人工智能时代高校意识形态安全风险防控研究

陈　苗,吴秀娜,林露萍

(右江民族医学院,广西百色　533000)

[摘　要]进入人工智能时代,高校意识形态工作面对的风险更加不可预测、不可控,面临着价值偏见、信任和意识形态管理等潜在危机和风险。为有效防控这些风险,本文从技术、制度和社会管理三个方面探讨意识形态安全风险的成因及传播机制,提出风险防控措施。宏观层面上应从技术上建立智能防御体系、制度上进行治理现代化改革、社会协同与环境营造上构筑韧性生态圈三个方面入手;微观层面上应建立“评估—预警—策略”体系,形成精准、可自我优化的操作流水线。

[关键词]人工智能;高校意识形态安全;风险防控

[作者简介]陈苗(1991—),女,广西田阳人,右江民族医学院党委组织部组织科副科长,硕士,研究方向:党建、高校思想政治教育和宣传思想。吴秀娜(1987—),女,广西三江人,右江民族医学院党委宣传部(党委教师工作部)思政科科长,博士,研究方向:党建、高校思想政治教育和宣传思想。通信作者:林露萍(1989—),女,广西靖西人,右江民族医学院附属医院百东妇产科副主任医师,医学硕士,研究方向:女性生殖、孕产妇保健管理。

[基金项目]本文系 2025 年广西高校大学生思想政治教育理论与实践研究课题“人工智能赋能高校意识形态安全治理的路径与机制研究”(项目编号:2025SZ026)的研究成果。

[DOI]https://doi.org/10.62662/llyj0202021　　[中图分类号]G641

[本刊网址]www.oacj.net　　[投稿邮箱]llyj2025@163.com

意识形态工作是“为国家立心、为民族立魂”的根本性工作。随着人工智能技术的迅猛发展,高校意识形态安全领域正经历着前所未有的变革,深刻影响着高校师生的价值观念与思维模式,如不加预防控制可能给高校意识形态安全带来不可控风险,我们应当把握数字化、网络化、智能化发展大势,把创新作为第一动力、把安全作为底线要求、把普惠作为价值追求。因此,高校需深入研究人工智能时代高校意识形态安全产生的原因及传播机制,建立起应对该风险的防控机制。

一、人工智能时代高校意识形态安全风险的潜在危机

(一)价值偏见危机

人工智能时代,只有科学认识和准确把握人工智能的自身特点和内在规律,才能正确运用这一手段最终实现人的自由解放。但技术的快速发展最先改变的是高校师生的思维路径和价值取向,没有对其特点和内在规律的深刻把握,就会产生价值偏见。智能推荐算法重构了师生接触知识与思想的渠道,通过精准的数据分析与用户画像技术,定向推送特定内容,形成信息茧房效应。这种基于用户行为数据的个性化信息分发机制,虽然提升了信息匹配效率,但可能导致师生群体的认知视野窄化,进而催生思维定式与价值偏见。比如因数字鸿沟和算法歧视衍生的偏见排斥风险,因智能推送和数据投喂催生的认知撕裂风险,因“计算宣传”和数字霸权引发的舆论操纵风险,以及因调控失位和公共失序导致的主流价值解构风险等。高校作为多元思想交锋和主流价值观培育的场所,其传统教育所倡导的辩证思维、批判性思考与社会主义核心价值观的引领作用,可能在与这种高效、个性化却隐含偏见的智能信息流的竞争中逐渐被侵蚀,造成学生价值认知的碎片化与内在冲突,动摇意识形态认同的根基。

(二)高校信任危机

人工智能时代,网络空间的高度开放性使各种社会思潮迅速传播,但质量参差不齐的信息不仅会影响高校师生的价值判断,还对主流意识形态的权

威性构成挑战,产生对高校的信任危机。尤其是技术赋权与算法黑箱的矛盾决定的决策过程的不可解释性加剧了意识形态领域的分化风险、师生对技术权威的信任危机。比如智能评阅系统对学术论文的评分标准若缺乏透明度,可能间接影响师生对客观评价机制的认知,甚至滋生技术异化焦虑。此外,算法系统若因训练数据偏差导致价值判断偏差,可能在教学评估、课程推荐等环节传递隐性偏见,这种技术媒介中的意识形态渗透具有隐蔽性,容易被师生忽视,从而造成价值观引导的系统性偏差。

