

# 生成式人工智能对大学生价值观引领的风险与应对策略

苟雅倩

山东财经大学马克思主义学院, 山东 济南

收稿日期: 2026年4月2日; 录用日期: 2026年5月2日; 发布日期: 2026年5月11日

## 摘要

伴随数智化浪潮的推进, 生成式人工智能也逐渐渗入教育领域的各个环节, 对大学生思政教育产生了深刻的影响。虽然生成式人工智能有丰富的教育资源并且在育人精准性方面有着显著提升, 但是随之而来的问题也不容忽视。如信息茧房、算法黑箱、人机情感淡化和学生过度依赖技术等伦理风险, 这些问题会影响大学生的价值观塑造和认知能力发展。所以, 基于价值引领视角, 分析生成式人工智能对大学生思政教育的伦理风险, 重点围绕价值导向偏差、主体能力弱化和情感认同淡化等问题展开探讨。只有把技术合理应用与正向价值引导结合起来, 才能让大学生思政教育在数智时代真正实现提质增效、高质量发展。

## 关键词

生成式人工智能, 大学生价值观, 伦理风险, 应对策略

# Risks and Response Strategies of Generative Artificial Intelligence in Guiding College Students' Values

Yaqian Gou

School of Marxism, Shandong University of Finance and Economics, Jinan Shandong

Received: April 2, 2026; accepted: May 2, 2026; published: May 11, 2026

## Abstract

With the advancement of the digital-intelligent wave, generative artificial intelligence has gradually

penetrated all aspects of the educational field, exerting a profound impact on the ideological and political education of college students. Although generative artificial intelligence provides abundant educational resources and significantly improves the precision of education, the accompanying problems cannot be ignored. Ethical risks such as the information cocoon, algorithmic black box, weakened human-machine emotion, and excessive student dependence on technology affect the formation of college students' values and the development of their cognitive abilities. Therefore, from the perspective of value guidance, this paper analyzes the ethical risks of generative artificial intelligence in college students' ideological and political education, focusing on issues including biased value orientation, weakened subjective ability, and diminished emotional identity. Only by combining rational application of technology with positive value guidance can ideological and political education for college students achieve quality improvement, efficiency enhancement, and high-quality development in the digital-intelligent era.

## Keywords

Generative Artificial Intelligence, College Students' Values, Ethical Risks, Countermeasures

Copyright © 2026 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

近年来,随着 ChatGPT、DeepSeek、豆包和千问等生成式人工智能的爆火,引发了大学生的广泛关注,也应用到了大学生以及高校的教学和学习当中。生成式人工智能在给大学生和教师带来益处的同时,也带来了风险和挑战。生成式人工智能依托数据、算法、模型等核心技术,为大学生提供了极具特色的信息和内容,提升了信息获取和理解的效率。正因如此,在信息筛选、价值导向等方面潜藏诸多风险,会对大学生的认知及价值观塑造产生严重的影响,必须予以重视。为了贯彻落实立德树人根本任务,在数智时代借助人工智能赋能大学生思想教育,对大学生的价值引领起到规范和导向的作用,本文从价值视角出发,探讨生成式人工智能对大学生价值观引领的风险及其应对措施。

## 2. 生成式人工智能与大学生价值观的概念界定

### 2.1. 生成式人工智能

生成式人工智能是指基于深度学习等技术,通过对大规模数据进行训练,能够自动生成文本、图像、音频、视频及代码等多种内容的人工智能系统。与传统的以分类、识别和预测为主的判别式人工智能不同,生成式人工智能更强调内容的“生成性”,即在学习数据分布规律的基础上,输出具有一定原创性和语义合理性的全新内容。以当前广泛应用的生成式人工智能系统 GPT-5.4 版本为例,其作为典型的大语言模型,能够在自然语言理解与内容生成方面展现出较强能力。例如,在学习场景中,学生可以输入“请解释某一理论”或“帮助撰写论文提纲”,系统能够基于海量语料生成结构清晰与符合逻辑的文本内容,从而辅助大学生的学习和写作。

### 2.2. 大学生价值观

价值观是指个体或群体在一定社会历史条件和文化背景下形成的,是对事物重要性与意义的根本看法和评价标准,是指导人们认识世界和改造世界的基本思想取向。它主要回答“什么是有价值的”“应

