

# 趋利避害：生成式人工智能嵌入高校思政教育的伦理风险与规制路径

杨婉婷<sup>1</sup>, 徐绍华<sup>2\*</sup>

<sup>1</sup>昆明理工大学马克思主义学院, 云南 昆明

<sup>2</sup>昆明理工大学学报编辑部, 云南 昆明

收稿日期: 2025年10月21日; 录用日期: 2025年12月10日; 发布日期: 2025年12月22日

---

## 摘要

生成式人工智能技术的纵深应用正加速重构高校思政教育生态, 凭借其跨模态学习能力、可扩展模型和强化学习机制等技术优势, 能够处理海量数据并生成模仿人类语言的交流内容, 极大地推动了高校思政教育的数字化转型。但因其技术内生性、工具功能性与规制滞后性等多重影响, 导致该技术的实践应用也面临着伦理关系异化、伦理价值偏差、伦理规范缺失以及伦理行为失当等潜在风险。为有效规制其伦理风险, 应基于趋利避害原则, 构建教育内部治理、平台伦理规范、制度协同防治的规制体系, 形成“人 - 技 - 制”综合治理的三位一体格局, 使生成式人工智能更好地为高校思政教育的高质量发展赋能增效, 并有效规避其潜在的伦理风险, 在动态平衡中解决好人的价值属性与技术的工具属性之间的矛盾冲突, 真正实现技术服务教育、技术促进育人的价值功能。

## 关键词

生成式人工智能, 思想政治教育, 伦理风险, 规制路径

---

# Seeking Benefits and Avoiding Harms: Ethical Risks and Regulatory Pathways of Generative AI Integration into Ideological and Political Education in Higher Education Institutions

Wanting Yang<sup>1</sup>, Shaohua Xu<sup>2\*</sup>

<sup>1</sup>School of Marxism, Kunming University of Science and Technology, Kunming Yunnan

---

\*通讯作者。

<sup>2</sup>Editorial Office of Journal of Kunming University of Science and Technology, Kunming Yunnan

Received: October 21, 2025; accepted: December 10, 2025; published: December 22, 2025

## Abstract

The in-depth application of generative AI technology is accelerating the restructuring of the ideological and political education ecosystem in higher education institutions. Leveraging its technical advantages—including cross-modal learning capabilities, scalable models, and reinforcement learning mechanisms—this technology can process massive datasets and generate human-like language content, significantly advancing the digital transformation of ideological and political education in universities. However, due to multiple factors including its technological endogeneity, tool-based functionality, and regulatory lag, the practical application of this technology also faces potential risks such as ethical relationship alienation, ethical value bias, lack of ethical norms, and inappropriate ethical behavior. To effectively regulate these ethical risks, a regulatory framework should be established based on the principle of maximizing benefits while minimizing harm. This framework should integrate internal educational governance, platform ethical standards, and institutional collaborative prevention, forming a trinity of “human-technology-system” comprehensive governance. Such an approach will enable generative AI to better empower and enhance the high-quality development of ideological and political education in higher education institutions while effectively mitigating potential ethical risks. It will resolve the inherent tension between the human value attributes and the technological tool attributes in a dynamic equilibrium, truly realizing the value function of technology serving education and technology promoting nurturing.

## Keywords

Generative Artificial Intelligence, Ideological and Political Education, Ethical Risks, Regulatory Pathways

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

生成式人工智能技术的加速演进，推动人类社会从“工具辅助”向“智能共生”深层转型。以ChatGPT、DeepSeek等为代表的生成式人工智能，依托其强大的自然语言理解与内容生成能力，已在教育、医疗等多个领域引发结构性变革。在高等教育领域的应用尤为广泛，正在重塑教学内容、学习方式与师生关系。根据《2023年教育中的人工智能统计数据》显示：约86%的教师认为含人工智能在内的技术应是课堂体验的重要组成部分[1]，可见生成式人工智能正在成为推动教育现代化的重要变量。而高校思政教育因其意识形态属性与技术应用的敏感性，已成为技术赋能与伦理争议并存的焦点场域，还衍生出价值偏移、主体消解、责任边界模糊等伦理风险。因此，通过前瞻性地阐释生成式人工智能嵌入高校思政教育所引发的伦理风险，揭示其生成逻辑与规制路径，具有推动高校思政教育数字化转型的重要意义。

