Modern Management 现代管理, 2025, 15(9), 123-129
Published Online September 2025 in Hans. https://www.hanspub.org/journal/mm
https://doi.org/10.12677/mm.2025.159253

Hans汉斯

生成式人工智能嵌入公共服务治理的实际效能与风险防范

吕泽州

西安交通大学公共政策与管理学院，陕西 西安

收稿日期：2025年7月31日；录用日期：2025年8月12日；发布日期：2025年9月15日

摘　要

生成式人工智能的快速发展正深刻重塑公共治理的技术范式与价值逻辑。本文基于技术治理理论框架，以“技术嵌入性－治理适应性”为核心分析维度，系统阐释生成式人工智能赋能公共治理的运行效能、风险挑战与政府规制路径。研究发现，生成式人工智能通过创新治理场景构建、赋能治理模式革新、重塑治理评估范式，显著提升了公共服务的精准性、协同性与动态调优能力。然而，其技术动态性也衍生训练异化、内容失范、应用失当等风险，并与传统治理体系的线性思维、割裂管理及科层滞后性形成深层矛盾，凸显“科林格里奇困境”。研究提出以“敏捷治理”为核心的应对方案，通过伦理责任明晰化、算力数据协同化、多元共治制度化等政策工具组合，实现技术赋能与治理调适的动态平衡。

关键词

生成式人工智能，公共服务，实际效能，风险防范

The Actual Effectiveness and Risk Prevention of Embedding Generative Artificial Intelligence into Public Service Governance

Zezhou Lyu

School of Public Policy and Management, Xi’an Jiaotong University, Xi’an Shaanxi

Received: Jul. 31st, 2025; accepted: Aug. 12th, 2025; published: Sep. 15th, 2025

Abstract

The rapid development of generative artificial intelligence is profoundly reshaping the technological paradigm and value logic of public governance. This article is based on the theoretical framework

文章引用: 吕泽州. 生成式人工智能嵌入公共服务治理的实际效能与风险防范[J]. 现代管理, 2025, 15(9): 123-129.
DOI: 10.12677/mm.2025.159253

of technology governance, with "technology embeddedness governance adaptability" as the core analytical dimension, systematically explaining the operational efficiency, risk challenges, and government regulatory paths of generative artificial intelligence empowering public governance. Research has found that generative artificial intelligence significantly improves the accuracy, collaboration, and dynamic optimization capabilities of public services by innovating governance scenarios, empowering governance model innovation, and reshaping governance evaluation paradigms. However, the dynamic nature of its technology also leads to risks such as training alienation, content deviation, and improper application, and forms deep contradictions with the linear thinking, fragmented management, and bureaucratic lag of traditional governance systems, highlighting the "Colin Gridge dilemma". The study proposes a response plan centered on "agile governance", which achieves a dynamic balance between technological empowerment and governance adjustment through a combination of policy tools, such as clarifying ethical responsibilities, coordinating computing power and data, and institutionalizing diverse governance.

Keywords

Generative Artificial Intelligence, Public Services, Actual Effectiveness, Risk Prevention

1. 引言

生成式人工智能的崛起加速了强人工智能时代的到来，在引发“奇点”期待的同时，也使技术治理面临“科林格里奇困境”的严峻考验。2024 年，世界人工智能大会暨人工智能全球治理高级别会议发表《人工智能全球治理上海宣言》[1]，明确共同促进人工智能技术发展和应用的必要性，党的二十大报告提出，“加快发展数字经济，促进数字经济和实体经济深度融合，打造具有国际竞争力的数字产业集群”[2]，在政府治理方面，党中央高度重视数字技术应用于国家治理现代化，统筹推进中国式现代化发展，2023 年中共中央印发《数字中国建设整体布局规划》，加快数字中国建设，构筑国家竞争新优势[3]，2023 年 8 月起施行《生成式人工智能服务管理暂行办法》，旨在促进生成式人工智能健康发展与规范应用[4]。随着生成式人工智能深度应用，我国社会治理范式进入技术重构的新周期，治理形态随之发生系统性变革，评估其实际效果及作用路径，深入分析其潜在风险与政府防范机制，有助于审慎接纳生成式人工智能带来的新的机遇与挑战。