(三)意识形态管理危机

对于人工智能带来的意识形态安全风险的精准研判和有效处理是风险防控的重要措施。当前,数据采集与分析技术的应用正在重塑高校意识形态管理的生态结构。一方面,深度伪造的有违主流意识形态的信息内容在悄无声息地侵入高校师生信息网络,不加以精准研判和预防会进一步影响高校意识形态安全,甚至政治安全。另一方面,智能监控系统通过多模态数据采集技术,能够实时捕捉师生的行为特征与思想动态,这种深度介入虽有助于提升校园治理效能,但过度的数据挖掘可能侵蚀个人隐私边界,降低高校师生思想交流积极性。基于自然语言处理的舆情监测工具可能对特定关键词进行敏感识别,使师生在表达观点时产生顾虑,这种技术规训效应可能抑制思想自由交流,动摇高校作为思想活跃阵地的根基。另外数据安全漏洞还可能使师生信息被非法利用,为境外势力进行意识形态渗透提供可乘之机。

二、人工智能时代高校意识形态安全风险的成因与传播机制

(一)人工智能时代高校意识形态安全风险的成因

1. 技术上的不可控

人工智能时代高校意识形态安全风险的产生,根植于技术本身的内在特性及其应用过程中的不可控性。算法驱动的信息分发机制通过个性化推送强化了“信息茧房”效应,使师生容易局限于同质化、片面化的观点环境中,潜移默化地削弱了对多元价值和主流意识形态的全面认知。自然语言处理、深度合成等技术能够高效生成逼真的虚假信息或带有倾向性的内容,极大降低了意识形态渗透和认知操纵的门槛与成本。同时,人工智能系统的“黑箱”特性导致其决策过程难以被理解与审查,其训练数据中可能隐含的偏见或价值倾向会不自觉地被复制和放大,进而通过智慧课堂、学术工具、校园信息平台等渠道,以“技术中立”的外衣影响师生的判断。此外,网络空间的泛在连接与数据流动性,使得外部意识形态干预能够借助人工智能技术实现精准化、隐匿化和规模化,传统基于边界防护的安全体系在应对此类技术赋能的渗透时显得力不从心。

2. 制度上的滞后

现有高校治理体系与管理制度在应对人工智能变革时存在管理上的滞后。面向人工智能时代的新型意识形态安全风险,许多高校尚未建立起与之匹配的、系统性的预防、监测与应对制度。网络意识形态风险往往隐匿于日常的海量信息之中,网络舆情一旦爆发极易发生裂变。传统的内容审核与舆情管理方式,在速度和规模上难以匹配 AI 生成与传播信息的效率。且部门之间职责分割容易导致协同不畅,形成管理盲区。在制度规范上,针对校园内人工智能技术研发、采购、应用和评估的伦理审查与安全标准往往缺位或模糊,缺乏对算法价值观进行引导和约束的明确规则。同时,现有的思想政治教育与课程教学模式,在内容、方法与载体上未能充分融合数字时代的特点,对学生在人机交互环境中形成的认知习惯与价值困惑回应不足,导致主流意识形态教育的吸引力和实效性面临挑战,制度层面的“防御工事”尚未完成适应性重构。

3. 社会环境的催化

风险源于宏观社会环境对高校场域的深刻塑造。首先,它提供了意识形态风险滋生的温床与能量。技术资本主义的商业逻辑和平台经济的流量法则,构成了一个鼓励注意力争夺和情绪极化的外部信息生态。高校师生身处其中,其接收的信息议题、讨论的框架乃至思考的习惯,都不可避免地受到这一生态的塑造。这为特定意识形态内容的传播提供了持续的“社会心理燃料”和“议题弹药”。其次,它放大了技术与制度层面的脆弱性。社会思潮的多元碰撞和国际竞争的意识形态维度,如同不断变化的“气候条件”,对高校内部的技术应用和制度设计构成了持续的压力测试。一个在平静环境中尚可运行的系统,在这种动态、高压的“催化环境”下,其漏洞和滞后性会迅速暴露并被放大为实际风险。最后,它模糊了风险边界,增加了治理难