## 2. 伦理异化：生成式人工智能嵌入思想政治教育的风险表征

伦理风险本质上是指伦理、道德层面的风险及价值选择的优先性问题[2]。生成式人工智能嵌入高校

思政教育引发的伦理风险，是指该技术在重塑高等教育实践过程中因技术理性与道德伦理的冲突，而对“人-机”“人-人”“人-社会”“人-自我”等伦理关系产生的多维度负面效应。尽管其通过创新教育形式、优化教学方法显著提升了高校思政教育的实效性，但科学技术有时表现为“异己的、敌对的和统治的权力”<sup>[3]</sup>，即在赋能教育现代化的同时，易引发伦理关系异化、伦理价值偏差、伦理关系失序以及伦理行为失当等复合风险。这些风险不仅挑战高校思政教育立德树人的根本任务，更可能侵蚀高校意识形态安全的价值根基。

## 2.1. 伦理关系异化：师生身份结构的去主体化趋势

高校思政教育以立德树人为根本任务，其本质是通过教师的引导、对话与情感感染，实现学生的价值认同与道德自觉。然而，随着生成式人工智能的深度介入，高等教育关系正由传统的“人对人”互动转向“人-机-人”的多主体动态协同结构<sup>[4]</sup>，人的主体性则被技术逻辑所遮蔽。

### 2.1.1. 教师的主导性逐渐削弱

生成式人工智能的自动化反馈、个性化推送与情境模拟能力使其参与到高校思政教学内容的定制、作业批改及答疑等环节。学生通过生成式人工智能获取“即时知识”与“定向资源”，逐渐视其为“智能教师”，教师角色则从“价值引领者”降格为“技术使用者”。此身份的转变不仅冲击了高校教师的职业认同与“教师即权威”的伦理基础，也弱化了其在价值传递中的核心地位。

### 2.1.2. 学生的学习路径趋于被动化与技术依赖化

生成式人工智能的算法偏好与行为数据驱动内容筛选与推送，虽强化了知识获取的便利性，却压缩了学生自主探索与批判性思维的空间。思政教育中原本通过师生间的互动激发价值辨析的环节，被“输入-输出”式的单向知识灌输所替代，教学人文性也随之削弱。长期依赖算法推送，学生易丧失多元化视角与独立判断能力，甚至“被迫”接受算法内嵌的价值立场。

## 2.2. 伦理价值偏差：算法逻辑对意识形态导向的扭曲

作为意识形态工作的有机组成部分，高校思政教育的核心任务是引导学生形成对国家制度、社会价值与发展道路的深度认同。而生成式人工智能在高等教育中的算法推荐和数据训练体系却可能带来意识形态的隐性偏移。

### 2.2.1. 生成式人工智能数据训练导致认知混淆与价值偏移

当前主流生成模型广泛采用以英文语料为主的开源数据集，其知识基础、话语范式与价值观念深受西方自由主义思想影响。在缺乏本土化语义调控机制的前提下，面对“民主”“自由”“公正”等关键词时，生成式人工智能常输出“普世主义”的话语框架式解释，忽略了中国式现代化的政治背景与历史实践。这种隐性输出虽非主观偏向，却会误导学生的价值判断，形成对社会主义核心价值观的认知混淆与价值偏移。