当追求什么”以及“如何进行价值选择”等问题，在人的认知、情感与行为中发挥着导向和规范作用。世界观、人生观、价值观共同构成个体稳定的思想观念体系，其中价值观在行为导向中具有核心作用。大学生价值观是大学生群体在特定社会历史条件、教育环境及自身认知发展阶段的作用下，逐步形成的相对稳定的价值判断体系与行为取向。其核心是大学生对人生价值、社会责任、个人发展及客观事物的根本看法，是个体基于价值认知、价值情感与价值意志，指导自身价值选择、规范行为实践的内在准则，既彰显青年群体的价值特质，也深刻反映社会文化与时代精神对青年成长的影响。

### 2.3. 生成式人工智能对大学生价值观的影响

在数字化与智能化深度融合的时代背景下，“Z世代”大学生成长于互联网与移动终端高度普及的环境之中，自幼接触智能手机、社交媒体及各类数字技术产品，具备较强的信息获取能力与技术适应能力。具体来看，生成式人工智能内部嵌入算法推荐和数据学习，能很好地贴合大学生的兴趣偏好，为他们推送需要的信息。积极地看，这类技术确实拓宽了大学生获取知识、表达想法的渠道，对提升学习效率、培养创新思维有一定助力；但从隐忧来看，这种紧盯兴趣推送信息的模式，也非常容易固化大家的认知，减少接触多元价值的机会，这反倒会给大学生价值观的养成带来深层次的结构性的影响。由于生成式人工智能从本质上来看是技术与资本进行结合的产物，更加注重通过生成的内容来实现经济效益最大化，而这恰好与思政教育的价值取向相悖，两者存在一定冲突[1]。由此可见，生成式人工智能作为当前极具代表性的时代技术，必然会给大学生价值观的塑造过程带来深远的影响。

## 3. 生成式人工智能对大学生价值引领的风险

### 3.1. 信息茧房固化认知结构与创新能力弱化风险

“信息茧房”这一概念是由美国学者凯斯·桑斯坦于2006年在其著作《信息乌托邦——众人如何生产知识》中首次提出。在数字化媒介与算法技术高度普及的现代社会中，个体往往会在算法推荐机制的引导与自身偏好选择的双重作用下，长期地接触与自身固有认知、价值立场和兴趣爱好高度一致的信息内容，而较少接触异质观点、多元视角和批判性信息。长此以往，每个人将像蚕蛹一样将自己束缚在自我偏好的信息城堡之中，无法逾越出自己的偏爱和喜好[2]。并且个体将被封闭在同质化的信息空间之中，形成相对固化的认知结构与价值判断模式，如同被包裹在无形的“茧房”之内，难以形成全面、客观、辩证的认知能力。生成式人工智能依托算法推荐机制，主动贴合大学生的兴趣偏好与价值立场，输出与其心理预期高度契合的内容，进而与预设结论构成自我封闭的认知循环，不断强化“信息茧房”效应，加剧大学生价值观念的固化与偏见。一方面，大学生身处于信息茧房之中，获取信息具有片面性以及窄化和固化思维，导致认知结构范围狭窄，这使得大学生们只能看到世界的一部分，而非世界的全貌，就如井底之蛙，只是局限于自己的一方天地。另一方面，生成式人工智能能够通过学生的提问直接给出表面化的答案，让学生的思维停留在表面，无法使学生深度思考。同时，生成式人工智能也滋生了学生的懒惰，遇到问题不思考，直接通过生成式人工智能获取答案，从而使学生的创新能力不断下降，形成思维的依赖。

### 3.2. 算法黑箱遮蔽价值判断与认知偏向风险

生成式人工智能底层模型的架构与运算逻辑极端复杂，呈现出典型的“黑箱”特征，外界极难看透其信息处理的内部轨迹。在此背景下，模型在数据训练、特征提取与内容生成过程中所隐含的价值偏好和意识形态，都具有较强的隐蔽性和不可知性，容易在缺乏有效检视的情况下被固化于系统运行逻辑之中。值得警惕的是，这种生成式人工智能模型还是以西方为主，存在着诸多种族歧视、反华对策、拜金

