生成式人工智能时代思政话语的意识形态风险及其引导机制

赵静怡

西南交通大学马克思主义学院 四川 成都 610000

【摘　要】：生成式人工智能通过语料训练与概率生成机制深度介入知识生产过程，推动思想政治教育话语生成方式由单一主体建构向人机协同模式转变。在这一过程中，由于训练语料结构和生成机制的影响，思政话语可能出现理论表达弱化、价值取向模糊以及语境表达抽象化等问题，从而引发一定的意识形态风险，相关问题主要与理论结构的生成方式、价值判断的对齐方式以及语境表达方式有关。为此，可以从数据内容优化、模型生成约束以及应用场景调整等方面构建协同引导路径，以提升思政话语生成的稳定性与有效性。

【关键词】：生成式人工智能；思政话语；意识形态风险；引导机制

DOI:10.12417/3041-0630.26.07.086

近年来，生成式人工智能（AIGC）的迅速发展，正在深刻改变人类的知识生产方式与话语传播格局，也为思想政治教育带来了新的机遇与挑战。随着教育数字化转型全面推进，人工智能已逐渐融入高校思政教学、学术研究与社会宣传等多个环节。习近平总书记在主持中央政治局第五次集体学习时指出，要深化教育数字化战略行动，加快建设全民终身学习的学习型社会、学习型大国，强调要充分运用数字技术推动教育高质量发展。[1]这一论述为人工智能技术与思想政治教育的融合提供了根本遵循，也为新时代思想政治教育创新指明了方向。但是，生成式人工智能与思政的融合可能会带来一定的意识形态风险。因此，本文从生成机制的角度出发，分析生成式人工智能参与思政话语生产所带来的影响，并重点探讨其在理论表达、价值导向和语言传播三个方面可能存在的问题及相应的引导路径，以期为思想政治教育的数字化转型提供参考。

1 生成式人工智能对思政话语生产方式的影响

生成式人工智能模型通过大规模语料训练和概率预测生成文本，使机器能够参与到信息表达和话语建构的过程中。在传统教育模式中，思政话语主要由教育者通过理论阐释和文本写作完成，而在人工智能环境中，文本生成逐渐呈现出“人机协同”的特点。生成式人工智能能够在较短时间内整合大量信息资源，提高文本生产效率，也为思政教育内容的拓展提供了新的技术条件。然而，从意识形态角度看，生成式人工智能并不是完全中性的技术工具。模型训练所依赖的语料来源、算法结构以及语义权重分配，都可能对生成文本的思想表达产生影响。马克思主义认为，意识形态的形成和传播始终与社会实践和社会关系相联系。在数字化时代，人工智能通过对既有语料的学习和重组，已经在一定程度上参与到话语再生产过程中，从而对思想表达的结构产生影响。这种影响并非单一维度展开，而是通过不同机制在话语生成过程中逐步体现。一方面，它可能改变理论表达的组织方式；另一方面，在多元语料融合的过程中，也可能对价值取向产生影响，同时还体现为语言表达在传播过程中的语境适配与情感呈现方式的改变。也就是说，生成式人工智能对思政话语的影响并不仅局限于某一环节，而是同时作用于理论表达、价值呈现和语言传播等多个层面。

2 生成式人工智能语境下思政话语生成的意识形态风险

生成式人工智能为思政话语生产提供了新的技术工具，但其生成逻辑也可能带来一系列意识形态风险。总体来看，这些风险主要体现在理论结构生成、价值对齐与语境传播三个方面。

2.1 理论结构生成中的碎片化风险

生成式人工智能通常以大规模预训练模型为核心，建立在 Transformer 架构及其自注意力机制之上，通过“预训练—微调—提示词引导”的流程，在海量未标注数据中学习语言与语义模式，进而构建概率生成模型，实现对高维语义信息的连续生成与相对可控的内容输出[2]，而不是对理论体系本身进行整体理解。因此，在处理复杂理论问题时，模型往往倾向于对已有语料中的常见表达进行组合和重组，从而形成结构完整但层次感不足的文本。在思政话语生成的情境中，这种情况表现为理论概念之间的联系更多建立在语言上的共现关系，而不是严格的逻辑推导上，容易出现概念并列、论证衔接不够紧密的问题。虽然生成的内容在形式上比较完整，但在理论阐释上往往缺乏清晰的递进关系和系统支撑，使得话语呈现出“结构完整但逻辑偏松”的特点。

