DOI:10.3876/j.issn.1671-4970.2026.05.005

生成式人工智能嵌入政府环境治理的逻辑与规制

秦　鹏,周继钊

(重庆大学法学院)

摘要:生成式人工智能的发展为政府环境治理提供了独特的技术支撑,前者能够迅速整理环境数据,以"数字客服官"的身份实现类人化表达,并辅助工作人员作出环境决策。生成式人工智能深度嵌入政府环境治理,在推动治理主体、治理过程、治理工具发生逻辑转变的同时,也暴露了阶段性缺陷:前期准备阶段存在价值和数据管理缺陷,实际运行阶段存在算法歧视和"算法黑箱"缺陷,后续处理阶段存在责任分配缺陷。敏捷治理理论的引入,使相应的规制路径更具全面性、适应性和灵活性。对此,应确立符合全面性的主体自觉、法律与技术协同共治、公平的基本原则,构建契合适应性的事前预防和事后应对相结合的适应性治理机制,创新符合灵活性的过程监管机制。其中,适应性治理机制中的事前预防层面应分为三层次:第一层次建立数据采集、验证、存储和保护制度,第二层次建立生成式人工智能应用清单制度,第三层次建立算法解释制度。事后应对层面,应建立生成式人工智能嵌入环境治理后的责任分配制度。过程监管机制则应包含三方面:一是建立生成式人工智能的监督、协同、问责部门,二是建立生成式人工智能嵌入环境治理的伦理验证、评估许可及投入使用的审计制度,三是建立社会组织与公众协同监管制度。由此能持续完善生成式人工智能嵌入环境治理的规制方案,以推动政府智慧环境治理目标的有效落地。

关键词:生成式人工智能;数字政府;环境治理;"算法黑箱";敏捷治理

当前,以人工智能技术为代表的新型科技迅猛发展,并被广泛应用于政府环境治理等行政治理场域。2023年印发的《数字中国建设整体布局规划》提出,要建设绿色智慧的数字生态文明,持续推进生态环境的智慧治理,发展高效创新的数字政务。同年,习近平总书记指出:"我国生态环境保护结构性、根源性、趋势性压力尚未根本缓解。我国经济社会发展已进入加快绿色化、低碳化的高质量发展阶段,生态文明建设仍处于压力叠加、负重前行的关键期。"[1]为此,习近平总书记着重强调:"深化人工智能等数字技术应用,构建美丽中国数字化治理体系,建设绿色智慧的数字生态文明。"[2]

生成式人工智能(Generative AI)的技术突破与迭代演进是政府提升环境治理能力的关键契机,但如何进一步理解并把握其与政府环境治理的互动融合关系,是环境治理领域必须正视的现实问题。生成式人工智能是指基于大语言模型(LLM),通过机器学习的方法,深度汲取数据并从中学习对象的特征,进而输出全新

引用本文:秦鹏,周继钊. 生成式人工智能嵌入政府环境治理的逻辑与规制[J]. 河海大学学报(哲学社会科学版),2026,28(5):41-53.

基金项目:重庆市社会科学规划重大项目(2024ZXZD019)

作者简介:秦鹏(1969—),男,教授,博士,主要从事环境与资源保护法研究。E-mail:qp6909@126.com

的、符合客户需求的、具备原创性的多模态内容(如文字、图片、声音、影像等)的技术形态,其具有超强的自然语言理解能力、多模态转变能力及知识内容生成能力[3]。在环境治理领域,生成式人工智能可对数据库中的各种语料进行深度学习,模拟多元环境治理场景,承担相应的文本处理工作;能够有效化解政府环境治理面临的治理成本攀升、治理风险类型增加、治理技术支撑不足及治理效果不佳等困境,提升智慧环境治理的水平和政府环境治理的效能。在智慧环境治理方面,算法嵌入政府环境治理能够提升政府环境治理的效率与精确性[4]。围绕生成式人工智能在政府治理中的应用问题,诸多研究成果指出,人机混合决策能够有效提供包括环境决策在内的政府决策的精度和准确度[5];ChatGPT 等生成式人工智能产品打破了传统政府行政服务模式中对话不对称的现状[6]。在敏捷治理方面,有研究将敏捷治理理论与生成式人工智能赋能政府治理相结合[7]。总体来看,将生成式人工智能嵌入环境治理的研究仍未引起学界的足够关注,且由于环境治理的对象、方式和目标与其他政府治理领域存在较大差异,因此有必要针对生成式人工智能嵌入政府环境治理展开深入研究,深挖生成式人工智能嵌入环境治理的优势和效能,在全盘考量相关风险的基础上,从敏捷治理的视角提出相应的规制方案。

一、生成式人工智能技术嵌入政府环境治理的可能空间

生成式人工智能技术具备学习、模拟和场景处理等能力,为其深度嵌入政府环境治理提供了可能。生成式人工智能的良好运用依赖真实、海量的环境数据喂养和场景的适配性,其技术迭代快、不确定性高等特性与环境治理过程中存在的科层架构、属地保护、程序刚性的体制特征之间存在一定张力,传统架构式、导向型的政府治理范式难以缓解这种张力。为此,亟须一种动态调适、能响应不确定性的治理框架。近年来兴起的敏捷治理框架恰好符合该要求,其能有效弥补并更好地助力生成式人工智能技术嵌入政府环境治理。敏捷治理理论由软件开发领域的“敏捷开发”方法演变而来,既有研究曾将敏捷治理引入数字政府[8]、新兴产业[9]、现代城市[10]等新兴技术或被新兴技术所渗透的领域,上述领域的共同特征是技术发展迅速,而传统治理框架的反应周期与技术发展速度不相匹配。同样地,环境治理也存在类似困境,该领域的突发性和不确定性与僵化程序之间的张力使得传统的“汇报-反应-执行”模式难以形成有效的“反应-问题解决”闭环。已有学者开始关注数字技术嵌入农村环境敏捷治理的实践逻辑,且已表明敏捷治理向环境治理领域的嵌入并非简单拼凑,而是需要回应环境治理模式发展需求的自然演进[11]。