主义以及极端自由主义。北京理工大学计算机学院研究者经过实验指出,“ChatGPT 生成内容中存在对于中国的大量偏见言论”[3]。这一行为严重影响到了大学生的价值观念和价值判断,数据算法里面嵌入的以西方主导的价值观念,控制住了大学生们的想法,不利于他们形成正确的世界观、人生观、价值观。生成式人工智能所采用的交互式对话模式存在固有的技术局限,容易在价值观教育与传播过程中产生导向偏差。由于该类技术本身不具备完善的内容甄别与自我纠偏机制,不能够有效地判断教育内容之中的价值立场是否与我国意识形态相一致,也无法主动剔除、修正与社会主义核心价值观相背离的信息与资源。在一味追求流量的情况下,很多低俗内容被推送给年轻人,让好内容越来越难被看到,形成了“劣币驱逐良币”的局面[4]。当生成式人工智能应用于大学生价值观教育实践时,甚至将错误的价值观念包装为合理观点,通过隐性渗透的方式嵌入互动对话过程,进而对我国意识形态的引导力与凝聚力造成削弱。

### 3.3. 情感疏离消解价值认同与人文关怀弱化风险

马克思曾指出:“人作为对象性的、感性的存在物,是一个受动的存在物;因为它感到自己是受动的,所以是一个有激情的存在物”[5]。情感既是人类特有的基础心理表征,也是驱动个体行为实践的关键内在动力。在大学生价值观培育与引领工作中,情感要素更是推动价值观念由认知层面转向内化践行的核心中介载体,而师生主体间的情感互动与精神共鸣,则是助推价值观教育实现深度升华、提升育人实效的重要催化因子。生成式人工智能是依托算法逻辑与数据驱动开展互动交互,缺乏人类主体独有的情感共情和人文温度,在赋能大学生价值观教育的过程中,易引发深度情感疏离问题。价值观教育并非单纯的知识传递与信息灌输,而是由理性引导与情感浸润的双向精神互动,需要教育者以真挚的情感共鸣、个性化的人文体察唤醒大学生的内心认同,这也是意识形态入脑入心的关键所在。这种深层次的情感疏离,会直接弱化大学生的情感共鸣,进而消解价值认同的根基。价值认同的建立离不开情感纽带的支撑,当教育过程缺失人文关怀与情感温度,难以触动大学生的内心世界,甚至会让大学生产生抵触、疏离的负面情绪。与此同时,人文关怀弱化风险进一步凸显,生成式人工智能无法针对大学生的个体差异和思想困惑开展精准化的引导,只能给出标准化的回应,忽视了大学生作为独立个体的精神需求与成长诉求。长此以往,价值观教育会丧失应有的人文底蕴与育人温度,无法实现对大学生思想的正向引领,甚至会加剧大学生的价值迷茫。

### 3.4. 技术依赖消解自我主体性与价值自主建构风险

“技术承载着价值,决定了它可以也应该成为主流价值的载体。”[6]生成式人工智能因其生成迅速、解答即时的特点,已成为许多学生获取思政知识的工具。然而,过度使用所引发的技术依赖问题,正逐渐消解大学生在价值认知与建构过程中的主体能动性,阻碍其自主完成价值体系的形成与完善,成为生成式人工智能在对大学生价值引领过程中迫切需要关注的伦理风险。大学生正处于价值观念形成与定型的关键阶段,自主思考、独立判断与主动建构是其形成正确价值体系的核心前提。而生成式人工智能通过“按需供给”的模式,为大学生提供标准化、现成化的价值答案,使部分大学生逐渐形成“有问题找AI”的思维定式,将本应自主完成的价值思考和观点辨析过程简单交由智能工具完成。当大学生在面对理想信念等价值相关问题时,不再主动调动自身思维进行深度探究,而是直接依赖人工智能的生成结果形成认知,其独立思考的意愿与能力会逐渐弱化,对智能技术的路径依赖会不断强化,最终导致自身在价值引领过程中的主体地位被弱化。技术依赖不仅会消解大学生的价值自我主体性,更会阻碍其价值体系的自主建构,使其价值认知呈现出碎片化和浅表化特征。生成式人工智能所生成的思政内容与价值观点,是基于海量数据的算法整合与浅层输出,缺乏对政理、学理、事理的深度阐释,也无法贴合大学生个体的成长经历和实践场景形成个性化的价值引导。大学生长期依赖此类标准化内容,难以对价值观念

