

生成式人工智能的价值观风险及其对思政教育的启示

罗昭辉

南昌大学共青学院 江西九江 332020

摘要：随着生成式人工智能（AIGC）技术的迅猛发展，其在教育、媒体、社交等领域的广泛应用日益深入，但其潜在的价值观风险也逐渐显现。由于训练数据源于海量网络信息，生成式AI可能内嵌并放大偏见、歧视、历史虚无主义、极端主义等非主流或错误价值观，在内容生成过程中潜移默化地影响用户认知，尤其对价值观尚在形成阶段的青少年群体构成挑战。本文系统分析生成式人工智能在价值观层面存在的主要风险，包括价值中立性假象、算法偏见隐蔽传播、主流意识形态弱化以及认知依赖导致的主体性弱化等问题。在此基础上，探讨其对思想政治教育带来的冲击与启示，指出思政教育面临话语权威弱化、内容边界模糊、学生认知干扰等新挑战。研究提出，应主动应对技术变革，推动思政教育理念与方法创新：加强人工智能伦理教育，提升学生的算法批判意识；构建“人机协同”的教学新模式，强化教师的价值引导作用；开发具有正确价值导向的教育专用AI工具；完善网络内容治理机制，净化技术应用环境。通过技术治理与教育引导双轮驱动，增强思政教育的时代性与引领力，筑牢意识形态安全防线。

关键词：生成式人工智能；价值观风险；思想政治教育；算法偏见；意识形态安全

1. 引言

在教育场景中，生成式AI不仅改变了知识获取的方式，也深刻影响着学生的学习行为、思维方式与价值判断。特别是对于正处于价值观形成关键期的青少年群体，AI生成内容的然而，生成式人工智能并非价值中立的技术工具。其生成内容的质量与倾向高度依赖于训练数据、算法模型与开发者的价值预设。由于训练数据多来源于互联网，不可避免地包含偏见、歧视、错误信息甚至极端意识形态内容，导致AI在回答涉及政治制度、历史事件、社会伦理等问题时可能出现价值偏差或错误导向。例如，对中国特色社会主义制度的误解性描述、对历史人物的片面评价、对社会热点事件的煽动性解读等，均可能通过看似“客观”的AI输出影响用户认知。此外，算法的“黑箱”特性与用户对AI权威性的盲目信任，进一步加剧了价值观风险的隐蔽性与危害性。高频接触可能潜移默化地塑造其对社会、政治、道德等问题的认知框架。思想政治教育作为落实立德树人根本任务的关键课程，肩负着引导学生树立正确世界观、人生观、价值观的重要使命。在生成式AI日益普及的背景下，传统以教师主导、教材为核心、课堂为阵地的教育模式面临挑战：学生可能更倾向于向AI提问而非与教师对话；AI生成的“个性化”答案可能削弱主流价值的统一性与权威性；

碎片化、娱乐化的信息表达方式可能降低学生对严肃政治议题的关注与思考深度。深入研究生成式人工智能的价值观风险，不仅关乎技术伦理与信息安全，更直接关系到意识形态安全与人才培养方向。思政教育必须主动回应技术变革带来的挑战，既要警惕AI可能带来的价值侵蚀，也要善于利用其技术优势创新教育方式。

2. 生成式人工智能的价值观风险分析

2.1 价值观风险的内涵

价值观风险指在社会文化传播、信息交互及技术应用中，价值观念偏差、冲突或扭曲，对个体思想认知、群体价值取向和社会意识形态稳定产生负面影响的潜在威胁。在生成式人工智能场景下，表现为算法输出内容蕴含错误、片面或有害价值观念，误导用户尤其是青少年的价值判断与行为选择。其具有隐蔽性、传播性和复杂性特征，隐蔽性体现为错误价值观以合理信息形式呈现，不易察觉；传播性指借助互联网快速扩散；复杂性源于算法黑箱与多元价值观念碰撞。影响涵盖个体、群体和社会层面，个体易价值认知混乱、群体价值取向分化、社会主流意识形态安全受威胁。