作者简介：赵静怡（2002 年 9 月—），女，白族，云南大理，西南交通大学马克思主义学院，硕士研究生，研究方向：社会主义在中国的早期传播。

2.2 价值对齐中的中性化漂移风险

生成式人工智能在处理多来源语料时，并不具备独立的价值判断能力，其输出主要受训练数据分布和对齐机制的影响。当语料中存在不同价值立场时，这些立场可能在模型内部产生一定的竞争，而现有对齐方法（如基于人类反馈的强化学习）通常更倾向于生成冲突较小、风险较低的表达。因此，在涉及政治制度、意识形态或价值评价的问题上，模型往往会对不同观点进行一定程度的折中处理，输出看起来较为中性，但立场表达会相对弱化。这种“中性化”并不代表没有价值倾向，而是通过降低表达的立场强度来提高兼容性。这样一来，在思政话语生成过程中，如果缺少明确的外部引导机制，模型可能会对不同价值内容进行平均化整合，使原本具有明确理论指向的表述变得更偏解释性或描述性，从而在一定程度上影响价值导向的清晰性。

2.3 语境传播的去人文化表达风险

马克思指出，“人是人的最高本质”[3]。生成式人工智能虽然具备较强的语言组织能力，但其表达主要建立在抽象语义空间之上，对具体社会情境与个体经验的依赖程度较低。这种生成方式使文本更偏向于一般性、可迁移性的表达结构，而较少体现具体语境中的情感结构与经验细节。在思政话语生产的过程中容易出现表达内容脱离具体社会生活场景的问题，理论阐释更多地停留在抽象逻辑层面，而无法将其与现实经验之间进行有效连接。同时，由于模型难以主动构建具体的叙事情境，其生成内容往往以较为规范、概括性的表达为主，情感表达和叙事层次相对有限，在一定程度上会影响话语的感染力和解释效果。从传播角度来看，这些内容一旦进入实际传播过程，如果缺少必要的语境说明和情境补充，不同受众在理解时就可能出现偏差，甚至导致认知差异进一步扩大。

3 思政话语意识形态风险的生成引导机制

面对生成式人工智能对思政话语权带来的挑战，引导机制的关键不在于简单限制或排斥技术，而在于发挥思政教育主体的能动性，将算法技术纳入马克思主义意识形态的价值框架之中，使技术应用既能够发挥功能作用，也能够服务于正确的价值导向，从而推动思政话语在新的技术环境中实现更有效的发展。

3.1 以政治导向优化算法治理

防范意识形态风险，首先要把思想政治教育的目标更早地纳入人工智能技术的设计和运行过程中，改变把技术视为“完全中立”的认识，从根本上认识到生成式人工智能并不是简单的工具，在数据选择、模型训练以及内容生成规则等环节中，都可能包含一定的价值取向，从而进一步筑牢思政话语生成的“数字底座”。[4]一方面，应在生成式人工智能的算法设计和训练语料中有意识地融入社会主义核心价值观、中国式现代化等主流价值理念，通过整合“马工程”教材、红色文化资源以及中国社会发展实践案例等主权性数据资源，提升主流意识形态内容在数据结构中的基础性支撑作用，使其不仅作为补充性材料存在，而是成为模型理解与表达的重要知识背景。在此基础上，引导模型在生成过程中更多依托中国理论话语体系与本土经验逻辑进行组织表达，从而在一定程度上避免外部语料结构对价值判断形成潜在干扰。另一方面，在技术应用和教育实践中，教育者不应只停留在使用者的角色，还应适当参与模型使用规范和生成规则的设计。通过调整提示词结构、任务指令和生成条件，可以对内容生成过程进行一定的引导。例如，在涉及社会制度、价值评价或历史叙事等问题时，可以设置具有明确理论导向的提示语，引导模型从主流意识形态的分析视角进行回应，而不是简单拼合各种解释，从而减少生成内容在价值立场上的偏移。