### 2.2.2. 生成式人工智能的信息偏好依赖弱化独立自主的实践抉择

生成式人工智能遵循“用户兴趣最大化”的原则，会根据用户使用情况生成信息回路。当系统根据学生历史提问、偏好行为进行算法塑形时，易形成封闭性极强的“信息茧房”，不断强化单一价值取向，将学生锁定在特定知识结构中，从而削弱其跨文化视角与批判性思维的生成能力<sup>[5]</sup>。如果长期依赖此类推荐机制，学生将难以跳出技术设定的“价值回音室”，逐渐失去真伪辨析与构建独立意识形态的实践能力。

### 2.3. 伦理关系失序：教育权责体系的结构性瓦解

伦理关系失序主要体现为生成式人工智能嵌入高校思政教育后，教育系统中原有的角色分工、行为边界与权责结构逐渐模糊，传统的教育伦理规范体系遭遇系统性冲击。当前，生成式人工智能教育应用的伦理治理尚处于初始阶段，主体责任、内容审查、后果追责等方面机制不全，极易引发伦理失序[6]。若不加干预，生成式人工智能介入高校思政教育将逐步重塑教育秩序的失衡状态。

#### 2.3.1. 师生的紧密接触和互动程度骤降

随着生成式人工智能承担了大量教学功能，如内容生成、答疑服务，学生对教师的依赖程度不断下降，教师对学生个体差异与成长需求的关注也随之淡化。教育过程日益呈现出“去人化”的倾向，师生间情感联系与价值共鸣的空间被压缩，高校思政教育赖以展开的“人际信任”基础受到侵蚀，教育伦理关系失去应有的情感和人格支撑。

#### 2.3.2. 教育责任主体边界的模糊性愈发凸显

当学生在生成式人工智能的辅助下接触到错误信息或受到不当引导时，责任主体的界定面临两难：应由教师承担教学监督责任、平台方对算法风险负责？还是学生自身承担选择行为的后果？由于生成式人工智能在现有的教育法律体系中尚未得到明确界定，其在教学过程中的地位处于“非人、非物”的模糊状态，由此造成伦理责任的割裂与责任追溯的困难。

#### 2.3.3. 技术系统与教育治理体系之间存在管理错位

多数高校将AI教学平台划归为技术部门管理，重点关注系统运行与技术服务，而教育主管部门对其内容来源、数据处理逻辑及输出机制缺乏实际的干预权。在此治理结构下，“技术逻辑”往往主导了教育流程安排，“教育逻辑”则被边缘化，造成教育目标与技术功能之间的系统性张力失衡。这种治理权与使用权的错配，进一步导致教育管理层难以对生成式人工智能介入后的教学行为进行有效调控与伦理监督。

### 2.4. 伦理行为失当：责任虚置与教育依赖心理的生成

在生成式人工智能深度赋能高等教育实践的背景下，部分师生逐渐表现出责任意识淡化与行为依赖倾向，并导致教育伦理行为滑坡现象频发。而人工智能教育场域中日益突出的“责任空窗期”现象[4]，在缺乏清晰责任划分和行为规范的情况下，伦理失当将演变为系统性风险，严重影响教育公平、教学质量与育人目标的实现。

#### 2.4.1. 教师层面：主体地位的责任弱化

受限于教学压力和技术素养差异，部分教师逐渐习惯并依赖生成式人工智能辅助完成备课、命题、批改等教学任务，将其由“教学辅助”演变为“代替教学”。于是，各类教育行为“外包化”现象随之产生，教师对教学内容和教学过程不再具有价值把关能力，其职业责任感与育人使命意识也随之弱化。

#### 2.4.2. 学生层面：“智能代笔”的隐性载体

部分学生借助生成式人工智能完成论文写作、作业提交乃至考试答题，形成“无需努力即可得分”的依赖心理，模糊了学术诚信的边界与学习成果的真实性。这种行为不仅削弱了学生思辨能力与道德判断力，使其对价值观的理解浮于表面，也容易导致其对社会现象的分析更加倾向于工具主义依赖。