度。社会整体对人工智能的伦理焦虑、技术崇拜或批判态度,形成了复杂的社会心理环境。这使得校园内部的风险事件极易与外部社会情绪共振,瞬间突破校园物理或网络围墙,演变为公共舆情事件。反之,社会的热点争议也极速“倒灌”校园。这种内外风险的快速传导与相互催化,使得传统的、基于清晰边界的内控模式面临严峻挑战。

(二)人工智能时代高校意识形态安全风险传播机制

人工智能时代高校意识形态安全风险的传播机制,是一个由技术逻辑驱动、制度滞后加剧、并在社会环境中催化放大的复杂动态系统。其核心在于,人工智能不仅仅是一种传播工具,更是一种重塑信息生产、分发与接受全过程的结构性力量,它深度嵌入高校的学术、管理与生活空间,从而建构了一套全新的风险生成与扩散路径。

从起点看,风险的传播源发于数据与算法的价值内嵌与偏见扩散。人工智能系统基于海量、往往未经意识形态净化的数据进行训练,其算法设计亦可能隐含着开发者的文化预设或商业平台的流量逻辑。当这些技术应用于学术资源推荐、智能教学辅助、校园信息推送乃至管理决策时,便以一种“技术中立”的自动化方式,持续、隐蔽地向师生输出特定的认知框架和价值排序。这使得风险的源头从传统明确的主体转变为弥散化、非意图性的技术过程本身,风险的生产实现了自动化与规模化。

风险的扩散则依赖于平台化校园生态与制度反应的“速度差”。高校日益依赖一体化的数字平台和智能应用进行教学、科研与管理,这为风险的快速渗透提供了无缝衔接的通道。一项风险内容可以瞬间跨越课堂、社交平台、行政系统等传统边界,在师生社群中形成指数级传播。然而,高校现有的意识形态管理制度,无论是内容审核、舆情监测还是危机应对,大多建立在相对缓慢、依靠人工判断、分部门处理的传统模式之上。这种技术传播的高速性、跨界性与管理制度的分割性、滞后性之间形成的巨大“速度差”与“覆盖盲区”,成为风险得以迅速蔓延和深化的重要条件。

最终,风险的生效与深化在人机互动与认知建构的社会心理层面完成,并受外部社会环境的强烈催化。风险的真正影响,发生在师生作为信息接收者与算法进行持续互动的过程中。人工智能提供的个性化、情感化、高沉浸度的信息环境,深刻影响着个体的认知习惯与判断力,可能削弱批判性思维和系统性的价值认同。而当校园内部的风险议题与外部社会矛盾、国际意识形态博弈相互勾连时,便会发生强烈的“共振效应”。外部社会环境包括平台的资本逻辑、社会的焦虑情绪、国际的斗争态势等如同催化剂,急剧放大校园内部风险的能级与破坏力,使其从局部管理问题升级为冲击高校信任根基和价值引领功能的系统性危机。整个传播机制因而呈现出一种内生性技术风险与社会性催化压力相互增强、循环反馈的复杂图景。

三、人工智能时代高校意识形态安全风险防控措施

(一)宏观层面

1. 技术上建立智能防御体系

高校应构建主动且内嵌价值观的智能防御体系,这要求超越被动封堵的旧有模式,转向对人工智能技术全链条的深度介入与引导。一方面,应自主研发或审慎引入具备意识形态风险感知能力的 AI 监管工具,实现对校园网络平台、学术数据库、智慧教学系统中多模态信息的实时扫描与智能研判,尤其加强对深度伪造内容与隐蔽性叙事框架的识别。通过算法模型的精准识别与动态学习,风险 预警的时效性可以得到提升,同时又能深度解析意识形态表达的隐性特征,这能充分发挥人工智 能在高校网络意识形态治理中的实践潜力。另一方面,应推动算法透明与问责,对校内应用的推荐算法建立伦理审查与定期评估机制,通过设置“主流价值参数”和干预“过滤气泡”效应,确保信息分发的多元与平衡。更重要的是,需利用区块链等技术为权威信息提供可追溯的认证,并建立动态的舆情模拟与预警系统,从而将技术从风险放大器转化为安全稳定器,在信息流通的源头与过程中植入免疫基因。