2.2 生成式人工智能的发展现状

生成式人工智能从早期基于规则的简单模型发展为深

度学习驱动的复杂神经网络模型，以 Transformer 架构为基础的大语言模型如 GPT 系列、文心一言等取得重大突破，在文本、图像、代码等领域应用广泛。其应用领域多元，涉及内容创作、智能客服、教育、医疗、娱乐等，在内容创作中可快速生成稿件，教育中能提供个性化辅导。全球市场规模持续高速增长，吸引大量资本与企业。但发展面临数据偏见导致内容歧视、缺乏常识与逻辑错误、伦理道德争议、易被用于恶意信息生成等挑战，为价值观风险产生埋下隐患。

3. 生成式人工智能对思政教育的影响

生成式人工智能为思政教育注入强劲动力，从多个维度重塑教学形态与价值引导效能。在教学资源生成方面，其依托庞大数据库与强大运算能力，能够实时追踪社会热点，深度挖掘新闻事件、社会现象背后的思政元素，快速生成兼具时效性与教育性的案例分析、教学课件等素材。教师可借此将抽象的思政理论与鲜活的现实案例紧密结合，使课堂内容更丰富立体，显著提升学生的学习兴趣与参与度。例如，针对突发的重大社会事件，生成式人工智能可迅速整合事件全貌、各方观点及相关理论链接，助力教师开展深度教学，引导学生透过现象看本质，深化对主流价值观的理解。在个性化教育推进上，生成式人工智能发挥关键作用。通过对学生学习行为数据、知识掌握程度、价值观念倾向等多维度信息的精准分析，其能为每个学生量身定制专属学习计划。无论是学习进度的调整、知识巩固方式的选择，还是价值观培育重点的侧重，都更贴合学生个体需求。这种精准化的教育模式，有助于因材施教，充分挖掘学生潜力，促进学生在符合自身发展节奏的基础上，形成正确且稳固的价值观体系。

4. 思政教育的应对策略

4.1 强化价值观教育

为应对生成式人工智能带来的价值观风险，需将价值观教育摆在思政教育的核心位置。在课程体系设计上，深入挖掘各学科蕴含的思政元素，构建全员、全过程、全方位育人的课程格局，让学生在专业知识学习中潜移默化地接受主流价值观的熏陶。同时，针对人工智能输出内容可能存在价值观偏差，开展专题教育活动，引导学生运用马克思主义立场、观点和方法，对人工智能生成的信息进行批判性分析，培养学生独立思考和价值判断能力。此外，

注重榜样力量的引领，选取在人工智能领域坚守正确价值观、作出突出贡献的人物事迹，融入思政教育，激发学生对主流价值观的认同与践行，帮助学生树立坚定的理想信念和正确的价值取向，增强抵御不良价值观侵蚀的能力。

4.2 完善教育监管体系

构建全方位、多层次的教育监管体系是防范生成式人工智能价值观风险的重要保障。在技术层面，开发专门的内容审核工具，运用人工智能技术对教学中使用的生成式人工智能输出内容进行实时监测和筛查，及时发现并过滤包含错误价值观的信息。建立动态评估机制，定期对生成式人工智能的算法和数据进行审查，评估其对学生价值观形成的影响，一旦发现问题及时督促整改。在制度层面，制定严格的教育领域人工智能应用规范，明确平台、教师、学生等各方在使用生成式人工智能过程中的责任和义务，规范使用流程和标准。加强与相关部门的协同合作，形成教育、网信、市场监管等多部门联动的监管合力，共同维护健康的教育网络环境，确保生成式人工智能在思政教育中的应用符合主流价值观导向。

4.3 创新教育方式

面对生成式人工智能带来的教育变革，创新思政教育方式是提升教育实效性的关键。充分利用生成式人工智能的技术优势，打造智慧化思政课堂，通过虚拟现实、增强现实等技术，创设更加生动、真实的教学情境，让学生在沉浸式体验中深化对价值观的理解和感悟。鼓励教师转变教学角色，从知识的传授者转变为学习的引导者和促进者，利用人工智能辅助教学，开展个性化辅导和探究式学习活动，激发学生的学习主动性和创新思维。同时，积极探索线上线下混合式教学模式，整合优质网络教育资源与传统课堂教学，拓展思政教育的时空边界。利用社交媒体、学习平台等开展互动式、参与式的价值观教育活动，增强学生的参与感和认同感，使思政教育更加贴近学生生活，适应时代发展需求，有效应对生成式人工智能带来的挑战。

5. 国际比较与启示

5.1 发达国家生成式人工智能的发展现状

在技术研发领域，美国、英国、日本等发达国家凭借强大的科研实力和资金投入，处于全球生成式人工智能技术前沿。美国拥有 OpenAI、谷歌等顶尖科技企业，其研发的 GPT 系列、Bard 等大语言模型不断刷新生成式人工智能

