

AI辅导员的价值引领风险及应对策略研究

陈蔚

四川外国语大学国际金融与贸易学院, 重庆

收稿日期: 2026年1月22日; 录用日期: 2026年2月23日; 发布日期: 2026年3月5日

摘要

随着人工智能技术与教育工作的深度融合, AI辅导员作为思政教育智能化的新兴形态, 在提升育人效率的同时也潜藏着价值引领风险, 包括西方国家利用技术优势施加价值影响、生产内容失真及使用者人工智能素养不足以及辅导员的教育主体地位弱化。为此, 应当构建以技术规制为核心的发展路径, 强化对大学生的人工智能素养教育, 提高辅导员思政教育工作主动性与价值引领力, 以期实现人工智能与思政教育高质量发展的有机统一。

关键词

人工智能, 辅导员, 价值引领, 思想政治教育

Research on the Value Guidance Risks and Coping Strategies of AI Counselors

Wei Chen

College of Finance and Economics, Sichuan International Studies University, Chongqing

Received: January 22, 2026; accepted: February 23, 2026; published: March 5, 2026

Abstract

With the deep integration of artificial intelligence technology and education, AI counselors have emerged as an intelligent form of ideological and political education. While improving educational efficiency, this development also carries potential risks in value guidance. These risks include Western countries leveraging technological advantages to exert value influence, distortions in AI-generated content, insufficient AI literacy among users, and the marginalization of counselors as primary educational agents. In response, it is essential to establish a development pathway centered on technical governance, strengthen AI literacy education for university students, and enhance the initiative and value-guidance capacity of counselors in ideological and political education. This approach

aims to achieve a harmonious integration of AI and high-quality development in ideological and political education.

Keywords

Artificial Intelligence, Counsellor, Value Guidance, Ideological and Political Education

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

人工智能在教育领域的运用广度与深度正被不断拓展,成为我国教育现代化的重要支点。中共中央、国务院印发的《教育强国建设规划纲要(2024~2035年)》指出,要“促进人工智能助力教育变革”、“打造人工智能教育大模型”、“推动思想政治工作和信息技术深度融合”。随着教育数字化战略的深入推进,人工智能辅导员(以下简称 AI 辅导员)作为技术赋能思政工作的前沿形态应运而生,它在优化高校学生事务管理、心理健康咨询、思政教育等方面展现出巨大潜力。人工智能辅导员深度融合了生成式 AI 技术与思政教育规律,能够提供精准化、个性化的教育服务,开辟了思政工作智能化转型的新路径。如何趋利避害,在发挥 AI 辅导员育人功能的同时,避免技术赋能背后的价值引领风险,是当前亟待探究的重要课题。

2. 人工智能辅导员的内涵

2.1. 高校辅导员的核心任务

教育部令第 43 号《普通高等学校辅导员队伍建设规定》明确规定,高校辅导员是开展大学生思想政治教育的骨干力量,是高等学校学生日常思想政治教育和管理工作的组织者、实施者、指导者。高校辅导员队伍是和高校教学、科研、管理同等重要的“第四支队伍”,是全学段思政教育的主力军。

2.2. 人工智能辅导员的特征

生成式人工智能(Generative AI)指通过人工智能技术,实现文本、图像、视频、音频等内容自动化生成的一类人工智能。2022 年 11 月,ChatGPT 的发布,标志着生成式人工智能实现了从单一领域的应用部署到多产业的技术赋能,也引发了全球的研究热潮。随后, Claude、GroK、DeepSeek 等生成式人工智能大模型不断取得突破性进展,生成内容质量持续提升,显著提高了人类社会各行业生产力水平。生成式人工智能所具有的知识处理与创造性输出的特性,对教育领域产生了直接且深刻的影响。陈宝生同志曾指出,现代教育的三个突出特征使得现代教育成为最先被 ChatGPT 替代的领域[1]。

AI 辅导员便是在这种背景下所催生出的一种生成式人工智能新形态。它可定义为:以 ChatGPT 类大语言模型为核心引擎,通过预训练语言模型、上下文学习及人类反馈强化学习等关键技术,模拟人类辅导员对话逻辑的智能辅助育人系统。AI 辅导员能够利用深度学习算法处理海量教育数据,面向大学生群体实现学业指导、职业规划、心理疏导及思想政治教育等工作内容的自动生成和输出。然而,技术的局限性和应用场景的复杂性,使得 AI 辅导员在融入思想政治教育生态时,反而会对价值引领这一核心职责构成挑战。