2. 理论框架：技术治理的嵌入性矛盾与科林格里奇困境

技术治理理论的核心在于解析技术系统与治理体系的互动关系，其分析框架可分解为三个关键维度：技术嵌入性、治理适应性以及二者动态调适过程中形成的科林格里奇困境。这一框架为理解生成式人工智能(GAI)在公共治理中的效能与风险提供了系统性视角。

从理论视角来看，生成式人工智能通过算法自主性、数据驱动性和应用渗透性三个维度深度嵌入治理体系，形成了“治理即计算”的新范式。在提升公共服务效率的同时，导致治理主体对技术系统的深度依赖。治理体系的适应性缺陷则在这一过程中表现得尤为突出。现行制度框架在面对技术变革时呈现出明显的响应迟滞，具体表现为法律规制的内容更新速度远低于技术迭代周期，跨部门协同机制难以应对技术风险的跨界传导特性，以及风险治理工具无法匹配人工智能系统的复杂性特征。技术－治理的速

度差在实践层面形成了多维度的冲突格局。在数据维度，训练数据集规模的指数级扩张与隐私保护制度的渐进完善之间产生了明显的时滞效应；在算法维度，神经网络模型的复杂化趋势与算法透明度要求之间形成了难以调和的矛盾；在应用维度，技术场景的快速拓展与治理能力的区域性差异导致了新的数字鸿沟。破解这一困境的核心在于建立技术发展与制度调适之间的动态平衡机制。敏捷治理框架的构建应当着眼于三个关键环节：首先是建立实时感知系统，通过嵌入式的监测网络压缩风险识别时滞；其次是创新政策实验机制，在可控范围内测试规制工具的适用性；最后是完善反馈调节回路，将技术演进的动态特征转化为制度优化的持续动力。这种治理范式的转型不仅需要技术层面的创新，更需要治理思维从控制导向向适应导向的根本转变。

理论框架的深化为后续风险分析和对策研究提供了系统性的认知基础。通过揭示技术嵌入性与治理适应性之间的动态互动关系，我们可以更准确地把握生成式人工智能在公共服务领域应用的实际效能边界，也为构建具有前瞻性的风险防范体系奠定了理论基础[5]。

3. 实际效能：技术嵌入驱动的治理范式跃迁

作为基于深度学习与海量数据训练的人工智能技术范式，生成式人工智能正成为公共服务治理转型的核心引擎，其通过重构数字政府组织架构、破解民生服务领域的治理瓶颈，借助智能化精准交互机制推动治理行为实现范式跃迁。

(一) 创新治理场景构建，拓宽公共服务边界

生成式人工智能正通过创新治理场景构建和拓展公共服务边界，成为推动公共服务治理现代化的核心驱动力。该技术以深度学习和大规模数据训练为基础，通过持续迭代升级赋能社会服务场景，实现精准适配与治理要素高效流通，在服务模式维度，实现从滞后性响应向前瞻性预判的质变跃升，在治理结构维度，完成由单一主体管控向多元协同共治的系统重构。具体而言，一方面，通过深度挖掘存量数据价值，在公共服务全链条实现“适数化”改造——解析数据特征规律、挖掘潜在关联价值，优化资源配置效率，如智能交通系统通过融合分析多源数据构建高精度预测模型；另一方面，着力提质扩容数据要素增量供给，依托全域数据感知网络构建、跨领域知识图谱融合以及多场景治理策略生成，解决传统数据共享安全顾虑，并确保各参与主体的合法权益，显著降低智慧农场、智能医疗等场景的运营成本，满足数字化转型对高质量数据资源的多元化需求。通过“存量活化”与“增量提质”的双轮驱动机制重构公共服务治理的范式与效能。

