生成式人工智能嵌入数字政府的数据安全风险及其法律规制*

邹焕聪，王　庆

（南京工业大学　法政学院，江苏　南京　211816）

摘　要：生成式人工智能深度嵌入数字政府意义重大，但其引发的全周期数据安全风险不容忽视。通过解构数据“收集—处理—存储—输出”四阶段风险，发现收集阶段存在安全与质量隐患，处理阶段面临算法歧视与黑箱挑战，存储环节有数据泄露与归责困境，输出阶段则涉及数据失真与侵权风险。借鉴欧盟“先监管后发展”和美国“技术驱动立法”经验，提出分阶规制框架：收集阶段实行分类分级与源头审查，处理阶段强化伦理审查与清单管理，存储阶段构建软硬法协同标准与公私合作责任模式，输出阶段引入沙箱监管与数字确权。以此推动数字政府建设，实现科技、风险治理与安全法治的协同共进。

关键词：生成式人工智能；数字政府；数据安全；行政法治；法律规制

中图分类号：D922；TP399　　**文献标识码：**A　　**DOI：**10. 19358/j. issn. 2097 - 1788. 2025. 07. 008

引用格式：邹焕聪，王庆. 生成式人工智能嵌入数字政府的数据安全风险及其法律规制［J］. 网络安全与数据治理，2025，44(7)：50 - 57.

Generative artificial intelligence embedded in digital government's data security risks and legal regulation

Zou Huancong, Wang Qing

(College of Law and Politics, Nanjing Tech University, Nanjing 211816, China)

Abstract: Generative AI's deep integration into digital government is significant, but its full-cycle data security risks cannot be ignored. Risks in the four stages of "collection-processing-storage-output" include: collection security/quality issues, processing-stage algorithmic bias/black boxes, storage leakage/accountability gaps, and output distortion/infringement. Drawing on EU "regulate-first" and US "technology-driven" approaches, a phased framework is proposed: classification/source review for collection, ethical audits/list management for processing, soft-hard law collaboration/public-private risk sharing for storage, and sandbox regulation/digital rights confirmation for output. This promotes digital government development and balances technology, risk governance, and legal safeguards.

Key words: generative AI; digital government; data security; administrative rule of law; legal regulation

0　引言

数字政府建设是“十四五”期间乃至2035年之前国家治理的核心任务[1]。数字政府借助人工智能技术赋能突破传统行政模式的效能瓶颈，在全面深化行政体制改革、推进国家治理体系和治理能力现代化，增进民生福祉等方面具有重大意义。然而，生成式人工智能的深度嵌入正引发“技术赋能悖论”——该技术虽在政务智能问答（如北京“京京”系统响应准确率达92%）[2]、政策模拟推演（如宜宾经济监测平台实现GDP预测误差率<3%）等场景展现应用优势，其引发的数据安全风险却呈现全周期扩散特性。

我国虽已初步构建由《网络安全法》《数据安全法》《个人信息保护法》组成的数据规范法律体系以及《互联网信息服务算法推荐管理规定》《互联网信息服务深度合成管理规定》《生成式人工智能服务管理暂行办法》（以

*基金项目：国家社会科学基金一般项目（18BFX055）；江苏省高校哲学社会科学研究重点项目（2018SJZDI035）

下简称《暂行办法》）等共同组成的生成式人工智能法律规制体系[3]，但在具体细节与衔接上存在不足之处，形成“技术赋能”与“规制真空”并存的结构性困境。

数据为生成式人工智能的运行提供“原料”，同时也是人工智能开发运用与发展的“基石”[4]。既有研究多聚焦人工智能伦理或政府数字化转型宏观路径，对于生成式人工智能技术特性与数据治理的适配性研究存在三重局限：风险识别缺乏全周期框架、规制依赖事后模式、责任分配拘泥于政企二元结构，尚未充分回应生成式人工智能技术特性与数据治理的适配性问题。

基于此，本文创新性采用技术规制理论解构生成式人工智能的数据风险，通过考察多个政务智能化项目的合规性案例，提出“技术归化法律”的新型治理范式。该研究为破解技术赋能与数据治理的价值冲突提供实证支撑，实现科技赋能与风险规避的逻辑统一，将我国数字政府的建设和发展纳入法治轨道。

1　生成式人工智能嵌入数字政府的数据安全风险样态

1.1　数据收集的安全风险

生成式人工智能大模型的应用，依赖于海量数据的学习与迭代。当其深度嵌入数字政府建设进程时，无论是依托人机互动为基础的问答式服务，还是基于数据分析搭建的辅助决策模型，数据收集都将成为各领域无法绕开的核心环节。然而，数据安全和数据质量构成收集阶段的双重困境，成为阻碍生成式人工智能落地应用的瓶颈。