2.3. 人工智能辅导员的发展现状

目前, 大多学者仍将人工智能辅导员视为高校辅导员的辅助工具, 强调其在事务处理、信息整合及基础咨询等领域的效率提升作用。如唐志凤认为 ChatGPT 有助于将辅导员从机械性工作中解放出来, 促进辅导员由事务型转变为专家型, 但无法替代辅导员的教育主体地位, 辅导员仍应当主动肩负起思政教育的主体职责[2]。但随着智能技术迭代与教育数字化转型深化, 国内许多高校已经部署了具有独立身份和完整功能的 AI 辅导员, 并使其成为专业的教育主体。如华中科技大学推出的思政教育智能体“爱华导”, 通过专项建设思政工作语料库, 让“爱华导”的交互语言充满思政味和温馨感[3]。蒋莉薇认为, 基于算法大数据的“AI 辅导员”, 能够以更符合青年特征的方式推送信息, 并以虚拟人的形式参与交流, 引导学生树立正确的三观[4]。

可见, AI 辅导员的定位正悄然突破工具边界向独立主体演进, 其直接承担思政教育任务的实践模式将重塑高校育人生态。未来, 随着生成式人工智能技术的发展, AI 辅导员将在思想政治教育领域发挥更加重要的作用, 但技术便利下潜藏着不容忽视的价值引领风险。有效防范化解此类风险、践行育人初心, 是贯穿 AI 辅导员发展始终的基本原则和核心要求。

3. AI 辅导员可能引发的价值引领风险

3.1. 西方国家利用人工智能技术优势施加价值影响

AI 辅导员所依托的生成式人工智能技术, 其研发与应用高度依赖技术平台的数据与算法优势。当前, 全球人工智能领域呈现激烈的战略竞争态势, 美国凭借其技术先发优势在基础模型研发上处于领先, 并通过人工智能技术优势对用户施加价值影响。

一方面, 人工智能技术发展始终受制于政治干预。美国将人工智能方面的竞争视为中美竞争的关键, 通过“小院高墙”策略来限制中国包括人工智能相关技术在内的关键科技领域发展, 来维持对其“竞争者”的技术优势[5]。西方政治活动中网络舆论经常被操控的事实也表明, 这类人机交互技术能够被恶意编辑, 编造虚假信息以服务政治目的。

另一方面, 不同人工智能大模型所输出的内容会因模型、算法、数据资源等的不同, 天然存在价值观差异。西方资本主义国家在生成式人工智能的开发、创新方面还存在一定优势, 许多生成式人工智能的核心技术仍由西方国家主导, 这些系统的底层架构——包括数据采集逻辑和算法设计原则等——不可避免地植入了西方资本主义的价值观念与思维方式。根据用户观察, ChatGPT 在交流中明显存在意识形态倾向: 其生成的观点往往更符合美式价值取向; 尤其在谈及中国问题时, ChatGPT 屡对中国发展道路和社会制度做出有失客观的评判[6]。

人们如果长期处于生成式人工智能应用环境中, 会不自觉地受到人工智能技术背后的开发者施加的影响。而以生成式人工智能技术为核心的 AI 辅导员, 在为大学生自动生成思政内容的同时, 也成为了西方价值传播的新通道, 对青年大学生的价值观塑造构成潜在威胁。此外, 别有用心之人还可能对语料库等进行攻击、污染, 使 AI 辅导员生成有害内容并推送给大学生。

3.2. 生成内容失真及使用者人工智能素养不足

生成式人工智能的内容生产机制, 本质上是对大规模数据的规律的深度探索与整合重组[7]。

一方面, AI 辅导员依据现有文本数据集中的语言统计特征, 借助算法模型进行可能性推断, 以生成爱国主义等教育内容。这种依赖于概率语言重组的生成方式, 不可避免地会引发信息真实性与可靠性的问题。如果训练所使用的数据先天存在不足或包含错误信息, 则生成结果往往出现系统性偏差。如果教

育工作者过于信赖信息技术，不加批判地接受其生成的内容，甚至奉为圭臬，就难以识别其中存在的事实性错误，并继续输出给学生。而当传统教育中基于真实历史的爱国主义教育材料与人工智能生成的内容明显矛盾时，学生会对教育内容的真实性提出疑问，这将严重损害爱国主义教育应有的公信力与说服力[8]。

另一方面，当前大量大学生群体仍不具备足够的人工智能素养，缺乏人工智能存在局限性和潜在偏差可能性这一基本认识，倾向于将算法输出内容等同于客观真理，从而弱化了自身的独立思考与价值辨别能力。基于对高等院校这一权威机构的信任，大学生们会认为学校推出的 AI 辅导员及生成内容具有同样的权威性与正确性。这会放大生成内容失真所带来的负面影响，不仅难以实现有效引导，反而可能强化认知偏差，阻碍正确价值观念的形成[9]。