1. 环境数据整理的迅捷便利

数据的收集、处理与再利用是数字政府建设的关键步骤,政府环境治理若要发挥成效,首要任务便是收集并处理各类环境数据。目前,环境数据收集的难点有二:一是收集难度高,因缺乏技术手段和行政资源支撑,政府难以在环境治理的重点、难点区域完成数据收集;二是收集涉及的政府部门繁杂,难以实现迅速且无障碍的数据共享。生成式人工智能在数据管理方面具有独特优势,能有效解决上述难题。其一,生成式人工智能可理解并分类大量环境数据,具有数据检索与组合优势,能够提升数据管理系统的效能。如深圳市政府数据开放平台专设人工智能语料数据开放专题,将语料数据分为预训练文本类、标注文本类、音频视频类、指令集图片类等类型,生成式人工智能可根据内容生成需要,汲取不同种类数据(如植物类数据集用于植物分类模型训练、植物分布及多样性研究等),从而针对性化解环境数据种类杂乱、不易整合等困境。生成式人工智能还能够通过全天候、自动化运行搭建动态数据整合机制,对气象变化、污染监测等关键环境数据进行实时收集、分类和整合。其二,生成式人工智能能够突破既有的数据检索“平台化”模式,改变数据垄断、层次分明的现状。如工作人员可实时查看环境数据而无需烦琐的申请程序,如此大大降低了个人获取数据门槛。其三,生成式人工智能可作为“数字分理员”,帮助完成数据分流,打通环境治理领域各业务部门数据壁垒,实

现环境数据横向和纵向的传递[12]。

值得注意的是,技术的应用不是盲目的,而是循序渐进、相互融合的。生成式人工智能的嵌入须逐步契合环境治理的技术需求,而敏捷治理的全面性、适应性和灵活性特征可以促进二者结合的实现。从全面性来看,敏捷治理主张多元参与,可调动政府以外的社会技术力量共同参与生成式人工智能训练数据的采集和共享;从适应性来看,敏捷治理强调场景适配,支持生成式人工智能根据现实场景使用不同语言与多国工作人员及时沟通,实现跨国和跨区域的环境协作治理;从灵活性来看,敏捷治理倡导治理方式弹性调整,不同区域的工作人员在将数据分类标准和数量传输给生成式人工智能时,可灵活地选择多种传输途径(如监测设备联网自动传输、手动输入等)。然而,环境数据来源繁杂,真假难辨,对生成式人工智能输出结果真实性影响巨大。因此,为保障数据的真实性和可靠性,须依托敏捷治理理论建立兼具全面性、适应性和灵活性的数据规范体系。

2. 连续性、类人化与专业化的功能展现

在传统政务服务中,政府往往须聘请相应领域的专家面向公众系统化讲解环境治理领域的复杂性、专业性问题,因此需承担较高的人力资源成本。生成式人工智能的人机互动特质能有效解决这一问题:其一,生成式人工智能可利用自然语言完成连续对话[13]。如重庆市人民政府官网开设智能问答平台,能较好地回复公众有关生态环境法典条文解释的信息咨询[14]。其二,生成式人工智能能够作类人化表达。经训练,生成式人工智能能够掌握人类自然语言的词语搭配和语气,且素材增加后,还能逐渐掌握人类在各种场景下的情绪变化,完成类人化交谈。如针对不熟悉环境信息公开制度的公众,生成式人工智能可用平实易懂的语言讲解公开范围、申请渠道与办理要求。其三,生成式人工智能学习能力强,可吸纳并输出专业知识。环境治理知识晦涩艰深,具备相应专业能力的从业人员数量有限,如环境信用评价领域能掌握环境负债、环境成本、污染防治等多维度专业知识的专业人才储备紧张,经过训练的生成式人工智能可给出较为专业化的回复,从而在一定程度上化解环境领域人才不足的困境。

可见,生成式人工智能已经具备承担环境政务咨询解答任务的能力,可承担类似政务客服的职能,成为面向社会公众提供环境政务咨询的“数字客服官”。但是,生成式人工智能仍须具备以下条件,才能胜任这一角色:一方面是保障生成内容的针对性、权威性和准确性。生成式人工智能在与公众进行一对一交互问答的过程中,应对所提供答案给出明确说明,涉及相关政策和法规的,须提供原文,以保障其作为政府“数字客服官”的权威性。对此,环境监管部门还须建立严格的内容审核机制,确保内容的准确性和针对性。另一方面是实现全过程服务闭环。生成式人工智能不仅要为办事群众提供答案,还应生成全过程的办事引导服务,帮助群众解决办事难题。

通过引入兼具全面性、适应性和灵活性的敏捷治理理论,生成式人工智能的专业化、类人化、连续性的功能优势得以更好地发挥。一是从全面性看,敏捷治理要求治理覆盖环境政务全流程,这为生成式人工智能的专业化功能提供了发挥空间。生成式人工智能可依靠知识汲取和规则嵌入功能进一步开拓数字客服应用场景,如提供线上排污许可申报等简易业务服务,从而实现环境业务流程的规范化办理和全过程自动化引导。二是从适应性看,敏捷治理要求对公众的多层次和差异化要求做到动态响应,这为生成式人工智能的类人化功能提供了施展余地。类人化功能通过多轮对话理解公众意图,使得动态响应更具针对性。生成式人工智能可依靠多轮对话与自然语言理解功能应对公众的各类提问,以提升政务服务的亲民性和及时性。三是从灵活性看,敏捷治理要求治理方式随治理场景变化弹性调整,这需要生成式人工智能的连续性功能加以支撑。适应性侧重对单一场景内需求变化的即时回应,灵活性则强调跨场景切换中治理手段的调整,而连续性功能的发挥则能通过跨场景追踪、连贯分析,防止因场景更替导致信息重置和服务中断,确保服务衔接一致。在此基础上,面对具体诉求的响应时,生成式人工智能通过推理分析功能对公众诉求进行文义拆解和逻辑辨析,从而精准地

满足公众诉求。因此,敏捷治理理论与生成式人工智能技术功能的耦合,不仅能拓宽数字客服的适用边界,亦能在功能调适过程中提升环境服务的响应效率和治理效能。

3. 环境决策的科学支撑

受主体专业性不足、专业环境数据繁杂庞大等因素制约,人类决策者往往难以获取并处理充足的有效信息,进而影响决策质量。基于全方位信息,生成式人工智能可辅助环境工作人员应对复杂环境问题,及时作出合理判断,避免随意行政,从而提升环境决策的科学性[15]。

首先,生成式人工智能可通过海量数据深度学习,将人类思维、偏好、逻辑、价值观等以代码的形式嵌入[16],借助较强的语言理解能力,识别环境工作者的意图与需求。其次,生成式人工智能擅长语言理解和信息汲取,可通过人机交互为环境工作者提供决策支持。如通过自然语言处理和自主学习,其可从海量环境数据中提取有助于解决环境问题的关键信息,为专业知识储备不足的环境工作者提供决策参考。最后,生成式人工智能可借助信息分析进行建模,预测工业、建筑等项目的环境影响,而该功能可用于环境影响评价领域,能提高环境影响评价的科学化、高效化和标准化水平。