(二) 赋能治理模式革新，提升服务精准对接

生成式人工智能深刻重塑公共服务供给模式，通过三重交互机制创新推动服务范式向智能对话式转型：在交互逻辑层面重构主客体关系，在媒介层面增强技术赋能，在路径层面优化响应流程，最终实现公共服务与公众需求的高效精准对接。其一，生成式人工智能技术通过大语言模型的深度语义解析与情境推理能力，构建起新型需求识别–响应机制，生成式人工智能通过构建智能体协同网络，赋能政府部门实时洞察社会需求的深层关联与动态演化趋势，首先，基于语义理解与模式识别技术，大幅提升对高频隐性需求的捕捉精度；其次，通过需求知识图谱的动态建模与可视化呈现，促进传统治理中长尾需求显性化；最终，形成“需求感知–图谱构建–服务预判”的闭环治理链条，推动公共服务从被动响应向主动预见转型。其二，生成式人工智能重构公共服务资源的协同配置逻辑，生成式人工智能技术可突破传统资源配置的时空约束，推动公共服务资源配置从经验驱动向数据驱动跃升[6]。从治理范式转型视角来看，技术应用从中心城市向基层社区呈波纹状扩散的同时，通过数字化网络节点产生乘数效应，最终形成“点–线–面”多维联动的公共服务资源共享新格局。具体表现为：一方面，公共服务需求高度集中的中心城市在模型训练过程中，需要将数据采集网络向基层治理单元延伸，并通过算法优化适配地方

治理情境；另一方面，技术下沉过程产生的数字溢出效应，为缩小区域间公共服务供给的质量鸿沟提供了持续的技术动能。双向互动机制既确保了人工智能模型的本地化适配能力，又通过技术势能的梯度转移促进了公共服务均等化发展。

(三) 重塑治理评估范式，实现服务动态调优

公共服务治理现代化的动态优化亟需构建科学敏捷的评估体系。生成式人工智能凭借其动态建模、多源数据融合和智能推演等技术特征，推动治理评估实现三重范式突破：从静态结果评价转向过程动态干预，从单一维度评估转向立体综合研判，从经验决策转向模拟预判，最终构建起全周期、自适应、智能化的新型评估生态系统。

生成式人工智能赋能公共治理的技术路径呈现三维立体架构：首先，在实时评估维度，依托自然语言处理与深度学习算法构建动态监测系统，通过参数自适应机制持续优化评估指标，实现对治理盲区的智能预警。典型如社区养老场景，通过老年人健康数据的实时解析，完成需求分级可视化呈现、服务方案动态调适及资源流转全周期追踪，使服务匹配效率提升 40%以上。其次，在跨域评估维度，突破组织壁垒构建多源数据融合平台，自动生成跨部门关联评估矩阵。以基层医疗为例，整合人口特征、诊疗记录与空间地理信息，精准定位医疗资源错配区域，为分级诊疗改革提供数据支撑。最后，在决策支持维度，运用大语言模型与知识图谱技术，模拟推演政策情景、预判供给矛盾并生成优化方案。“监测－分析－决策”的技术闭环，标志着公共治理进入智能迭代新阶段。

4. 潜在风险：科林格里奇困境的多维呈现

生成式人工智能风险谱系已发生根本性嬗变——从初始阶段的技术性缺陷(如算法偏见、数据偏差等单点问题)，逐步演化为涉及技术伦理、制度适配和社会接受度的系统性治理挑战。这一风险形态的跃迁，既渗透于算法训练、内容生成与应用落地的技术全链条，又折射出治理思维、工具与体制的适应性困境。技术层面的数据侵权、算法歧视与内容失范等问题，与治理层面的线性思维桎梏、工具滞后及体制僵化相互交织，凸显了人工智能嵌入公共治理的复杂性与矛盾性。