#### 2.4.3. 管理层面：制度防线的空缺隐患

生成式人工智能平台普遍具备开放性与匿名性特征，若缺乏有效审查，一些极端言论、虚假内容会

以“生成结果”的形式出现在学生面前，造成其对国家制度、社会现象的误解甚至不信任。这种风险看似源于技术的中立性，实则是平台监管不足、教师审核松懈与制度防线空缺的共同结果。

### 3. 生成机理：生成式人工智能嵌入思想政治教育的风险逻辑

生成式人工智能深度嵌入高校思政教育的实践过程中，相关伦理风险呈现出多源并发、持续演化、系统关联的复合特征。这些风险既源于生成技术的内在局限，也关联教育制度的适配滞后，更与师生的实际使用行为密切相关。从整体来看，生成式人工智能引发的伦理风险主要源自内生性、功能性与实践性三重逻辑机制的交互作用。

#### 3.1. 内生性机理：算法局限性与思政教育价值引导之间的矛盾张力

生成式人工智能本质上是一种依赖海量语料与统计规则进行语言生成的技术，其核心逻辑是“内容预测”而非“价值判断”。这使其在服务高校思政教育时，难以自然承担价值引导、意识形态塑造等教育目标，进而形成技术逻辑与教育逻辑之间的矛盾张力。

##### 3.1.1. 算法的历史数据依赖性限制了内容的时效性适配

高校思政教育强调紧扣时代脉搏、回应社会现实，具有高度的政策敏感性与理论更新要求。但当前主流生成模型(如 ChatGPT、文心一言等)大多基于静态语料库进行训练，数据更新滞后，难以有效覆盖党的最新理论成果和社会热点议题<sup>[5]</sup>。<sup>81</sup>当教师依赖此类工具生成课件时，极易出现内容与现实脱节、理论表达滞后等问题，从而削弱课堂的引导性与权威性。

##### 3.1.2. 训练语料的文化偏向与价值立场隐含风险

在全球数据结构仍以西方话语体系为主的背景下，ChatGPT 在面对政治制度、社会公平或意识形态等议题时，常呈现“价值中立”或“普世主义”式表述，回避集体主义、制度自信等中国语境中的核心价值表达。此种“算法中性化”现象实质上反映了技术生成背后的话语权结构，既削弱了高校思政教育的政治引领功能，也潜在影响学生的价值判断与意识形态认知。

##### 3.1.3. 模型的“通用优化”策略难以满足思政教育中的“因材施教”需求

高校思政教育强调因材施教与个体思辨能力的培养，但生成式人工智能模型倾向于推送平均化、标准化内容，忽视学生的个体差异。长此以往，学生将陷于被动接受的模式，思维路径趋于固化。这种“去个性化”的生成模式，与高校思政教育致力于培养全面发展人才的根本目标形成结构性冲突。

### 3.2. 功能性机理：工具理性主导下思政教育的人文价值弱化

生成式人工智能凭借其高效、精准、可扩展的工具特性，正在加速渗透教育实践，成为教学组织与知识生成的重要参与者。然而，当工具理性逐步主导教学逻辑时，高校思政教育中“以价值育人”为核心的功能定位也面临被技术逻辑边缘化的风险。

##### 3.2.1. 教师对技术的依赖日益加剧导致教育行为出现“工具化倾向”

在高校教学实践中，许多教师开始广泛使用生成式人工智能进行课件制作、案例构建乃至讲解内容生成，形成对技术的功能性倚赖。这虽提升了教学效率与课程资源的生成能力，却在无形中削弱了教师的价值引导职能，使其逐渐从“育人主体”退化为“技术协调者”和“内容审核者”<sup>[7]</sup>。高校思政课堂原有的价值引领、思想碰撞与情感感染等“人文特征”也将随之淡化，教学过程的深度与情感温度显著下降。