2. 制度上进行治理现代化改革

针对当前人工智能迅速发展的趋势,高校亟需进行一场系统性的治理现代化改革,以跨部门协同与全流程管理应对风险的弥散性。核心是建立一个整合宣传、网络信息、教务、科研、学生工作及国际交流等职能的“意识形态安全综合治理中心”,打破数据与行政壁垒,实现风险研判、决策与处置的一体化联动。引导高校师生突破“互联网仅是工具”的认知,使其认识到网络的政治属性和网络行为的价值负载,不再成为缺乏独立判断能力的信息“搬运工”。首先,在制度设计上,应通过制定规则

明确在科研攻关、教学应用、服务管理中各主体的责任边界，并将意识形态安全评估设为相关技术项目立项、采购与验收的必经环节，如我国《生成式人工智能服务管理暂行办法》明确要求，提供和使用生成式人工智能服务要坚持社会主义核心价值观。其次，必须将风险防控深度融入育人主渠道，一方面，革新思想政治教育的内容与方式，开发聚焦“人工智能与社会”“数字伦理与价值判断”的课程模块，提升师生的数字公民素养与批判性思维能力；另一方面，将教师的专项培训常态化，使其成为校园风险治理的前哨与引导者。通过以上制度化、常态化的设计，构建起权责清晰、响应迅速、教育前置的刚性防线。

3. 社会协同与环境营造上构筑韧性生态圈

高校网络意识形态治理是一项复杂的系统工程，要做好这项工作，需要统筹建设好校内科研、教学、管理治理平台，合理调动整合跨学科治理资源。因此应主动突破“象牙塔”边界，在更广阔的社会与技术生态中寻求合作、引导共识，构筑韧性生态圈。一方面，需与负责任的科技企业、研究机构建立合作，共同研发安全可控的教育技术产品，并参与行业标准的制定，从产业上游施加积极影响。另一方面，应积极搭建公共对话平台，组织专家学者向社会解读人工智能的伦理与安全议题，澄清公众误解，引导理性认知，营造有利于技术向善的社会氛围。在国际维度，高校应在坚持自主可控的前提下，深化有安全底线的国际学术交流，主动参与全球科技伦理对话，贡献中国高校的治理智慧，并在此过程中提升师生在国际舞台辨析与抵御意识形态渗透的能力。最终，通过联结产业、引导公众、对话世界，高校能够将自身的安全治理体系锚定于一个健康、清朗且具有支持性的宏观社会生态之中，实现内控与外联的辩证统一。

（二）微观层面

宏观层面的“技术—制度—社会”协同为高校意识形态安全风险防控提供了立体的能力域和资源库，但仍需要高校建立微观层面的“评估—预警—策略”体系，构成一个精准、可自我优化的操作流水线。二者相辅相成才能构建起一套既高瞻远瞩又落地生根的人工智能时代意识形态安全风险综合防控体系。

首先，风险评估指标体系的设立是整个防控工作的科学基石与感知器官。它需要将宏观层面的风险成因转化为一系列可量化、可监测的具体指标。在技术层面，可设立校内主流信息平台中 AI 推荐内容的意识形态一致性指数、深度伪造内容校内传播峰值与响应时间；在制度层面，可设立跨部门风险会商频率与决议执行率、师生数字伦理课程覆盖率与效果评估值；在社会环境层面，可监测关联社会热点议题在校园社群中的情感极化度、国际学术合作中涉及敏感技术领域的频次与合规性审查强度。这套多维度、细颗粒度的指标体系，使得原本模糊的风险得以被具象化感知和常态化扫描。