的技术高度，在自然语言处理、图像生成等核心技术上持续突破。英国注重产学研结合，剑桥大学、牛津大学等高校与企业深度合作，推动生成式人工智能在医疗诊断、金融分析等专业领域的技术创新。日本则聚焦人工智能与制造业的融合，利用生成式人工智能优化生产流程、设计创新产品。在应用推广方面，发达国家将生成式人工智能广泛应用于教育、媒体、艺术等领域。如美国的教育科技公司利用生成式人工智能开发个性化学习系统，为学生定制学习方案；媒体行业借助其快速生成新闻稿件、视频内容；艺术领域则涌现出人工智能创作的绘画、音乐作品，极大地拓展了生成式人工智能的应用边界。同时，发达国家还通过政策支持和资金引导，积极构建人工智能产业生态，加速技术成果转化，进一步巩固其在生成式人工智能领域的领先地位。

5.2 发达国家思政教育的应对策略

面对生成式人工智能带来的价值观风险，发达国家在思政教育方面积极探索应对策略。在课程体系构建上，美国将人工智能伦理、数字素养等内容融入中小学和高校课程，通过开设专门课程或在相关学科中渗透，培养学生对人工智能生成内容的批判性思维和价值判断能力。德国在职业教育中强化职业道德与人工智能伦理教育，引导学生在未来工作中正确使用人工智能技术，避免价值观偏差。在监管机制完善方面，欧盟出台严格的数据隐私保护法规和人工智能伦理准则，对生成式人工智能在教育领域的应用进行规范，要求平台对输出内容进行审核，确保不传播有害价值观信息。同时，发达国家注重加强教师培训，提升教师运用人工智能开展思政教育的能力和应对价值观风险的素养。例如，英国定期组织教师参加人工智能教育应用与伦理培训课程，帮助教师掌握引导学生正确看待和使用生成式人工智能的方法，通过教师在课堂中的主导作用，有效抵御生成式人工智能带来的价值观冲击，保障思政教育的正确方向。

结论：

生成式人工智能以前所未有的速度融入社会各领域，在带来便捷与创新的同时，也引发了不容忽视的价值观风险。其输出内容中错误、多元的价值观念，如暗流般冲击着用户的认知体系，尤其对价值观尚未定型的受教育群体造成干扰，动摇着思政教育的价值引领根基。然而，风险与机遇并存，生成式人工智能也为思政教育革新提供了技术动能，在教学资源生成、个性化学习等方面展现出积极价值。未来，随着技术的持续迭代，思政教育需保持敏锐的洞察力与积极的应变力，持续深化理论研究与实践探索，在与生成式人工智能的良性互动中，不断提升育人实效，确保主流价值观在数字时代薪火相传、熠熠生辉。

参考文献：

- [1] 胡华. 智能思政：思想政治教育与人工智能的时代融合 [J]. 思想教育研究, 2022,(01):41–46.DOI:CNKI:SUN:SI XI.0.2022-01-007.
- [2] 林峰. 人工智能时代思想政治教育的价值定位与发展 [J]. 思想理论教育, 2020,(01):79–83.DOI:10.16075/j.cnki.cn31-1220/g4.2020.01.013.
- [3] 颜佳华, 李睿昊. 人工智能驱动的高校思政课教学模式创新论纲 [J]. 湘潭大学学报(哲学社会科学版), 2022,46(03):118–124.DOI:10.13715/j.cnki.jxupss.2022.03.003.
- [4] 李华君, 王沛佳. 人工智能时代高校思政课的智慧化教学创新与建构路径 [J]. 中国大学教学, 2021,(11):35–38.DOI:CNKI:SUN:JXCY.0.2021-11-007.
- [5] 王健, 郑旭东. 新时代信息化促进高校思想政治教育的思路、框架与建议 [J]. 电化教育研究, 2022,43(01):100–105. DOI:10.13811/j.cnki.eer.2022.01.013.

作者简介: 罗昭辉, 1994年9月, 男, 汉族, 江西赣州人, 南昌大学共青学院, 研究生学历, 助教, 思想政治教育

项目基金: 九江市社会科学基金纪念中国人民抗日战争暨世界反法西斯战争胜利 80 周年研究专项课题