进一步地,生成式人工智能的数据来源具有多源性,而敏捷治理的全面性强调治理覆盖多领域、多主体,二者的结合有助于满足环境监管部门的多种治理需求。如,生成式人工智能可辅助工作人员快速认知核污染、化石能源污染等陌生治理领域,从而制定合理的应对方案。为避免技术辅助演化为技术依赖,生成式人工智能的规范应用还要遵从以下三点要求:一是提升算法的可解释性和可追溯性。环境决策的算法生成链条应当做到可追溯和可验证。二是人类应成为决策的最终主导者。环境治理涉及重大利益,生成式人工智能仅可提供较为合理的备选方案,但难以对伦理、发展等多要素进行权衡。为此,人类须综合考量生态环境、经济发展、社会道德等因素,选择最合适的环境治理方案。三是明确决策错误的责任分配。生成式人工智能会因数据、算法等问题生成错误的内容,对此,须构建明确的责任分配制度。

二、生成式人工智能嵌入政府环境治理的逻辑转变

1. 主体逻辑转变:环境监管部门主导至多元协同合作

受环境治理强制性与单向性影响,传统环境治理曾长期以“命令-控制”模式为主,环境治理被视作环境监管部门的专属职能,后者是环境治理的核心主体。在该模式下,环境监管部门以环境风险的有效管制和防控为治理目标,企业、社会组织与公众作为被管制和被防控的对象,处于治理的边缘。这种治理格局使治理责任集中于环境监管部门,其他主体缺乏参与渠道,不仅割裂了部门间的联动关系,也阻塞了政府和其他社会主体的合作通道。

生成式人工智能的应用改变了既往的环境治理形态:凭借信息处理、内容生成和平台协作能力,其为多元主体协同治理提供了条件。环境信息的充分共享和利用是多元主体高效协作的重要前提,生成式人工智能可广泛收集整理环境治理信息,对信息作分级分类整合后呈现给治理者,有助于缓解环境治理中的信息不对称问题。部门利益冲突是制约环境治理多元协同的关键因素,生成式人工智能凭借人机交互与多模态内容生成技术,通过结构化对话机制系统归集各部门的差异化诉求,生成兼具包容性与协调性的权责配置方案,为化解利益分歧引发的协作障碍提供支持,但制度层面的利益协调仍需依赖常规行政过程推进。进一步地,生成式人工智能能够突破物理空间束缚,搭建虚拟的网络空间,为环境治理对话提供跨地域的公共空间,并可保存对话内容,通过信息沉淀推动环境公共话题的长久讨论。该对话平台也可作为社会公众发表意见、提出建议的渠道,有利于政府及时了解公众的环境疑问与诉求,并予以回应[17]。

2. 过程逻辑转变:“被动-末端”治理至“主动-全过程”治理

在“以经济发展为先”的政策导向下,政府环境治理曾长期存在“先污染,后治理”的倾向,较多采用“被动-末端”治理策略,随着技术发展,“被动-末端”治理逐渐展现出劣势:一是

污染后果的修复成本高昂。“被动-末端”治理是对已发生的环境污染进行事后治理,而生态修复成本通常居高不下。以2015年天津港“8.12”重大爆炸事故为例,事故暴露出危险品仓储监管的多环节缺位,大量化学危险品泄漏造成了周围土壤、空气的污染,该事故虽属安全生产事故,但后期生态修复成本却远高于事前预防投入。二是污染处理不彻底易引发二次污染。当前,污染处理技术难以完全消除污染物,治理后的残留物极易造成二次污染。以污水处理为例,处理污水时产生的污泥中富集病原体、重金属等污染物,而污泥处置费用高昂,若监管缺位,污泥未得到妥善处理,其中的污染物进入环境,易造成二次污染。三是末端监管难以覆盖企业生产全过程。即便法规明确要求企业配置污染处理设备,但受限于生产规模和设备的技术水平或环保设施建而不用,生产环节的污染难以彻底清除。以家纺行业为例,某些家纺企业印染车间的废气处理设施未达验收标准即投入使用,废气收集管道存在旁路直排,致使废气未经有效处理就被排放。

“主动-全过程”治理将生产的各阶段纳入环境监管,能够使预防、监管、发现、修复等环节紧密连接,实现环境的有效治理,而生成式人工智能的应用加快了环境治理由“被动-末端”治理迈向“主动-全过程”治理的进程。这一进程可分为事前预防、事中监管和事后救济三阶段:一是在事前预防阶段,生成式人工智能凭借自然语言理解能力和信息搜集能力,助力环境数据整理。环境风险预测有赖于信息搜集和多元主体沟通,信息掌握得越全面,预测结论就越精准。生成式人工智能能够采集公众反馈信息,筛选环境风险相关内容并整理成文,为决策提供参考,其信息搜集的广泛性和迅捷性有助于提升风险预防的效率和科学性。风险预防能够减少环境污染事件的发生概率,从而降低生态修复成本。二是在事中监管阶段,生成式人工智能可辅助监管人员全天候回收、整理并分析环境风险信息,并对已经发生或将要发生的环境风险向治理者发出警示,有助于督促企业在生产环节落实污染处理要求。三是在事后救济阶段,生成式人工智能能够依托知识生成能力,快速生成问题解决方案。在环境风险发生后,救济方案是否精准高效直接影响治理成效,生成式人工智能可在自主学习既有环境治理知识的基础上,根据治理者提供的相关信息,提出具有针对性的治理方案,提升环境风险救济效率[18],由此能够帮助环境工作人员选择适配当前环境危机的解决方案,在一定程度上降低因治理技术落后导致二次污染发生的概率。

3. 工具逻辑转变:模糊治理至精细化治理

人工智能技术的发展推动环境治理者转变思维、更新方法,促进治理范式从经验驱动向数据驱动变革。以往,环境治理者多依靠经验和直觉作出决策,而生成式人工智能类工具的嵌入可实现数据实时采集,为决策方式从经验驱动向数据驱动提供条件,从而满足环境治理精细化、个性化、区域化和动态化的需求。

生成式人工智能嵌入环境精细化治理,须做到数据精准筛选和环境问题精准施策:一方面,环境数据的精准识别与分类是开展精准治理的前提。传统人工识别分类对信息完整度要求高、主观性强、成本高、标准不统一,往往难以支撑精准治理,政府不得不较多依赖模糊治理,从而加剧了环境治理中不充分、不平衡的矛盾。另一方面,精准施策应以治理对象的需求为导向,依据治理对象的反馈和需求制定方案。环境治理囊括的行业、区域、个体存在差异,治理对象的需求呈现复杂性、多样性、弥散性等特征,需求分散且差异大,经验判断难以逐一回应,工作人员往往较多依赖经验判断,难以实现精准施策。自适应性是生成式人工智能的基本技术特征,生成式人工智能可根据对系统和环境的感知,自主调整生成策略和决策倾向,以回应分散性和多样性的环境治理需求[19]。以ChatGPT为例,其可根据环境治理场景与要求调整生成内容。有学者建构了政府精准购买公共服务的大数据应用模型,主张通过监测数据的动态交换及流动过程,精准获取服务对象的实时需求及质量反馈,完善后续购买计划[20]。生成式人工智能大模型从大规模无监督的数据中学习环境治理知识,可降低针对性搜集治理对象需求信息的重复成本,通过知识转移匹配特定治理任务,减少重复训练和数据标注流程,