其次，风险预警机制的设计是连接感知与行动的中枢神经与决策支撑。它基于风险评估指标体系所采集的动态数据，通过建立风险模型、设定阈值和构建响应预案，实现从监测到预警的质变。当特定非主流话题的校内网络声量指数与校外关联社媒同步热议指数同时突破阈值，且内容情感分析模型显示对抗性情绪快速上升时，预警机制应自动触发，向综合治理中心发送分级警报，并同步推送初步的关联数据分析和历史类似案例处置参考。这一机制的核心在于利用 AI 和大数据技术，实现从海量信息中自动识别风险模式，将事后处置前置为事中干预甚至事前预测，极大地缩短了决策响应时间。

最后，风险防控策略的制定与执行是防控闭环的末端输出与效能检验环节。它直接对应于预警机制发出的指令，并根据风险评估所揭示的风险性质与根源，调用和组合宏观层面所构建的各种资源与能力。针对一起由 AI 生成虚假学术信息引发的意识形态争议，防控策略可能立即启动一个组合包，技术层面，指令安全工具溯源并限流该信息，同时推送权威澄清信息；制度层面，依据预案，由宣传部门牵头发布官方说明，教务部门通知相关师生，同时启动针对此类新型虚假信息的制度修订调研；社会协同层面，协调合作媒体进行事实报道，并计划在下期科技伦理工作坊中加入该案例教学。所有策略的执行效果，又会作为新的数据反馈回风险评估指标体系，从而开启下一个循环的优化与迭代。

参考文献：

[1]习近平. 习近平著作选读(第一卷)[M]. 北京：人民出版社，2023.

[2]习近平向 2024 年世界互联网大会乌镇峰会开幕视频致贺[N]. 人民日报，2024-11-21.

[3]廖军，景星维. 人工智能时代高校意识形态风险防控

能力研究[J]. 西南交通大学学报(社会科学版),2026,27(1):108-122.

[4]余杰. 人工智能时代的意识形态风险及其化解[J]. 思想理论教育,2022(12):84-89.

[5]杨杰. 人工智能时代高校网络意识形态安全风险及其防范研究[J]. 传播与版权,2025(17):86-89.

[6]陈海萍. 扎根理论视域下高校网络舆情危机应对探究[J]. 高校辅导员学刊,2025,17(3):69-76,99.

[7]邱丹. 人工智能工具赋能高校教师网络意识形态安全建设[J]. 数字技术与应用,2025,43(12):29-31.

[8]王方,王楠. 网络时代高校思想政治教育对象的特征与启示[J]. 高校辅导员学刊,2021,13(4):26-30.

[9]生成式人工智能服务管理暂行办法[EB/OL]. https://www.gov.cn/zhengce/zhengceku/202307/content_6891752.htm,2025-4-30.

[10]曹静,杨正宇,喻娟. 人工智能时代高校网络意识形态治理效能提升研究[J]. 东华大学学报(社会科学版),2025,25(4):5-7.

Research on Preventing and Controlling Ideological Security Risks in Colleges and Universities in the Age of Artificial Intelligence

CHEN Miao, WU Xiu-na, LIN Lu-ping

(Youjiang University of Ethnic Medicine, Baise Guangxi 533000, China)

Abstract: In the age of artificial intelligence, ideological work in universities faces increasingly unpredictable and uncontrollable risks, including potential crises and risks such as value bias, trust issues, and ideological management challenges. To effectively prevent and control these risks, this paper explores the causes and transmission mechanisms of ideological security risks from three aspects: technology, system, and social management, and proposes risk prevention and control measures. At the macro level, efforts should be made to establish an intelligent defense system technically, carry out modernization reforms in system governance, and construct a resilient ecosystem in social collaboration and environmental creation. At the micro level, an "evaluation-warning-strategy" system should be established to form a precise and self-optimizable operational pipeline.

Key words: artificial intelligence; ideological security in universities; risk prevention and control