##### 3.2.2. 学生对生成式人工智能生成内容的过度信任催生学习行为的机械化与思辨力的弱化

在生成式人工智能的广泛辅助下，不少学生将其作为撰写作业、发言乃至思考问题的主要工具，长

此以往极易陷入“技术舒适区”。这不仅削弱了其对复杂政治议题的敏感性和判断力，也使得本应充满“思想交锋”的思政教育情境，被生成内容的“中性输出”与“安全表达”所取代，从而导致学生批判意识、公共表达与价值选择能力的弱化。

### 3.2.3. 生成式人工智能系统设计中的“政治中立性”与“风险规避性”虚设掩盖了敏感性或争议性的价值议题

这种“避实就虚”的生成机制虽在一定程度上保障了内容把关的安全性，却也压缩了高校思政课堂中引导学生面对矛盾、厘清立场、凝聚认同的教育空间[8]，不利于其形成坚定的政治立场与清晰的价值方向。

## 3.3. 实践性机理：认知水平不足与制度配套缺位带来的风险放大

生成式人工智能在教育实践中所引发的伦理风险，也深受师生及管理者的认知水平与制度环境影响。在实际应用过程中，技术认知不足、伦理判断能力薄弱以及制度机制滞后共同构成了伦理风险放大的现实根源。

### 3.3.1. 教育主体的风险意识与内容把关能力普遍不足

不少高校教师在使用生成式人工智能时，过度强调其便利性与效率，忽视了内容审核与价值导向的重要性。一些教师将生成内容直接用于课堂讲授，未进行必要的甄别与再加工，导致带有立场偏差、事实错误或话语歧义的内容进入教学现场，进而干扰学生的价值判断与政治认知。

### 3.3.2. 制度层面的配套缺失加剧了伦理治理真空

尽管生成式人工智能已广泛应用于高等教育场景，但多数高校尚未建立明确的生成式人工智能教学工具使用规范，缺乏系统性的伦理审查机制与责任划分标准。教师在操作中往往“各自为政”，对于使用边界、审查责任及风险预警缺乏统一共识[9]。一旦引发争议，难以在制度层面进行有效追责，也无法及时进行干预和修复，从而形成治理“空窗期”。

### 3.3.3. 生成式人工智能“算法黑箱”的不透明特性进一步加剧了风险的不确定性[10]

生成式人工智能模型的生成逻辑高度复杂，输出内容缺乏可解释性和可追溯性，使得教学过程中一旦出现政治偏差或敏感内容，难以厘清源头与责任主体。这种技术不确定性和制度滞后性的叠加，导致伦理风险呈现隐蔽化、复杂化与难控化趋势。

## 3.4. 赋能潜力：生成式人工智能的教育促进功能

尽管生成式人工智能在嵌入高校思政教育过程中带来伦理风险与价值偏移，但从教育发展的长远视角看，其在认知促进与教学创新方面仍具有积极意义。

### 3.4.1. 赋能学生认知发展

凭借强大的语言理解与知识整合能力，生成式人工智能可作为学生的“认知脚手架”，辅助其在复杂议题探究中实现知识迁移与结构化思维。通过智能问答与情境模拟，学生能在多元语境中比较不同的价值立场，提升批判思维与价值辨析能力，从而由“被动接受”转向“主动思考”。同时，其信息整合能力也为高校思政教学拓展了跨文化理解的空间，帮助学生在全球化视野中强化对中国特色社会主义道路与制度自信的认同。

### 3.4.2. 助力教师教学创新

生成式人工智能的应用有助于教学精准化与课堂创新。其在数据分析、案例挖掘与互动设计中的功

能，能帮助教师识别学生需求、优化教学内容、提升课堂参与度。因而，在强化伦理规制的同时，更应坚持“以用促管、以管促用”的治理理念，通过制度引导与伦理教育并举，培育师生的数字智慧与技术自觉，使生成式人工智能真正成为推动高校思政教育提质增效的有效工具。