从而在多样化治理场景中保持施策的针对性。

三、生成式人工智能技术嵌入政府环境治理的内在缺陷

生成式人工智能嵌入政府环境治理可分为前期准备、实际运行及后续处理三阶段,生成式人工智能虽能在不同阶段为政府环境治理带来显著便利,但其在具体应用时仍潜藏诸多隐忧,有必要对其内在缺陷进行系统性检视。

1. 前期准备阶段

(1)价值权衡难以实现

政府环境治理须对部分环境事务作出价值判断和选择,在传统的环境治理模式下,价值权衡环节本就存在诸多问题,引入生成式人工智能参与治理后,这些问题并未消解,反而可能衍生出新的治理困境。

其一,政府环境治理中的价值判断高度依赖专业知识、工作经验与决策场景,难以由生成式人工智能完全取代。环境治理不仅涉及自然环境,还与社会环境的变化密切相关。自然环境和社会环境及其相关工作经验难以完全数据化,生成式人工智能缺乏身体化实践和具体场景的嵌入,难以理解跨区域性环境治理背后的利益纠纷、解决技巧等,而这些问题的解决有赖于治理主体在场景中体会和认知的经验性知识[21]。因此,生成式人工智能更适合应对知识交叉少、条件较为明晰的治理场景,但不能代替治理主体作出价值判断[22]。即便其作出类比推理,相关技术支撑仍不充分,输出内容的可靠性也缺乏充分保障。

其二,环境治理中各利益主体间的价值冲突,难以通过算法机制得到妥善权衡。立法者难以在事前确立绝对优先的价值基准,环境事务的价值权衡往往需在具体场景中事后展开。在具体场景中,需先考量争议背后不同利益代表所涉及的社会、经济、政治等要素,才能作出价值抉择。已有研究表明,生成式人工智能算法决策模型难以准确作出上述价值判断,即便具备类人化表达功能,其价值判断仍需基于算法作出,决策流程仍是“数据吸收-分析处理-输出结果”。以评分系统为例,其规则是事前建构的价值分配方案,难以随环境价值权衡的变化即时调整[23],一旦出现和事前设计场景不符的价值冲突,算法机制就可能输出偏离环境正义的决策。

(2)数据管理难以规范

生成式人工智能所学习的数据来源复杂、质量参差,数据的管理也欠规范。一是数据真实性难以保障。环境数据验证机制目前尚不健全,加之预训练数据来源复杂,使生成式人工智能输出信息真假难辨,影响环境决策的科学性和可靠性。随着公众获取环境政务信息的门槛显著降低,若模型依靠现有语料生成错误环境信息并发布至政务信息公开网站,缺乏专业知识的公众难以辨别真伪,一旦错误信息被广泛传播并误导公众,便可削弱政府公信力[24]。二是对数据的保护不到位。一方面,重要领域数据保护不足。生成式人工智能预训练数据库需纳入环境治理各领域数据,其中核工业、重要化工行业等领域的生产数据及自然资源资产数据与国家发展安全息息相关,若未采取特殊保护措施,便存在数据泄漏风险。另一方面,对企业与个人数据的保护不足[25]。环境监管部门收集企业和个人的数据可能面临多重风险:其一,采集环节缺乏有效同意机制。部分企业在收集数据训练模型时未征得个人同意,个人也缺乏渠道查询企业数据存储库中涉及的个人数据。其二,使用期间存在过度、无序采集问题。如将生成式人工智能嵌入环境信用评价,可能会涉及企业财务数据及相关个人信息的采集,一旦泄露,便可能造成企业财产损失和个人权益损害[26]。

2. 实际运行阶段

(1)算法歧视缺陷

生成式人工智能的模型算法由技术专家设计,其嵌入环境治理的训练数据也依赖技术专家的收集与输入,一旦技术专家的价值预设、认知局限及数据输入环节存在偏差,易导致多方面的算法歧视风险:一方面,算法歧视风险源于设计过程。受学历背景、生活经历、专业知识等局限,算法工程师在设计模型时常会自觉或不自觉地将不公平因素嵌入模型设计(如工程师的价值预设会通过权重设定、规则编写等环节进入模型)。以生成式人工智能嵌入环境、社会和治

理(Environmental, Social and Governance,下文简称 ESG)制度为例,当前 ESG 评价标准尚未形成国际统一规范,工程师在设定 ESG 中环境治理、社会责任和公司治理的权重时,会将偏见思维转化为代码嵌入模型,导致 ESG 评价结果不公。而 ESG 评价中环境指标权重的设定,同样会影响环境治理相关评价。另一方面,数据输入环节也可能引入歧视。即便技术专家中立地设计算法,环境治理工作者也可能出于私益,选择性地输入环境数据,如在环境影响评价领域,规划环评所需的环境质量、污染源分布、排放量等基础资料由生态环境主管部门掌握。在生成式人工智能嵌入环境影响评价流程后,掌管相关资料的环境影响评价部门可能会出于私益将有利于该项目的环评数据输入模型,从而导致环评结论有失公允,构成算法歧视。

(2)“算法黑箱”缺陷

生成式人工智能技术仍未克服“算法黑箱”的缺陷[27],主要有两种表现形式:一是因技术或商业秘密本身而形成的不透明,二是因技术的专业性所导致的相对黑箱。从技术或商业秘密保护看,嵌入政府环境治理的生成式人工智能算法多由商业机构设计开发,为维持其市场优势地位,公司会对算法的代码和运行逻辑等采取严格保密措施。从技术壁垒看,算法的专业性极强,普通公众缺乏相应算法知识,难以知道商业公司或其他主体可能以何种手段侵犯其正当权利,也难以行使陈述、申辩、听证等程序性权利。不仅如此,算法往往需要成千上万行乃至上百万行源代码的支撑,才能在复杂条件下执行人类预设的各种规则,未深度参与编程的程序员也难以完全理解算法内部的决策代码。“算法黑箱”使决策理由难以获得,进而损害公众的程序性权利。正如有学者指出的:“算法先发制人的能力,既损害了我们获得听证的权利和质疑那些试图对你不利的人的能力,也削弱了我们获得律师帮助、得到公开的决策理由以及在某些情况下对决定寻求上诉或提出司法审查的权利”[28]。可见,“算法黑箱”不只是认知层面的难题,更是对公众程序性权利的侵害。