(一) 技术层面：嵌入性风险的三重裂变

生成式人工智能的技术架构主要基于算法训练、内容生成和模型应用三大核心环节，这一基础技术范式在推动治理创新的同时，也构成了风险演化的内生性根源。

从训练异化风险来看，生成式人工智能的核心技术机理在于深度学习的算法架构，通过海量数据训练和高性能算力支撑，构建具有超大规模参数的神经网络模型，在技术实现路径上，系统通过持续的参数优化与模型微调(fine-tuning)，使算法模型在特定任务场景中逐步逼近最优性能表现。这一方面极易引发数据侵权风险，训练数据可能来自公有领域仍处于版权保护有效期的作品；另一方面加剧算法歧视风险，即意识偏见风险，2024 年联合国教科文组织发表研究报告认为，“生成式人工智能存在性别偏见与种族刻板印象倾向等歧视风险”[7]，这意味着，“我们无法抗拒算法，但是我们可以规训算法；我们需要算法，但是我们真正需要的是可以信任的算法”[8]。从内容失范风险来看，一是内容可信度风险，生成式人工智能深度融入公共服务领域，在提升政府响应效能和决策科学化水平的同时，也引发了值得警惕的治理异化现象。从技术治理的辩证视角来看，技术在推动治理模式智能化转型的过程中，其固有的技术依赖性与算法不透明性正逐步削弱治理体系的自主性。以典型的算法黑箱问题为例，公共部门在应用 AI 决策时面临双重困境：一方面难以穿透神经网络的可解释性屏障验证决策合规性，另一方面受制于技术复杂度无法构建清晰的问责链条。这种技术层(算法模型)与应用层(治理实践)的权责断裂，本质上反映了人工智能赋能与治理自主性之间的深层张力。二是内容侵权风险，意大利曾于 2023 年强制暂停全域 ChatGPT 应用，坚持生成式人工智能“监管至死”，最终陷入技术发展与安全的两难境地。从风险演化

的维度审视，生成式人工智能的应用失当风险呈现出显著的层级跃迁特征，相较于微观层面可观测的生成内容风险(如事实性错误或价值偏见)，其宏观治理场域更易诱发具有扩散性和复杂关联性的系统性风险，当生成式人工智能被嵌入公共服务的决策链条时，单个应用节点的偏差可能通过行政系统的传导放大，最终演变为影响治理体系韧性的结构性风险。一方面，诱发算法诚信风险，“当生成式人工智能一路过关斩将、成为人类社会不可或缺的一部分时，其可能揭开人类与‘类人’进行深度社会互动的一幕”[9]，技术主导特征下的自动化决策、智能推送等技术路径不仅重构政民互动的基本模式，而且存在公共服务偏离人文关怀底色等分歧，导致核心价值让位于技术理性。另一方面，放大数智鸿沟风险，学界普遍认为，数字鸿沟经历了三次范式转变，从早期的接入问题到用户素养、再到当下数据驱动的智能算法，早期以硬件设施和数字素养为核心的能力差距，正在重构为包含算法理解力、数据治理能力和智能决策水平等维度的新型数智鸿沟，更深层次地反映了治理主体在人工智能时代适应性与创新力的系统性分化，从而加剧公共服务获取的马太效应。

(二) 治理层面：适应性危机的制度根源

生成式人工智能的迅猛发展在赋能公共治理的同时，也暴露出传统治理体系的深层困境：静态思维难以匹配技术动态性，割裂管理无法应对风险跨界性，科层调控显现明显滞后性。制度刚性与技术局限形成“双重瓶颈”，风险分级标准模糊与协同机制碎片化凸显体制创新需求。亟需构建敏捷、整体、包容的新型治理框架以实现智能治理的良性嵌入。