## 4. 三维协同：生成式人工智能嵌入思想政治教育的风险规制

面对生成式人工智能在高校思政教育中引发的复合伦理风险，单靠教师个体自律已难以有效应对，亟需构建多元协同、系统联动的综合治理机制。应以教育主体、平台技术与制度规范三维协同为核心，形成“人-技-制”三位一体的伦理防控体系，在趋利避害中充分释放其“认知脚手架”的积极潜能，确保高校思政教育的价值引领功能不被技术异化。

### 4.1. 主体规制：加强教育系统内部伦理治理能力，重塑教师主导地位

在生成式人工智能深度介入高校教学场域的背景下，教师作为思政教育的首要责任人，应重新确立其在价值传导与风险防控中的主导地位，构建兼具技术适应力与伦理判断力的专业化教师队伍。

#### 4.1.1. 提升教师抗风险能力

教师的技术判断能力与价值审查能力是抵御风险的首要屏障[4]。高校可通过专题培训班、教学沙龙、案例研讨等形式，使教师掌握生成式人工智能的基本原理、运行机制与潜在风险，理解其内容生成的逻辑特点与局限性；并增强其在教育伦理、意识形态安全及舆情引导等方面的敏感度与责任感。

#### 4.1.2. 明确教师内容把关责任

生成式人工智能的输出不具备自主价值判断能力，因此必须由教师进行审查、筛选与再加工。教师在使用AI生成教学资源时，应结合教学目标、课程内容与学生特征，进行适配性审视，确保其内容符合政治导向与教育立场[11]；同时，高校应将此能力纳入教学评价体系，引导教师在技术支持下重建专业权威。

#### 4.1.3. 鼓励教师科学借力人工智能

生成式人工智能可作为启发式工具服务于教学创新，但不能取代教师在政治引导、情感沟通与道德塑造中的核心功能。高校可探索构建教师生成式人工智能伦理培训模块，涵盖技术认知、教学创新与风险反思等单元，帮助教师在技术应用中实现从“被动适应”向“主动引导”的转变。通过系统培训与案例教学的结合，教师能够在技术应用中保持政治敏感与伦理理性，确保AI赋能过程始终服务于立德树人的核心目标。

### 4.2. 技术规制：规范平台技术伦理标准，完善算法可控与审查机制

生成式人工智能在高等教育中的伦理风险，很大程度上来源于其技术设计逻辑的“黑箱性”与算法运行机制的不透明性。这种不确定性使其输出内容难以预测，内在的价值导向与偏见倾向难以精准识别与有效规制[12]。因此，亟需从模型训练、内容生成与平台监管等环节建立技术层面的规范机制，确保技术输出符合意识形态安全与价值引导的根本要求。

#### 4.2.1. 应在模型训练阶段引入系统性价值筛选机制

高校可设立由思政教师、技术专家与伦理学者共同组成的“AI内容安全审查组”，实施“人工审核-算法检测-动态评估”三步流程，对生成内容进行价值取向与语义偏差分析。通过建立“教育语义安全数据库”与师生反馈系统，实现风险识别的常态化与动态化。

#### 4.2.2. 建立生成式人工智能教学平台的备案管理与技术审计机制

高校在引入第三方生成式人工智能教学平台前，应由党委宣传部、网络安全办与教务系统联合评估

平台功能边界、生成逻辑及数据存储路径[13]。平台方应定期提交生成内容日志及使用报告，接受高校内部监管与教育主管部门的外部抽查，形成持续性、闭环式的监管体系。

#### 4.2.3. 借鉴实践成功的共性治理范式

部分高校已探索“人工标注－技术推荐－教师复核”的三级审核模式，有效实现了人机协同与风险防控的平衡[14]。该机制可作为高校推广的可行治理范式，在确保技术应用效率的同时保障教育内容安全。

### 4.3. 制度规制：健全协同治理体系，推动大思政格局下的协同应对

生成式人工智能引发的伦理风险具有系统性与结构性特征，治理工作不应局限于课堂环节的局部应对，而应在制度设计、主体联动与治理体系建设层面实现统筹推进，构建覆盖“政策－平台－个体”的协同治理结构[15]。