3. 后续处理阶段

生成式人工智能嵌入政府环境治理后续处理阶段,会引发一个关键问题:在问责阶段,生成式人工智能能否成为合法的治理主体?针对此问题,已有研究存在不同认知。生成式人工智能应用时间较短,迭代迅速,相关法律规定比较概括、模糊、实践性和可操作性不足;同时,我国行政追责制度本身也存在责任认定标准模糊、责任类型复杂、责任主体界定不清等缺陷[29]。在此背景下,如何确立与生成式人工智能嵌入政府环境治理这一人机交互参与治理相配套的追责制度,变得更为棘手。

一方面,存在担责主体的资格合法性问题。生成式人工智能能否成为环境治理的行政主体尚存争议,这会进一步引发行政主体的合法性危机。另一方面,存在担责主体选择的模糊性问题。生成式人工智能的环境决策在很大程度上取决于操纵者所输入的语料,操纵者欲实现特定目的,会输入相应的要求和指令,生成式人工智能会根据操纵者的要求和输入的数据生成相应的答案[30]。然而,操纵者虽可决定输入内容,却难以精准把控模型输出结果和环境决策走向,决策意图和输出结果之间存在断裂,传统“谁决策,谁负责”的担责原则因而容易失效。是否应由生成式人工智能担责,其应在何种范围内担负何种责任,究竟应由技术开发者、使用者还是环境监管部门担责等问题都需要进一步商榷和考虑。

四、生成式人工智能嵌入政府环境治理风险的规制路径

为解决生成式人工智能嵌入环境治理过程中存在的前述难题,应当引入敏捷治理框架。敏捷治理强调全面性、适应性和灵活性,与以风险识别和处置为核心的风险治理有所不同,敏捷治理既契合生成式人工智能嵌入政府环境治理的治理需求,也与我国在新兴技术领域推行的包容审慎监管有共通之处。为此,规制路径的设计应融入敏捷治理的三维框架:就全面性而言,应当确立生成式人工智能嵌入政府环境治理的基本原则,借助法律原则的周延性和适度模糊性,在宏观层面覆盖治理全流程、全主

体;就适应性而言,应当划分已知和未知风险,构建事前预防和事后应对相结合的治理机制,以提升对问题的响应速度;就灵活性而言,应当创新监管制度,在生成式人工智能嵌入政府环境治理过程中保留必要的监管弹性。在敏捷治理框架融入的同时,还需关注科技革新趋势,需立足中国环境治理经验,从而提出生成式人工智能嵌入政府环境治理的规制路径。

1.基于敏捷治理的全面性要求,确立生成式人工智能嵌入政府环境治理的基本原则

其一,主体自觉原则。主体自觉原则立足人本主义立场,强调理性、全面地看待生成式人工智能应用的优势与风险,以人的理性约束技术发展。该原则要求政府、技术设计者与公众分别履行监管自觉、设计自觉与权利自觉,从外部约束、内部优化与权利救济3个层面引导技术向善。首先,政府环境监管部门应对生成式人工智能嵌入环境治理的过程展开全方位、全过程的监管,防范因技术嵌入而导致的行政权力异化,尤其应避免算法决策替代行政裁量。其次,技术设计者和开发者应通过优化内部模型算法逻辑、规避主观偏见影响,保障生成式人工智能所输出环境决策的公平性和可靠性。最后,企业和民众应主动了解生成式人工智能的运行机理和决策逻辑,若自身权益因技术漏洞或设计者私心受损,应及时向治理部门反映,或通过法律途径维权。

其二,法律与技术协同共治原则。科学技术与政府环境治理相结合,呈现出技术标准与法律规范相互交织、深度相融的特点,仅凭法律制度的构建与实施,难以取得理想的治理效果。环境治理与科技进步息息相关,生成式人工智能的嵌入既能给环境治理带来机遇,也潜藏诸多风险,这些风险若缺乏有效规制,可能给环境治理秩序和公共利益带来损失。传统的法律制度尚未设置针对生成式人工智能的纠错机制或截断机制,该功能在相当程度上需借助技术手段实现,即以代码制约代码,以技术牵制技术。因此,在技术快速进步的今天,需要引进法律与技术协同共治原则,既要依托技术手段阻断他人利用技术损害公众和社会利益的可能,也应在制度层面做好监管和保护。

其三,公平原则。《互联网信息服务算法推荐管理规定》中明确提出公正公平的原则[31]。生成式人工智能嵌入政府环境治理,需以善治为目标、维护公共利益,其公平维度分为两个层面:一是个体公平和群体公平,二是过程公平和结果公平。生成式人工智能嵌入环境治理须同时兼顾上述两个维度的公平价值诉求:一方面,个体公平强调技术嵌入环境治理过程中尊重每个个体的数据权利和隐私权,在收集环境数据时应避免个人隐私泄露与恶意滥用。群体公平要求技术设计标准公平对待各类群体,不将地理位置、资产多寡、企业属性、行业类别等可能隐含歧视的因素纳入算法设计。另一方面,过程公平要求环境治理遵循正当行政程序。但生成式人工智能嵌入后,传统行政程序难以完全适用,应建立适应生成式人工智能治理模式的行政程序,保障环境行政过程公平。结果公平要求生成式人工智能的输出不受算法歧视、数据偏差等不当因素的干扰,确保结果公平公正。

2.基于敏捷治理的适应性要求,构建事前预防和事后应对相结合的适应性治理机制

生成式人工智能嵌入政府环境治理可能引发伦理、技术及法律等多重缺陷,单一治理机制难以作出有效应对,因此应立足敏捷治理的理念,构建事前预防和事后应对相衔接的适应性治理机制,使治理手段随缺陷类型与风险程度动态调整[32]。

(1)事前预防

事前预防体系的构建包括3个层次:

第一层次是健全生成式人工智能嵌入政府环境治理的数据规范制度。生成式人工智能嵌入政府环境治理的核心在于深度开发利用数据,借助自然语言处理、人工神经网络分析、机器学习等技术优化环境治理服务。环境数据是生成式人工智能运行的基础语料,其采集、验证、存储和保护应予以重视,并建立系统的规范体系。

首先,健全环境数据采集制度。生成式人工智能具有自动化学习与生成特征,数据的质量在很大程度上决定输出质量。数据来源、类型与采集方式的多样性和复杂性使所汇集的数

据较为杂乱,质量难以保证,为此应统一数据采集标准和流程,并对来源与类型加以规范,同时建立质量核验环节。统一采集还有助于保障数据来源的合法性,避免企业、个人数据泄露及隐私权受到侵害。因此,建议可由环境监管部门统一向合规第三方购买采集服务,既统一标准,也便于审核数据来源合法性。