从治理思维来看，一是存在风险类型化治理思维的偏差，过度依赖对技术风险的分类预设与逐一应对模式，试图通过类型化推演和主观风险假想来构建治理框架。一方面，囿于实证主义的风险认知维度，仅针对可观测的“实在性风险”设计对策；另一方面，过度延伸至基于主观想象的风险场景，导致治理策略的建构缺乏坚实的现实基础。二是存在技术分离型治理思维的桎梏，根本困境源于对技术风险动态性的认知不足。现有治理模式采用技术要素分离的管理思路，将数据、算法、平台等核心组件进行割裂式管控，碎片化治理架构导致技术风险的系统性特征被人为分解、风险治理权责在部门间离散化分布、治理效能被技术要素的流动性所消解，过度依赖制度的功能性调节、经济性补偿和秩序性维护等静态手段，难以应对生成式人工智能风险跨域流动、动态演变的本质特征。三是存在线性控制型治理思维的滞后，其以政府资源集中投入为特征，通过行政力量主导的外部调控机制实现治理目标，本质是工业时代治理思维的延续，造成风险传导的多向性消解单向管控的有效性、技术迭代的快速性超越行政调控的响应速度、系统关联的复杂性模糊干预措施的靶向定位。结构性错配迫切要求重构适应智能时代的敏捷治理体系。从治理工具来看，一是制度工具自身存在不适应性，从制度工具设计理念维度看，单一的技术中心主义治理范式难以应对风险的多维复杂性，现有制度设计过度聚焦于技术使用者(“人素”)的规制，而忽视技术系统与社会环境的交互影响，从制度工具供给维度看，技术发展的不确定性与立法过程的滞后性形成结构性矛盾，表现为专门立法的时效性赤字与治理成本攀升并存，技术快速迭代与制度渐进调适之间存在难以弥合的速度差。二是技术工具有待创新深化，公共部门现有的技术管控能力与生成式人工智能风险的复杂性、不确定性特征之间存在结构性落差。具体表现为，政府部门主导开发的风险识别－监测－控制技术系统，在应对生成式 AI 特有的涌现性风险和跨域传导效应时存在三重局限，风险识别维度难以覆盖技术应用的长尾场景、监测机制无法实时捕捉风险的动态演变、控制手段缺乏应对非线性风险的弹性空间。从治理体制看，一是技术风险分级治理具有僵化趋向，一方面，欧盟《人工智能法》创新性采用基于风险等级的分类监管框架，将人工智能风险谱系划分为不可接受风险、高风险、有限风险和轻微风险四个监管层级。导致其风险分级科学性的认知局限导致分类依据存在争议，同时，风险等级评估标准的模糊性使得监管对象难以准确归位，制度设计与技术现实的适配落差，反映了传统规制工具应对智能技术复杂性的局限性；另一方面，跨部门协作机制不足，组织架构的碎片化导致治理真空地带，

各部门权责边界模糊不清，同时信息孤岛效应造成数据资源难以共享，引发重复建设和资源错配。二是忽略不同治理主体差异，生成式人工智能风险治理体系内不同主体在治理动机与优势等方面存在明显差异。一方面，传统科层制治理体系介入新技术领域，在治理负责人定位、治理战略规划能力等方面存在指向性较弱的问题；另一方面，面对生成式人工智能引发的多层次风险谱系(从微观个体权益到宏观公共安全)，现有治理体系表现出双重适应性不足，或监管滞后导致风险积累，或干预过度陷入“科林格里奇困境”。

5. 防范机制：敏捷治理的政策工具

生成式人工智能的快速发展在为公共服务治理带来革命性变革的同时，也使得传统的治理体系面临前所未有的“科林格里奇困境”。这一困境的本质在于技术迭代的指数级增速与治理体系线性演进之间的结构性矛盾，具体表现为数据采集的实时性与隐私保护立法的滞后性、算法决策的复杂性与行政问责的透明性要求、技术应用的涌现性与风险分级体系的刚性框架之间的多重张力。为破解这一困境，需要构建一个贯穿短、中、长期的多层次敏捷治理体系。