3.2 重构“人本主导”的教学共同体

马克思指出：“人不是抽象的蛰居于世界之外的存在物，人就是人的世界，就是国家，社会。”[3]思想政治教育的主体是人，对象也是人，因此要在生成式人工智能赋能思政教学的过程中确立人本主导的地位。首先，教育者应逐步实现从“知识传递者”向“价值引领者”的角色转型。[5]在人工智能广泛介入教学活动的背景下，教师的功能不再仅仅体现为信息供给，而更多体现为价值判断与方向引导能力的强化。因此，教师不仅要具备基本的技术理解能力，也要能够对生成内容进行有效识别与判断，强化在教学过程中的审核意识与意识形态把关能力。在此基础上，应进一步发挥教师在情感引导和价值塑造方面的作用，使其情感表达与人工智能偏理性的输出形成互补，从而避免教学过程过度依赖或被算法逻辑单方面主导。其次，应进一步激活学生作为价值自觉主体的主动性。在智能技术深度嵌入学习过程的条件下，学生容易在便利的信息获取中弱化独立判断能力，因此有必要通过“问题驱动”的教学方式，引导学生在具体情境与真实问题中使用和审视人工智能工具，使其在持续的人机交互过程中形成批判性思维能力，而非被动接受由算法推送的既有结论。在强调技术应用效率与教学效果的同时，也需持续强化教育活动中的主体性意识，使个体始终被理解为具有自主判断与价值选择能力的主体，而非技术系统的被动承载者，从而防止教育对象在技术逻辑中被隐性遮蔽或弱化。

3.3 提升思政话语的现实回应能力

思政话语的生命力在于能够解释现实并回应现实问题，生成式人工智能的引导机制如果脱离这一基本逻辑，就容易在表达上形成技术化的抽象叙述，从而削弱其现实针对性与传播有效性。因此，在思政话语生成过程中，有必要克服算法文本可

能存在的脱离群众经验的“去人文化”倾向，使其重新回到具体社会实践与真实生活情境之中。首先，教育者应从单一的知识传授者转变为学习引导者、问题解决者、创新思维的激发者，[6]从受教育者的个体差异、认知水平以及现实关注点出发，对教学内容进行更具针对性的组织与表达，使不同学习者能够在差异化语境中获得相对适配的理解路径，从而在一定程度上提升思政话语的可接受性与内在吸引力，并激发其持续学习的兴趣与认同基础。其次，在教育内容层面，教育者应进一步强化理论与现实之间的联系，使思政话语能够更直接地嵌入社会生活实践之中。具体而言，可以引导生成式人工智能在文本生成过程中更多结合现实社会议题与具体发展情境，将抽象的理论逻辑转化为可感知、可理解的现实表达方式，使话语不仅具有规范性表达结构，同时也具备一定的生活经验基础与现实指向性，从而减少脱离语境的空泛化表达，增强其在受众中的亲和力与解释力。最后，在治理与优化机制层面，可以探索建立动态化的“学生思想反馈网络”，通过对学习过程中的互动数据与行为反馈进行持续收集与分析，在人工智能辅助下实现对生成内容的精细化标注与风险识别，提升学生对信息内容及其思想取向的理性辨识能力，使其具备一定的计算思维与反思意识，从而摆脱信息茧房的限制。同时，能够在唯物史观的视域下，对资本主义的当代表现形态进行审慎分析，进而形成对自由主义、历史虚无主义等错误思潮的理性批判能力。[7]

总体而言，生成式人工智能正在深入思想政治教育实践之中，这不仅是技术层面的变革，更是教育话语体系与价值生成方式的重塑。在这一进程中，如何在技术发展与价值引领之间保持协调，在效率提升与人文关怀之间实现平衡，关系到思想政治教育的根本方向与长远发展。面向未来，应在开放技术应用的同时，始终坚持正确的价值导向与教育立场，使技术真正服务于人的发展与价值塑造，推动思想政治教育在数字化进程中实现更具内在张力与解释力的转型与提升。

参考文献：

[1] 习近平.扎实推动教育强国建设[EB/OL].求是网,2023-09-15[2025-11-1].https://www.qstheory.cn/dukan/qs/2023-09/15/c_1129862386.htm.

[2] 胡飒,秦梦琪.生成式人工智能赋能高校思想政治教育的三重向度[J].思想教育研究,2025,(10):51-57.

[3] 马克思,恩格斯.马克思恩格斯选集(第一卷)[M].北京:人民出版社,2009.

[4] 王成伟,邓黎.生成式 AI 驱动高校思政课的潜在风险与规制路径[J].学校党建与思想教育,2026,(06):54-57.

[5] 李超民,张帆.生成式人工智能与思想政治教育主客体关系的重塑[J].吉首大学学报(社会科学版),2025,46(04):103-110.

[6] 胡刚.生成式人工智能时代高校思政课话语伦理风险及其治理路径[J].黑龙江高教研究,2025,43(09):8-15.

[7] 冯琳,倪国良.基于生成式人工智能的思想政治教育数字化转型[J].思想教育研究,2024,(02):46-53.