3.3. 辅导员的主体地位弱化

人工智能的使用淡化了辅导员在教育过程中的主体地位。高校辅导员工作的核心在于人际互动，本质上是围绕学生开展的思想政治教育与管理工作的。而这一过程不仅依赖于理性的知识传递，更离不开情感的共鸣与信任的建立。教育主体与受教育者之间的情感联结，是实现教育目标的心理基础，也是辅导员话语产生实效的关键支撑。沟通作为辅导员实现教育目标的关键媒介，其使用情境、信息内容与表达形式均对学生的成长产生深远而持久的影响。

随着生成式人工智能的深入应用，AI 辅导员能够响应学生的多样化需求，为其提供学业辅导、生活建议与心理支持等，这使得学生越来越多地依赖人工智能而非辅导员解决问题。师生间传统的“辅导员-学生”直接交流模式将转变为“辅导员-AI 辅导员-学生”间接交流模式。该转变冲击了原本基于共情与信任而形成的教育情感共同体，在一定程度上淡化了辅导员作为教育主体的存在感，话语影响力和实效性受到削弱，原有的教育引导作用难以充分发挥。

在此情况下，辅导员对 AI 辅导员生成内容进行及时识别与干预的难度也大大增加。最终可能动摇辅导员在价值引导过程中的权威性与有效性，对育人工作的整体质量构成挑战。

4. 对价值引领风险的应对策略

4.1. 构建以技术规制为核心的发展路径

在 AI 辅导员的技术路径选择上，应优先采用 DeepSeek 等自主可控的国产大模型。一是其打破了国外生成式人工智能的技术垄断，能够避免国外技术开发者在技术中强化价值逻辑。二是其针对中文语境进行了优化，更熟悉中国文化、历史、社会习俗和价值观，能更好地解决中文自然语言处理难题[10]。三是 DeepSeek 等国产人工智能大模型采用的逻辑思维链更加接近人类的思考模式，能够自行识别、纠正错误内容，为思政教育提供更安全可靠的技术支撑。除了出于对价值安全的考量以外，使用国产大模型也是解决数据隐私这一核心合规问题的关键。具体而言，应大力推动 AI 辅导员系统的“本地化部署”。这意味着将大模型部署于高校或教育主管部门自建的服务器上，实现“数据不出校园”。所有学生的交互数据在本地服务器上完成处理与存储，从根本上杜绝了数据跨境流动或被第三方商业平台滥用的风险，为遵守《个人信息保护法》和《数据安全法》提供了坚实的技术基础[11]。

在确保数据隐私的前提下，对 AI 辅导员的规制重点是对生成内容的精准控制。这需要通过以下两项关键技术实现：其一，是构建基于检索增强生成技术(Retrieval-Augmented Generation)的思政教育专项知识库。该知识库的构建需明确标准：在来源上，须严格收录马克思主义经典著作、国家政策文件、权威史学资料及校本规范等，确保信息准确；在结构上，需对语料进行切片、向量化处理并标注主题元数据，以支持精准检索。RAG 技术机制能确保 AI 辅导员的回答始终基于预设的权威知识片段进行生成，从而将输出内容“锚定”在安全、可靠的范围内，有效规避内容失真与事实性错误。其二，是实施系统性的提

示词工程(Prompt Engineering)。通过精心设计系统指令,为AI设定明确的角色与行为规范。例如,初始提示词应明确“你是一名中国高校的AI辅导员,需提供符合社会主义核心价值观的引导,对不确定或敏感问题引导咨询真人辅导员”等。此举能将教育要求编码至AI交互逻辑中,实现动态规制,确保输出内容的准确性与教育性。

4.2. 强化对大学生的人工智能素养教育

生成式人工智能的输出内容深受训练数据、价值预设与文化背景的影响。当前,社会对该技术的认知仍处于不断调适阶段,不同用户群体对其使用方式和伦理边界尚存在较大认知差异。因此,必须加强对大学生群体的思想引导和教育规范,帮助其树立科学的技术使用观念,坚守正确的价值观。

高等院校,尤其是设置了AI辅导员的高等院校必须重视大学生群体在人工智能时代的思想奠基工作,提升学生的生成式人工智能素养教育。通过开设《AI技术原理与局限性》《AI应用伦理与法律法规》等必修课程,帮助学生从技术本质上理解AI技术的双面性,避免陷入盲目依赖或滥用误区,使学生牢固树立科技向善的价值理念。同时,始终坚持基础知识与核心能力的培养,使学生既善于利用技术赋能学习,又葆有独立思考和价值判断的根本能力。要将人工智能伦理、信息鉴别力和正确价值观念有机融入思政课程,通过案例教学、情景模拟等多种方式,增强青少年对生成式内容的批判思维和自主判断能力,要不断强化学生的法治意识和道德自律能力,使其清晰认知技术使用的边界与禁区,自觉抵制虚假信息和错误思潮。

教育应坚持“守正”与“创新”相统一。在积极运用新技术的同时,绝不能放松对学生基本认知能力和价值观的塑造。这要求我们在人工智能赋能教育的过程中,始终把“育人”置于核心地位,实现技术工具与教育初心的同频共振。