#### 4.3.1. 加强制度顶层设计，构建规范化治理框架

教育主管部门应尽快制定《高校生成式人工智能教学使用管理办法》专项制度，明确平台、教师与管理部门的权责边界。高校也应制定《高校生成式人工智能教学应用伦理指南》，明确师生使用AI的责任条款，包括AI生成内容符合国家意识形态安全与数据保护要求；AI生成成果在教学与研究场景中须注明技术来源；AI工具应服务于价值塑造与思维启发，不得取代教师主导地位；建立校级伦理委员会定期评估AI教学使用情况等内容。通过制度建设与常态化监管并举，可实现伦理治理的可持续性与可复制性。

#### 4.3.2. 强化过程监管与平台运行监督

高校教务系统应设立生成式人工智能教学工具使用备案模块，对教师使用频率、内容类型与功能用途进行全过程记录与分析，建立数据可视化的伦理风险预警机制[16]。对于涉及高校思政课程等意识形态敏感环节，应配置专门审查人员，确保课程内容的政治正确性、表达规范性与价值合规性。

#### 4.3.3. 推动多元主体协同参与治理

高校应依托“大思政”工作格局，联合党委宣传部、信息技术中心、学生事务部门及社会力量，形成“校内共治一校外协同”的联动机制。政府相关部门应引导平台企业与教育机构共同参与伦理标准建设，明确教育数据边界、内容审核规范与知识产权归属，共同筑牢技术伦理风险的社会防线。

#### 4.3.4. 注重家庭与媒体的协同引导功能

家长与媒体在学生认知建构与行为规范中扮演着关键角色，应共同推动形成“科学使用、合理审查、理性思辨”的生成式人工智能使用文化。通过校内外合力营造技术治理与价值育人并重的环境氛围，增强学生的信息素养与伦理判断力，共同筑牢技术伦理风险的社会防线。

### 4.4. 实证支撑：AI嵌入高校思政教学的现实反馈

为验证上述规制路径的可行性，笔者于2025年5月在昆明理工大学开展了“生成式人工智能在高校思政教育中的应用与伦理感知”问卷与访谈研究，共回收有效问卷312份（教师102份、学生210份）。结果显示，78.4%的教师认为AI在教学中存在内容审核与政治导向识别难题，65.7%的学生感知到AI的“价值中立”倾向，43.2%的学生存在不同程度的技术依赖心理，56.9%的教师认为AI生成内容削弱了课堂的思想深度，72.8%的学生在AI辅助写作中难以把握价值导向，反映出“人机协同”与“价值引领”间的现实张力。案例观察发现，将AI答疑系统引入思政课堂后，虽提升了学生学习效率，但也出现了算法偏向问题。例如，学生检索“制度自信”等概念时，系统生成“普世主义”式话语，而非中国化语境下的价值表达。该校随后实施的“人工审核－算法预警－教师复核”三阶段机制，有效降低了内容偏差风险，

验证了“人 - 技 - 制”三维协同治理思路的可行性。实证表明，生成式人工智能在高校思政教学中的伦理风险具有可识别性与可调控性。其有效治理应以教师主导、技术审查与制度支撑为核心，形成动态共治格局，唯有在制度引领与伦理自觉的共同作用下，技术赋能方能真正转化为教育育人的正向力量。

## 5. 结论：在趋利避害中化解矛盾与互动育人

生成式人工智能嵌入高校思政教育所引发的伦理问题，并非其本身所固有的“恶”，而是在缺乏规范引导、制度保障与价值审查的情境下，技术逻辑与教育伦理之间张力外显的结果。这一问题的关键，不在于是否应用技术，而在于如何科学、合理、合伦理地使用技术。尤其是在高校思政教育领域，教育具有极强的政治属性和价值引领功能，任何教学工具的使用都不能脱离意识形态导向和立德树人的根本任务。因此，面对生成式人工智能技术带来的挑战与变革，教育者更应主动适应、理性引导、规范使用，把握新技术介入冲击高校思政教育效果的本质，在动态平衡中解决好人的价值属性与技术的工具属性之间的矛盾冲突[17]，在趋利避害中真正实现技术服务教育、技术促进育人的价值功能。

