智能伦理素养[17]。中国特色 AI 伦理教育的本质要求是培育兼具科技能力与伦理素质的数字公民。跨学科整合传统文化、社会实景、技术治理，让学生与人工智能形成和谐共处、相互调节的道德共同体[18]。传统文化与 AI 伦理对话，将传统“慎独”精神与隐私保护结合，推动人机协同、算法透明，克服技术理性与价值理性的分歧。以缓解技术应用中的不平等风险[19]。遵循技术调节、数字向善、责任分布、完善制度的伦理实践理路，更好地实现教育数字化转型[20]。

算法时代，技术发展的终极目标是保持人类的自主性、目的性和伦理性及创造性。人工智能伦理教育的本土化实践与发展，已经远超技术适配的范畴，成为呼应全球 AI 发展的制度性中国方案。

参考文献

- [1] 白钧溢,于伟. 人工智能伦理教育的三重境界[J]. 中国电化教育, 2025, (03): 11-19.
- [2] 高玉华,王明阳. 人工智能医疗应用中的伦理挑战与医学人文教育对策[J]. 中国医学人文, 2025, 11 (06): 45-48.
- [3] 陈翠荣,崔红岩. 博弈论视角下教育人工智能伦理风险治理困境及化解策略研究[J]. 中国地质大学学报(社会科学版), 2025, 25 (03): 131-144. DOI:10. 16493/j. cnki. 42-1627/c. 2025. 03. 001.
- [4] 郭庆,吴砥. 国际视野下人工智能教育应用伦理风险与治理策略[J]. 比较教育研究, 2025, 47 (01): 56-64. DOI:10. 20013/j. cnki. ICE. 2025. 01. 06.

- [5] 郝成. 上海自贸区将建立人工智能技术的伦理道德和治理框架[EB/OL]. 中国经营报, (2023-12-08) [2025-07-14]. <https://finance.eastmoney.com/a/202312082928409241.html>
- [6] 闫瑞峰. 生成式人工智能赋能思想政治教育的伦理风险与化解路径——基于教育对象主体性视角的考察[J]. 思想教育研究, 2025, (06):34-42.
- [7] 微软亚洲研究院. 价值观罗盘: 如何让大模型与人类价值观对齐?[EB/OL]. 微软研究论坛, (2024-12-10) [2025-07-14]. <https://www.microsoft.com/en-us/research/articles/value-compass/>
- [8] UNESCO. Recommendation on the Ethics of Artificial Intelligence[EB/OL]. (2021-11-24) [2026-06-28]. https://unesdoc.unesco.org/ark:/48223/pf0000380455_chi.
- [9] 闫瑞峰. 生成式人工智能赋能思想政治教育的伦理风险与化解路径——基于教育对象主体性视角的考察[J]. 思想教育研究, 2025, (06):34-42.
- [10] 王佑镁, 倚杨莹, 柳晨晨. 基于风险矩阵的教育人工智能应用伦理风险评估[J]. 现代远程教育研究, 2024, 36(06):3-10.
- [11] 杨俊锋, 褚娟, 张斌贤. 人工智能教育应用伦理规范指标构建研究[J]. 电化教育研究, 2024, 45(10):19-27. DOI:10.13811/j.cnki.eer.2024.10.003
- [12] 林欣潼. 广州共研人工智能赋能探索跨学科新路径以1+8模式普及人工智能通识教育[N]. 广州日报, 2025-05-15.
- [13] 曾光, 宋以华, 黎新华. 职业教育数字化转型中人工智能应用的伦理风险研究[J]. 教育与职业, 2025, (02):13-21. DOI:10.13615/j.cnki.1004-3985.2025.02.012.
- [14] 王悦, 董刚, 黄海旻. 南京探索人工智能教育新路径[N]. 中国日报, 2025-05-23
- [15] 何利, 易芝. 工科背景下工程伦理意识培养——《网络规划与设计》课程思政探索与实践[J]. 计算机技术与教育学报, 2022, 10(04):61-65.
- [16] 曹莹, 马宇翔等. 新工科背景下工程伦理教育的创新与实践[J]. 计算机技术与教育学报, 2025, 13(02):90-96.
- [17] 谢娟. 人工智能与教育融合创新何以“伦理先行”——兼论生成式人工智能教育应用的伦理路径[J]. 现代远程教育研究, 2024, 36(06):11-19.
- [18] 孙田琳子, 金约楠, 寇笑迪. 跨学科视域下人工智能伦理教育重构: 知识观、学生观与教学观[J]. 中国电化教育, 2024, (04):45-51.
- [19] 袁婧, 吴飞. 人工智能时代知识生产逻辑的转向与教育应对[J]. 中国远程教育, 2025, 45(07):20-34. DOI:10.13541/j.cnki.chinade.2025.07.008.
- [20] 王斌伟, 付圣莹. 人工智能教育应用中的伦理风险及其应对[J]. 学术研究, 2025, (04):1-9.