进行系统性和深层次的理解，无法将外在的价值引领要求内化为自身的价值认知，更难以通过自主思考与实践体验完成价值观念的筛选与整合。同时，过度的技术依赖会让大学生丧失价值探索的主动性，在面对多元社会思潮与复杂价值选择时，缺乏独立的价值判断能力与自主的价值抉择意识，容易被人工智能生成的观点所裹挟，导致自身价值体系的建构陷入被动与混乱，难以形成符合社会主流要求和适应个人成长发展的稳定价值观念。

## 4. 生成式人工智能对大学生价值引领的治理路径

### 4.1. 打破算法茧房桎梏，培育多元认知与创新思维

在运用生成式人工智能助力大学生价值观教育的过程中，只有强化意识形态引领，加大数据的“喂养”程度，健全算法纠偏机制，才能真正打破算法茧房桎梏，牢牢把握大学生价值观教育的正确方向[7]。在信息推送方面，生成式人工智能平台不能一直陷在“只按兴趣匹配”的算法惯性里，反而要搭建起“正向适配 + 反向补位”的双向推送模式。一方面，保留基于用户画像的正向推送思路。提供契合学生场景的个性化内容，这样能够满足大家自主学习的需求，也为后续的价值引领打好共情基础；另一方面，更要发挥反向推送的纠偏作用。借助人机交互的数据，在捕捉学生的价值、道德和社会等问题上的盲区，从而能够推送拓宽他们视野等内容。这种反向推送并不是生硬的道理灌输，而是紧扣大学生真实的思想困惑，用多元视角和辩证分析，打破信息茧房带来的认知固化和偏见，引导学生跳出单一和同质化的信息闭环。在全面正向的信息中完成对自我的反思和观念的更新。例如，平台监测到学生长期浏览“躺平”“内卷”这类消极内容，不能顺着搜索内容继续推同类素材，而反向推送青年扎根基层创业和大学生参与西部志愿服务这类主流价值类型，让大学生重新理解奋斗与成长的意义，进而消解认知偏差。

### 4.2. 破解算法黑箱困境，塑造理性价值判断与认知能力

算法作为核心运行机制，直接决定了信息的生成与传播。信任是人际往来的精神纽带，更是思政教育的情感底色，然而算法的黑箱特质与不可解释性，却从根本上瓦解了思政教育赖以存续的信任根基[8]。首先要做的，就是推动算法朝着透明化与规范化的方向发展，降低技术给大学生带来的认知隐患。目前，生成式人工智能在教育领域的应用还在快速发展，算法本身具有专业性强和逻辑复杂等特点，普通学生弄懂这套机制是相当困难的。针对这一点，需要靠完善技术规范和相关制度，让算法在合理范围内具备可解释性和可追溯性。比如，在教育应用中，做好信息来源标注和优化内容生成强化结果审核校验，这都能提升 AI 生成内容的可信度。除此之外，思政教育还要进行理性的引导，提升大学生对算法信息的辨别能力。当前 AI 内容泛滥成灾，学生获取信息的渠道变得越来越多，但信息质量却参差不齐，很容易被算法推荐带着走。这就要求思政教育不能只停留在知识灌输，更要侧重能力培养，帮学生建立起批判性思维和理性判断的意识。实际教学中，应该结合案例分析、情境研讨和问题导向等授课方式，引导学生主动辨析 AI 生成的内容，从而能够练就甄别信息真伪的能力，避免盲目轻信算法输出的结果，在繁杂的信息环境里守住理性底线。

### 4.3. 化解人机情感疏离，强化主流价值认同与人文关怀

大学生正处在价值观成型的关键阶段，价值认同和人文关怀这两点本身就格外重要。要做好这部分工作，首先要坚持以人为本，突出教师在思政教育中的情感引领作用。就算生成式 AI 用得再普遍，教师的角色也绝不可能被技术取代，反而要从单纯的讲课授知，转向做学生的价值引路人和情感倾听者。教师通过言传身教和日常谈心，能够补足情感疏离的短板。所以高校要优化课堂设计，引导教师投入更多的真情实感，从而让思政教育不光有理论厚度，更有人文温度。其次，要强化价值导向。生成式人工智

能系统的设计,不仅涉及技术问题,更涉及价值取向问题。正因如此,把生成式 AI 融入思政教育的过程中,一定要把正确的价值观嵌入技术应用的底层逻辑中,从根源上规避价值偏差问题。与此同时,也要顾及学生的个体差异,把人文关怀贯穿教育全过程,通过个性化引导从而增强学生的归属感和认同感。其实只有让技术带着人文温度,思政教育才能真正做到“润物无声”。比方说高校可以通过情感工作室这些线下阵地,组建辅导的队伍,定期排查学生的心理和情感状态,做好跟踪记录和分类疏导,既守护好学生的身心健康,也帮大家树立起积极正向的价值观念。