其次,健全环境数据的验证制度。环境数据应汇入大模型数据库用于预训练。为保障数据质量,须建立事前验证和事后验证制度:事前验证指数据入库前,由对应领域的专家对数据的关联性、真实性、合法性进行核验判断;事后验证则指若输出环节发现虚假、错误环境决策内容,则在回溯数据采集阶段开展二次复核,并追究相关环节责任。

最后,建立环境数据存储和保护制度。环境治理领域关乎国家利益和安全,应设立严格的数据分类分级管理制度,保护重要领域数据与敏感信息安全[33]。具言之,绝对保密数据(如核工业、自然资源分布和数量等与国家安全密切相关的数据)须由政府掌握绝对控制权,禁止技术开发商接触含此类数据的处理系统,合作开发过程中涉及此类数据的环节,须由政府全程管控。重要数据(多为企业和个人重要信息)的保密等级介于绝对保密数据与一般环境数据之间,须通过数字密钥管控查看权限,如需向大模型开放,须上传数据拥有者(企业或个人)的同意书,以实现环境利益、国家利益、个人权益三者的平衡。对保密等级较低的数据来说,可在履行必要的审核程序后,向大模型数据库开放,以提升环境决策的科学性。然而,不论保密等级高低,相关数据库均应设置严密防护,此举既能避免数据泄露侵害国家安全、企业和个人隐私,也能防范外部攻击者通过侵入数据库,掺杂虚假数据等方式,干扰模型运行,破坏环境治理工作[34]。

第二层次是建立生成式人工智能应用场景清单制度。分析生成式人工智能输出环境决策的特点和环境治理特质可知,其并非适用于政府环境治理的所有领域,为实现技术应用风险可控,须建立明确的应用场景清单制度,将政府环境治理清单划分为生成式人工智能完全决策、人机协同治理、禁止生成式人工智能嵌入三类。清单制订标准需系统考量生成式人工智能的生成和类比推理能力、治理领域的复杂程度、涉密程度等要素,以确定技术嵌入政府环境治理的深度和广度。

具体来说,三类场景的划分与清单确定可依以下标准展开:其一,以生成式人工智能的生成能力和类比推理能力为标准,将生成式人工智能能够汲取数据、快速生成文本决策并准确解决问题的场景纳入清单。其二,以环境治理场景的复杂程度和明确性程度为标准,将程序规定明确、需比对大量类似案例的政府环境治理事项交由生成式人工智能决策。其理由在于,此类事务程序清晰,有类似案例可比对,决策结论可预测,生成式人工智能算法更为适用,其优势更容易发挥。此外,应建立清单动态调整机制,适应技术发展的趋势。随着技术迭代升级,生成式人工智能的可靠性和安全性有望提升,届时,可逐步将其应用于更多治理领域与事项。对新增清单应用事项,应通过技术听证会等形式吸纳专业意见,保障清单调整的科学性。

第三层次是建立算法解释制度。“算法黑箱”是引发治理风险的主要原因之一,然而,算法解释不等于将算法全面公开,一旦将算法全面公开,可能会引发市场不正当竞争,算法也易被恶意行为者利用,干扰环境治理进程,损害公共利益。因此,应围绕解释主体、解释内容及解释边界三要素,结合环境治理的具体场景展开制度设计。

首先,应明确解释主体。商业公司不宜作为算法解释的主体,解释权不应掌握在算法创造者手中,但当算法决策实质影响环境治理中的权利义务分配时,其便具有公权力行使的属性。公权力的行使主体是政府,因此政府应当作为算法解释的主体。政府经人民授权行使环境治理的权力,其利用算法实施环境治理的行为应该接受公众的监督,一旦缺乏算法专业知识的公众难以理解算法决策的依据,此时的政府有必要就此作出合理的解释。若算法解释错误,可能会造成其他社会主体难以承受的公共损失,因此,由政府主导解释并辅以第三方监督

更为合适。政府可设立专门的算法解释部门,当公众、企业或社会组织对某些环境决策有异议时,可向该部门申请解释。其次,应明确解释内容。算法解释的内容不应是一般算法架构或代码技术的解释。解释内容应结合环境治理的具体场景展开,如在环境预警场景中应解释算法承担的角色与具体作用。在具体的解释场景中,环境主管部门应组织算法专家、环境治理专家参与,并出具相应的解释报告,以增强解释的权威性和可说服性。最后,应明确解释边界。公众权益需要保护,商业公司的技术秘密同样受法律保护,且算法的源代码可能涉及商业公司核心技术,不加限制地解释算法,既无益于保护公众的权益,也易造成商业秘密外泄。算法解释应保持必要限度,由商业公司、政府和第三方监督机构共同界定解释边界,即明确解释范围与不能披露的核心源代码等。同时,政府可设立算法审计委员会,将三方主体纳入委员会,对具体解释边界协商确定,必要时引入专家论证程序,以期在保障公众的知情权与商业公司的正当权益之间取得平衡[35]。

(2)事后应对

事后应对是指建立生成式人工智能嵌入政府环境治理后的责任分配制度。生成式人工智能嵌入政府环境治理后,其责任分配不能僵化地局限于单一主体,而应与其技术特点相适应:一方面,应考虑在承认生成式人工智能可独立承担法律责任的前提下,如何分配法律责任(当前相关法律文件中,生成式人工智能并不被认可作为独立承担法律责任的归责主体,本研究只是探讨假设其可以作为归责主体时的责任分配)。另一方面,应考虑当生成式人工智能不具备责任能力时,如何分配法律责任。针对第一种情况,生成式人工智能独特的自然语言处理系统与算法构造,使其法律责任认定既不同于一般环境法律责任承担,也有别于传统部门法的因果关系认定,评判其责任分担比例,应综合考虑生成式人工智能作出环境决策时受他人操控的程度、技术故障等因素。

即便生成式人工智能可独立承担责任,其责任仍可与研发公司、操控者、监管者按比例分配,具体可按致损原因分为以下情形:一是生成式人工智能因自身底层逻辑混乱或运算失误作出错误的环境决策的,由研发公司依合同承担违约责任,政府有关部门对治理相对人承担相应行政责任。二是因生成式人工智能的操控者操控失误导致决策错误的,由操控者承担主要的责任。无法查明具体操控者的,由操控部门集体承担责任。三是若环境治理行为出现错误,但无证据证明主要的责任承担主体是生成式人工智能、操控者还是研发者,可依公平原则由三者按照合理的比例承担责任。四是因监管部门的失误和疏忽造成损失的,主要行政责任由监管部门承担。