## 基金项目

教育部人文社会科学研究西部和边疆地区项目“新时代高校思想政治理论课立德树人规律研究”(24XJA710005); 昆明理工大学马克思主义学院“中外思想政治教育理论与实践比较研究团队”建设项目。

## 参考文献

- [1] DATACAMP (2024) AI in Education: Benefits, Challenges, and Ethical Considerations. <https://www.datacamp.com/blog/ai-in-education>
- [2] 张彦. 价值排序与伦理风险[M]. 北京: 人民出版社, 2011: 60.
- [3] 马克思, 恩格斯. 马克思恩格斯文集: 第 8 卷[M]. 北京: 人民出版社, 2009: 358.
- [4] 张娟, 邓卓明. 生成式人工智能赋能思想政治教育的伦理风险与防范[J/OL]. 教学与管理, 2025(18): 80-89. [https://xueshu.baidu.com/usercenter/paper/show?paperid=143m0gh0sp6800r0qq790tr0j9086980&site=xueshu\\_se](https://xueshu.baidu.com/usercenter/paper/show?paperid=143m0gh0sp6800r0qq790tr0j9086980&site=xueshu_se), 2025-04-10.
- [5] 任凤琴, 董子涵. 风险与超越: 生成式人工智能赋能思政教育的伦理分析[J]. 重庆邮电大学学报(社会科学版), 2023, 35(6): 80-89.
- [6] 邓含双. 人工智能赋能思想政治教育的伦理风险及规避路径[J]. 林区教学, 2024(4): 7-11.
- [7] 米华全. 智能思政伦理风险的生成逻辑、表现形式及防控机制[J]. 中国电化教育, 2023(2): 111-117.
- [8] 周刚. 生成式人工智能嵌入思想政治教育的伦理风险及其治理路径[J]. 唯实, 2024(5): 7-11.
- [9] 冯子轩. 生成式人工智能应用的伦理立场与治理之道: 以 ChatGPT 为例[J]. 华东政法大学学报, 2024, 27(1): 61-71.
- [10] 苗芳艳. 风险、成因与化解: 语境主义视角下生成式人工智能的数据操纵、算法偏见与伦理审视[J]. 昆明理工大学学报(社会科学版), 2025, 25(2): 49-56.
- [11] 姜华, 王春秀, 杨暑东. 生成式 AI 在教育领域的应用潜能、风险挑战及应对策略[J]. 现代教育管理, 2023(7): 66-74.
- [12] 胡洪彬, 朱星宇. 生成式人工智能与思想政治教育: 机遇、挑战及其应对[J]. 中共山西省委党校学报, 2024, 47(2): 122-128.
- [13] 吕丹阳, 郎元柯, 范柏乃, 等. 生成式人工智能在公共服务中应用的机遇与挑战[J]. 电子科技大学学报(社科版), 2024, 26(3): 35-45.
- [14] 殷文轩. 生成式人工智能赋能大学教学的作用机制与推进路径——基于斯坦福大学的案例分析[J]. 教育探索, 2025(4): 33-39.
- [15] 胡祥秋. 生成式人工智能应用于思想政治教育的意义、风险与对策[J]. 科教文汇, 2024(13): 40-43.
- [16] 杜江舟. 高校教务管理系统信息化建设研究[J]. 科教文汇, 2023(21): 13-15.
- [17] 刘箫峰, 张锦霖. 生成式人工智能冲击高校思政教育的三维探赜[J]. 国家教育行政学院学报, 2023(12): 66-75.