#### 4.4. 消解智能技术依赖, 重塑价值自我主体性与自主建构能力

生成式人工智能靠着便捷的内容生成与智能答疑的优势,在给大学生带来高效学习便利的同时,也很容易让大家养成“技术依赖”的习惯。不少学生过度仰仗智能工具给出的现成答案,懒得自主思考和独立判断,也就失去了主动实践探索的劲头。针对这类问题,首先要牢牢守住主体性原则,保障大学生在价值引领中的主体地位。技术说到底只是人类认知和价值的延伸,想要破除技术依赖,最核心的就是把人的主体性和能动性放在技术之上。一方面,要在技术设计阶段就筑牢“人控技术”的底层逻辑,不能让算法打着“个性化推荐”的幌子,把学生困在信息茧房里,滋生思维惰性;另一方面,也要借助制度约束和教育引导,让大学生认清一个事实,就是智能工具终归只是辅助手段,并不是可以替代思考的“万能答案库”。其次,必须搭建隐私防护制度,筑牢学生隐私安全防线。一方面,要落实数据采集“最小必要”原则。高校可以联合技术平台,只收集学习辅助和价值引导直接相关的必要数据,不会采集人脸信息、心理状态和社交轨迹这类敏感隐私数据。例如,智能学习平台只需采集学生浏览价值类内容的时长,不会过多记录对话细节。这样既能保证技术适配性,也能切实维护学生的隐私权益。另一方面,要强化隐私保护的监管和问责力度。可以成立专门的技术审查小组,定期对 AI 平台上的数据进行安全审查,查看使用全流程是否合规。只有这样才能让大学生在安全的技术环境里,摆脱对技术的依赖,从而敢于主动表达想法和自主建构正确价值观念。

### 5. 结语

生成式人工智能伴随着社会的变迁和人类的需要迅速发展,并且在大学生思政教育中的使用也日益频繁,这为创新教育和提升育人质量提供了重要的途径。但是在技术赋能的过程中,信息茧房、算法黑箱、人机情感疏离和学生过度依赖技术等问题逐渐浮出表面,影响了大学生的价值判断和认知结构,也给思政教育育人功能带来新挑战。本文基于价值引领视角,分析生成式人工智能在思政教育中的应用不能仅停留在技术层面,也要兼顾价值导向与育人目标。在具体实践中,要推进算法透明与规范建设,化解技术不确定性引发的认知风险。此外,也要强化教师情感引领与人文关怀,弥补人机交互的情感缺失;加强大学生科技伦理与媒介素养教育,提升他们在理性认知和价值判断的能力,守住自我主体性与批判性思维。总体而言,生成式人工智能给大学生思政教育带来机遇与挑战并存的局面。只有坚持价值引领,统筹技术理性与价值理性,才能发挥技术正向作用,推动思政教育在数字时代创新发展。

### 参考文献

- [1] 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]. 教学与管理, 2025(18): 80-89.
- [2] 龚莉红. 基于“信息茧房”理论的意识形态话语权研究[J]. 河海大学学报(哲学社会科学版), 2019, 21(5): 35-40, 106.
- [3] 张华平, 李林翰, 李春锦. ChatGPT 中文性能测评与风险应对[J]. 数据分析与知识发现, 2023, 7(3): 16-25.
- [4] 王瑞, 吴蒋雪. 生成式人工智能对青年价值观影响的逻辑理路、双重镜像与应对策略[J]. 辽宁行政学院学报, 2025(6): 82-89.

- [5] 马克思恩格斯文集(第1卷) [M]. 北京: 人民出版社, 2009: 211, 320.
- [6] 卢俞成, 陈佳虹, 张艺琼. 生成式人工智能赋能大学生价值观教育的价值向度、风险挑战及应对策略[J]. 继续教育研究, 2025(8): 86-92.
- [7] 谢俊, 彭福莹. DeepSeek 类生成式人工智能融入思想政治教育的机遇、挑战及应对策略[J]. 思想教育研究, 2025(8): 115-122.
- [8] 本报评论部. 用主流价值纾解“算法焦虑” [N]. 人民日报, 2018-06-20(005).