针对第二种情况,即当生成式人工智能不具备独立承担责任能力时,从因环境治理决策违法或不当而受损的治理相对人这一角度思考,政府环境治理部门应承担主要行政责任。民众将环境治理权授予政府,政府作为合法治理主体,应为治理行为担责,这也是维护政府公信力的需要。以第三方技术研发公司为担责主体,既缺乏明确法律依据,实践中追责成本也过高,环境治理相对人通常不掌握生成式人工智能的底层逻辑和设计机理,难以提出确实的证据证明应由研发公司担责。若由环境治理部门担责,可促使政府投入更多的人力、物力、财力参与生成式人工智能监督系统的设计、测试和实施,提升生成式人工智能监督系统的公正性和科学性。当然,为弥补应用生成式人工智能可能带来的损失,也可构建风险分担机制,由政府环境治理部门从项目资金中划拨一部分投入专项基金,设定观察期,在此期间,若生成式人工智能运转良好、未损害国家和社会公共利益,则逐步返还剩余资金;若发生损害,则以该资金赔偿被侵权人的损失。

3. 基于敏捷治理的灵活性要求,创新生成式人工智能嵌入政府环境治理的过程监管制度

过程监管涉及生成式人工智能嵌入政府环境治理的监管主体及其职责范围,生成式人工智能技术具有独特性和复杂性,需利用敏捷治理的灵活性特征,从多方面展开过程监管。

一是建立生成式人工智能的监督、协同、问责部门。生成式人工智能具有强专业性,当前,监管机构相对分散、专业化程度较低,对生成式

人工智能发展的支持、引导和监管长期处于较为分散与被动的状态。随着生成式人工智能技术的快速迭代,可预见其将迎来技术爆发阶段,因此,环境治理机构应设立独立的生成式人工智能监管部门,配备兼具人工智能技术和环境专业能力的复合型人才。在职能上,除对应用于政府环境治理的生成式人工智能进行日常监管外,该部门还应规范其使用范围,对数据来源、使用场景、应用领域、隐私保护等作出明确规定。

二是建立生成式人工智能嵌入环境治理的伦理验证、评估许可及审计制度。首先,应建立伦理认证机制。人工智能伦理认证是指以透明性、隐私保护、算法公平、责任可追溯为裁量基准,将社会广泛认可的价值原则转化为可操作的评估指标和方法。此举应由政府环境治理部门聘请第三方机构,对生成式人工智能嵌入环境治理的技术和产品进行独立认证,认证通过后,由政府环境治理部门允许其应用于环境治理实践[36]。其次,完善技术使用前的评估许可制度,根据技术特点和不同环境治理场景的重要程度,分别采取强制评估、随机抽查及依申请评估等方式。对在重要领域中使用的内部模型算法,应当在正式运行前,由政府环境部门聘请专业第三方机构进行技术安全与场景适配评估。最后,对已投入使用的生成式人工智能,应由专门监督机构对其嵌入环境治理的过程进行监管,并根据其参与领域的重要程度,采取定期审计、随机审计及依申请审计等方式,检测其输出的内容是否存在偏误、缺陷和歧视。治理相对人对生成式人工智能输出的决策、处罚、方案等内容有异议的,可向监管部门申请审查,监管部门也可依职权启动审查程序[37]。

三是建立社会组织、公众协同监管制度。生成式人工智能嵌入环境治理涉及复杂的“技术+环境”的互动,具有跨界融合、群智开放、人机互动等特性,单一政府监管难以覆盖,需引入多元主体协同。同时,技术嵌入过程也为社会组织和公众提供了参与监督的机会,如自然之友等自然保护组织可通过模拟测试、问卷调查等方式进行监督[38]。社会公众是监督的重要力量,其中,人工智能专业人士可发挥专业优势积极参与监督。如受网络安全“漏洞赏金”机制启示,X(原推特)开展了算法偏见赏金竞赛,借助人工智能领域人才识别图像裁剪算法的潜在风险[39],这一机制对环境治理中算法偏见的公众监督具有借鉴意义。环境治理的许多领域(如环境信用评价、环境影响评价、ESG 评价等)与企业的利益相关,行业协会可设立人工智能技术咨询机构,就评价机制中的技术问题提出意见。为保障普通公众有效行使监督权,政府机构和社会组织还应设立面向社会公众的人工智能技术咨询部门,解答技术问题,防止公众监督流于形式。此外,在生成式人工智能拟嵌入环境治理的某一具体领域前,可召开技术听证会,召集专家、社会公众与相关企业等利益相关人,集中探讨技术嵌入的可能性、合理性及风险承担,从而作出是否引入该技术的决策。

五、结　语

随着生成式人工智能技术的快速发展,将其嵌入政府环境治理,对提升环境治理成效、降低环境治理成本具有重大意义。但技术发展不等于无条件的信任和支持,在生成式人工智能参与环境治理仍存在不确定性的背景下,应及时构建有效的法律规制方案,最大程度保障生成式人工智能在环境治理中稳妥应用。本研究尝试提出相应规制方案,但后续研究仍有进一步探讨的空间:如何处理独立的生成式人工智能监管部门与现有人工智能监管部门之间的关系、如何判定生成式人工智能输出环境决策或方案是否违法或失当等。未来,应立足技术发展与应用实践,持续完善生成式人工智能嵌入环境治理的规制方案,以推动政府智慧环境治理目标的有效实现。

参考文献:

[1] 习近平在全国生态环境保护大会上强调　全面推进美丽中国建设　加快推进人与自然和谐共生的现代化[N]. 人民日报,2023-07-19(01).

[2] 习近平. 习近平谈治国理政:第五卷[M]. 北京:外文出版社,2025:333.

[3] 王静,张红. 人工智能在数字政府建设中的应用及其法治化路径——基于 DeepSeek 在数字行政中应用场景的分析[J]. 中外法学,2025,37(3):625-643.