(一) 重点建立快速响应机制

通过实施四色分级预警体系，针对不同风险等级制定差异化的响应策略：对红色等级的公共安全事件实行 30 分钟内强制干预，对橙色等级的群体歧视风险启动 72 小时算法审计流程，对黄色等级的数据泄露风险实施 7 日内合规检查，对蓝色等级的潜在伦理冲突则采取常态化监测。同时，在雄安新区、深圳前海等试点区域开展预测性立法试验，通过监管沙盒机制测试《生成式 AI 禁用场景清单》[10]和动态合规标准，为全国性立法积累经验。这些措施需要配套建设 AI 风险监测平台，并修订《生成式 AI 服务管理暂行办法》以明确分级响应条款。

(二) 工作中心转向治理技术夯实

一方面要推进国家级智能算力网络建设，按照“中心–边缘–终端”三级架构配置资源，确保到 2025 年基层算力覆盖率达到 70%以上；另一方面要建立多级压力测试体系，在国家 AI 安全实验室开展常规、极端和灾难三级场景模拟测试，要求关键领域 AI 模型必须通过抗干扰性 ≥ 90%、恢复时间 < 5 分钟等严格标准才能投入使用[11]。这些措施需要与“东数西算”工程相衔接，并制定《公共算力使用管理办法》予以保障。

(三) 目标：构建多元共治的治理生态

通过设立数字权利保障基金，按照 AI 企业营收的 0.5%提取资金，用于数据侵权赔偿和算法歧视救济，同时成立由多方代表组成的数字权利仲裁委员会来管理基金使用。此外，建立“1 + 3 + N”模式的算法治理委员会，整合中央部门、技术机构和公民代表的力量，通过季度联席会议、技术评估报告和算法听证会等形式实现协同治理。这些制度创新需要修订《数据安全法》并制定《算法治理委员会工作条例》来提供法律支撑。

这一“三步走”敏捷治理战略，通过分级预警控风险、算力基建强基础、多元共治建生态的递进式路径，最终目标是实现技术可控性、治理灵活性和社会包容性的有机统一。确保短期措施为中长期改革奠定基础，同时通过持续监测评估机制及时调整治理策略，形成适应生成式人工智能发展特性的动态治理范式。

参考文献

[1] 人工智能全球治理上海宣言[N]. 人民日报, 2024-07-05(04).

[2] 高举中国特色社会主义伟大旗帜 为全面建设社会主义 现代化国家而团结奋斗——在中国共产党第二十次全国代表大会上的报告(2022 年 10 月 16 日) [M]. 北京: 人民出版社, 2022.

[3] 中央政府门户网站. 中共中央 国务院印发《数字中国建设整体布局规划》[EB/OL]. 2023-02-27. https://www.gov.cn/xinwen/2023-02/27/content_5743484.htm, 2024-07-10.

[4] 中央政府门户网站. 生成式人工智能服务管理暂行办法[EB/OL]. 2023-08-30. https://www.gov.cn/gongbao/2023/issue_10666/202308/content_6900864.html, 2024-07-10.

[5] 汪波, 周其鑫. 迈向整体智治的数字乡村治理: 实践逻辑、制约症结与转型策略[J]. 南昌大学学报(人文社会科学版), 2024(3): 113.

[6] 新华社. 联合国报告: 生成式人工智能加剧性别偏见[EB/OL]. 2024-03-08. http://www.news.cn/world/20240308/197475120cfc4efea698bd1738dd707c/c.html, 2024-08-20.

[7] 袁康. 可信算法的法律规制[J]. 东方法学, 2021(3): 5-21.

[8] 谢新水. 论数智社会人工智能诚信问题对公共秩序安全的冲击[J]. 浙江学刊, 2024(4): 109-119, 239-240.

[9] 钟祥铭, 方兴东. 数字鸿沟演进历程与智能鸿沟的兴起——基于 50 年来互联网驱动人类社会信息传播机制变革与演进的视角[J]. 新闻记者, 2022(8): 34-46.

[10] 中共中央关于进一步全面深化改革 推进中国式现代化的决定[N]. 人民日报, 2024-07-22(1).

[11] 中共中央办公厅国务院办公厅印发《关于加强科技伦理治理的意见》[J]. 中华人民共和国国务院公报, 2022(10): 5-8.