4.3. 提高辅导员在思政教育工作中的主动性与价值引领力

AI辅导员的应用为高校辅导员的教育主体地位和教育效果带来挑战。因此,高校辅导员必须提高自身在思想政治教育工作中的主动性和价值引领力。

对于前者而言,辅导员需强化担当意识,积极提升技术素养,主动深入理解人工智能技术的运作逻辑与影响机制,在科学掌握人工智能技术的基础上,与AI辅导员共同开展思想政治教育。因此,AI辅导员并非完全替代传统辅导员,而是与其形成功能互补、深度融合的新格局。AI辅导员可高效完成政策宣导、信息解答、简单思想政治教育与心理健康教育等事务,而辅导员则应更专注于价值观塑造、情感沟通与危机干预等高阶育人工作,形成“人工智能初步响应-辅导员深度介入”的协同机制。该模式不仅显著减轻辅导员的事务性负担,使其更聚焦于思想引领、心理疏导和个性化育人,还可借助AI的数据分析能力提升思政工作和危机干预的精准性与实效性。

对于后者而言,辅导员还应不断增强话语的说服力与感染力。一方面,要以扎实的理论内容和鲜明的价值立场引导学生,使思想政治教育言之有物、有理有据,真正实现“理论掌握群众”;另一方面,应注重贴近学生实际,采用更具亲和力和时代感的话语方式,深入学生的思想世界与真实需求,在情感共鸣中实现价值传递和思想引领。此外,辅导员须主动扮演“信息守门人”角色,加强对AI生成内容的鉴别与筛选,确保教育内容的政治性和真实性。为此,应通过引进相关人才和加大培训力度,增强辅导员队伍对生成式人工智能的技术认知与批判能力,使其能够在人机协同中始终保持教育的主导性和判断力[12]。

5. 结语

AI辅导员的发展是教育数字化转型的重要体现,既带来了前所未有的便利,也带来了一些价值引领

风险。本研究通过剖析技术应用过程中潜藏的风险，以构建防范化解策略。未来 AI 辅导员的健康发展，必须始终坚持技术为育人服务的原则，在创新应用中持续完善风险治理机制，强化国产可控技术支撑，提升师生人工智能素养。唯有将育人初心贯穿于技术研发、部署与应用全流程，才能真正实现人工智能与思政教育的深度融合，为培养担当民族复兴大任的时代新人提供坚实保障。

参考文献

- [1] 陈宝生. ChatGPT: 教育的未来和未来的教育[J]. 人民教育, 2023(Z3): 6-13.
- [2] 唐志凤, 杨海云. 高校辅导员遇上 ChatGPT: 契机、危机及应对[J]. 教育探索, 2024(12): 33-38.
- [3] 中华人民共和国教育部. 华中科技大学构建 AI 赋能“三元互动”思政教育新范式[EB/OL]. 2025-06-09. http://www.moe.gov.cn/s78/A12/gongzuo/moe_2154/202506/t20250609_1193457.html, 2025-07-24.
- [4] 蒋莉薇, 李佩洁. 虚拟人在高校智慧思政体系中的应用探析[J]. 网络安全与数据治理, 2024, 43(5): 46-51.
- [5] 黄日涵, 高恩泽. “小院高墙”: 拜登政府的科技竞争战略[J]. 外交评论(外交学院学报), 2022, 39(2): 133-154+8.
- [6] 牛子牛. 生成式人工智能的意识形态风险与治理策略: 基于政治经济学批判方法的考察[J]. 人文杂志, 2025(1): 65-75.
- [7] 刘成, 莫生叶. 生成式人工智能应用中的意识形态安全边界厘清、风险监测及规制理路——基于 DeepSeek、Manus、ChatGPT、Sora 的思考[J]. 重庆大学学报(社会科学版), 2025, 31(6): 236-249.
- [8] 姚遥. 生成式人工智能融入爱国主义教育的维度、风险及优化路径[J]. 思想教育研究, 2025(6): 121-127.
- [9] 谢慧, 周建华. DeepSeek 与高校思想政治教育: 影响机理、运用风险及应对策略[J]. 湖南科技大学学报(社会科学版), 2025, 28(2): 104-111.
- [10] 李晓婷, 谢玉进. 从 ChatGPT 到 DeepSeek: 生成式人工智能介入思想政治教育的机理及其进路[J]. 当代教育论坛, 2025(5): 99-107.
- [11] 王润斯, 赵革, 胡晓龙. 基于 DeepSeek-RAG 的人工智能辅导员系统框架研究[J]. 南京开放大学学报, 2025(2): 1-6+39.
- [12] 曲建武, 刘越. 准确把握数字化时代高校辅导员工作的“变”与“不变” [J]. 思想政治教育研究, 2023, 39(6): 157-163.