[4] 秦鹏,周继钊. 算法嵌入政府环境治理的正向效应、潜在风险与规制路径[J]. 中国人口·资源与环境,2024,34 (6):80-89.
[5] 张邦辉,李一博. 赋能与使能:数字技术驱动生态环境治理现代化转型的重要路径[J]. 探索,2025 (3) : 96-114.
[6] 胡刘,牛婷婷. 生成式人工智能发展前景下公共性的消解及重建[J]. 南通大学学报(社会科学版) ,2024,40 (4) : 25-35.
[7] 王春英,姚亚妮,滕白莹. 生成式人工智能嵌入敏捷政府建设:影响、风险与应对[J]. 北京行政学院学报,2024(3):73-83.
[8] 李汉卿,孟子龙. 数字政府建设何以实现敏捷治理:多维度展开及其不确定性克服[J]. 求实,2022(5):26-37.
[9] 薛澜,赵静. 走向敏捷治理:新兴产业发展与监管模式探究[J]. 中国行政管理,2019(8):28-34.
[10] 赵静,薛澜,吴冠生. 敏捷思维引领城市治理转型:对多城市治理实践的分析[J]. 中国行政管理,2021(8):49-54.
[11] 杜焱强,詹昕颖. 数字技术如何驱动农村人居环境敏捷治理? ——一项双案例研究[J]. 南京工业大学学报(社会科学版),2024,23(4):42-57.
[12] 于文轩. ChatGPT 与敏捷治理[J]. 学海,2023 (2):52-57.
[13] Chatterjee J,Dethlefs N. This new conversational AI model can be your friend,philosopher,and guide and even your worst enemy [J]. Patterns, 2023, 4 (1):100676.
[14] 重庆市人民政府网. 智能问答[EB/OL]. (2025-11-05) [2026-06-26]. https://www. cq. gov. cn/dhc/cq/index. html.
[15] 刘丹鹤,孙嘉悦. 人工智能规制政策制定的风险与治理[J]. 人文杂志,2023(2):121-128.
[16] 尹克寒. ChatGPT 的发展对情报信息工作的影响及启示[J]. 图书馆理论与实践,2023(3):15-22.
[17] 曾宇航,史军. 政府治理中的生成式人工智能:逻辑理路与风险规制[J]. 中国行政管理,2023,39 (9):90-95.
[18] 郑少华,张翰林. 数智时代环境智理的内涵与构建路径[J]. 甘肃社会科学,2023(5):169-176.
[19] De Lemos R,Giese H,Müller H A,et al. Software engineering for self-adaptive systems: a second research roadmap [C]//Proc. of the software engineering for self-adaptive systems II. Berlin, Heidelberg:Springer-Verlag,2013:1-32.
[20] 董正爱,张黎晨. 数字技术嵌入环境风险预防:范式变迁、规范解构与法典表达[J]. 湖南科技大学学报(社会科学版),2026,29 (2):89-98.
[21] 陈锐,孙庆春. 道德计算是否可能——对机器伦理的反思 [J]. 科学技术哲学研究,2020,37 (4):74-80.
[22] Harry Surden. Artificial intelligence and law: an overview〔J〕. Georgia State University Law Review, 2019(4):1305-1337.
[23] Corbett-Davies S, Pierson E, Feller A, et al. A computer program used for bail and sentencing decisions was labeled biased against blacks. It's actually not that clear [J]. Washington Post, 2016,17.
[24] 秦鹏,周圣佑. 区块链环境风险的逻辑缘起、规制困境及其制度纾解[J]. 东南大学学报(哲学社会科学版),2023,25(3):80-89.
[25] Deng J, Lin Y. The benefits and challenges of ChatGPT:an overview[J]. Frontiers in Computing and Intelligent Systems,2023,2(2):81 -83.
[26] 李海平. 个人信息国家保护义务理论的反思与重塑[J]. 法学研究,2023,45(1):74-90.
[27] 樊博. ChatGPT 的风险初识及治理对策[J]. 学海,2023(2):58-63.
[28] Kerr I,Earle J. Prediction,preemption,presumption: how big data threatens big picture privacy [J]. Stanford Law Review Online,2013,66(65):65-72.
[29] 杜辉,赵琛. 论《生态环境法典》中的国家条款[J]. 南京大学学报(哲学·人文科学·社会科学),2026,63 (4):102-113.
[30] 林园春. 人工智能驱动政府治理模式变革的逻辑与策略[J]. 中州学刊,2022(9):93-98.
[31] 互联网信息服务算法推荐管理规定[J]. 中华人民共和国国务院公报,2022(9):71-75.
[32] 董正爱,邱窈裔. 生态环境法典中预防为主原则的规范解释与适用路径[J]. 齐鲁学刊,2026,(4):79-94.
[33] 朱千一. 论数据分类分级的目的、原则与规则[C]//上海市法学会. 上海法学研究集刊——智慧法治学术共同体文集,2023:8-19.
[34] 郭小伟. 生成式人工智能数据风险治理模式评析与机制完善[J]. 科技进步与对策,2026,43 (8):1-12.
[35] 臧雷振,陈浩. 生成式人工智能算法风险及社会治理挑战[J]. 中共中央党校(国家行政学院)学报,2025,29 (1):43-53.
[36] 王鹏飞. 污染环境罪治理中生态修复机制的多维

建构[J]. 中国人民公安大学学报(社会科学版),2025,41(5):51-61.
[37] 刘冬妮."数字悬浮":政府数字化治理的结构性困境及其矫治策略[J]. 社会主义研究,2026(4):108-115.
[38] Secretary of state for science & innovation and technology, a pro-Innovation approach to AI regulation[EB/OL]. (2023-03-29)[2023-04-10]. https://assets. publishing. service. gov. uk/government/uploads/system/uploads/attachment_data/file/1146542/a_pro-innovation_approach_to_AI_regulation.
[39] 胡晓萌. 从"算法偏见赏金"说起:科技伦理治理的众包模式[J]. 数字经济,2022(6):18-23.

Logic and Regulation of Embedding Generative Artificial Intelligence into Government Environmental Governance / Qin Peng, Zhou Jizhao (School of Law, Chongqing University)

Abstract: The development of generative artificial intelligence (AI) provides unique technical support for government environmental governance. Generative AI can rapidly organize environmental data. As a "digital customer service agent", it can realize human-like expression and assist staff in making environmental decisions. As generative AI is deeply embedded in government environmental governance, it drives logical transformations in governance subjects, governance processes, and governance tools, while bringing stage-specific deficiencies: in the preliminary preparation stage, there are deficiencies in value and data management; in the actual operation stage, there are deficiencies of algorithmic discrimination and "algorithmic black boxes"; and in the subsequent processing stage, there are deficiencies in responsibility allocation. The introduction of agile governance theory renders the regulatory path more comprehensive, adaptive, and flexible. Therefore, the basic principles of subject consciousness, collaborative governance between law and technology, and fairness, which conform to comprehensiveness, should be established; an adaptive governance mechanism combining ex-ante prevention and ex-post response, which conforms to adaptability, should be constructed; and a process supervision mechanism, which conforms to flexibility, should be innovated. The ex-ante prevention level of the adaptive governance mechanism should be divided into three levels: The first level establishes a system for data collection, verification, storage, and protection; the second level establishes a system for the application list of generative AI; and the third level establishes an algorithm explanation system. The ex-post response level should establish a responsibility allocation system after generative AI is embedded in environmental governance. The process supervision mechanism should include three aspects: establishing supervision, coordination, and accountability departments for generative AI; establishing an audit system for ethical verification, evaluation and licensing, and operational use of generative AI embedded in environmental governance; and establishing a collaborative supervision system for social organizations and the public. In this way, the regulatory scheme for embedding generative AI in environmental governance can be continuously improved, thereby promoting the effective realization of the goals of smart government environmental governance.

Key words: generative AI; digital government; environmental governance; "algorithmic black box"; agile governance theory