1.1.1　数据收集过度的隐患

在政府传统的履职过程中，行政机关对于办理事项所需数据材料都会履行严格告知义务，并采取公示原则，执行标准化的办公流程[5]。但是在生成式人工智能嵌入后，除行政相对人自行提供的数据外，可能会“未经告知”“未获同意”就对其社交数据、位置数据等进行收集分析。有学者所言“收集的数据越多，互动的程度越高，用户的体验感也会越强，但这种几乎无止境的收集方式对于现行数据安全规范而言无疑是巨大的挑战。”[6]对于这种收集方式，部分地方政府以“提升城市治理智能化水平”这一模糊表述作为收集合法性的依据，未就数据收集与具体治理目标的关联性进行充分论证、作出明确的解释。这一做法实质上违背了《个人信息保护法》确立的“目的限定”原则，而数据收集目的正当性模糊，势必给数据安全埋下隐患。

换言之，即便平台在与行政相对人互动时进行了明确的告知、获得了对方的同意，也仍然无法扭转“数据脱离支配”的状况。原因在于生成式人工智能模型能通过信息主动爬取来加重数据采集风险，这是生成式人工智能的原生风险。从技术原生风险传导至数字政府应用场景，使得生成式人工智能模型在收集数据时催生出极为强烈的“数据饥渴”效应，导致过度收集问题。这种过度收集不仅增加了相对人负担，还严重违反了“最小必要”原则，使个人、企业、国家的数据安全立于危墙之下。

1.1.2　数据收集的质量风险

政务数据作为最优质的数据资源，蕴含着丰富的经济价值、社会价值、政治价值，是数字政府建设重要的数据来源[7]。生成式人工智能技术嵌入数字政府的“表现”，在很大程度上取决于“喂养政务数据”的质量。从行政效益视角出发，只有质量可信的数据才能提高生成式人工智能模型的准确性和可靠性，提升生成式人工智能嵌入数字政府的应用效能，而一旦相关政务数据质量遭到破坏就会影响生成式人工智能输出内容的准确性和可靠性，从而制约政务服务的效果，误导政府决策，扰乱国家、市场和社会秩序。

具言之，一方面，低质量数据需投入大量人力、算力进行清洗与校验。在数据收集阶段，有相当一部分的资源用于数据校验，严重挤占了数据处理的投入，显著增加了数据治理成本。另一方面，低质量数据会引发数据污染，污染传导的链式效应会让初始数据通过模型训练产生的风险成倍增加。总而言之，数据收集的质量风险不仅与数字政府提高资源配置效率的目标背道而驰，削弱了数字政府建设的社会效益，而且从根源上危及了数字政府建设的总体安全。

1.2　数据处理阶段的数据安全风险

在生成式人工智能技术应用中，数据作为基础要素，为整个系统提供支撑。运用一系列技术方法对数据进行重构，通过复杂的数学模型与逻辑规则，深度剖析数据的内在关联，将原始数据转化为有意义、具决策价值的信息，驱动生成式人工智能嵌入数字政府达成预期效用。

1.2.1　算法歧视导致决策“不公平”

在算法的不断优化中，当输入数据存在歧视倾向，结果就会产生不公平的判断[8]。从行政法的基本原则出发，行政主体在履职过程中须对行政相对人保持平等对待，确保决策与行为的无差别性及非歧视性。在数字政府情境下，政府通过生成式人工智能提供政策咨询、福利分配、决策辅助等服务，传统行政法上的主体范畴由“行政机关—行政相对人”转向“智能系统—行政相对人”，如果算法存在歧视，算法会通过迭代将这种歧视固化为“最优决策模式”，形成“数字马太效应”，导致强者愈强、弱者愈弱的现象，进一步加剧社会不平等。从

行政公平理论审视，这样的算法歧视实质是算法行政权力对公民平等权的侵蚀[9]。算法决策看似“价值中立”的背后，实则凭借数据编码方式重构了行政权力的运行逻辑。

1.2.2 算法黑箱挑战行政“透明度”

生成式人工智能的算法黑箱将行政公开的程序蒙上了一层“阴影”[10]。算法黑箱是指生成式人工智能内部运作机制，如处理逻辑、决策过程，对用户而言是不透明或无法解释的现象。但行政程序的公开性则要求行政行为作出的过程是公开透明的，即“公平必须公开地、在毫无疑问地被人们所看见的情况下实现。”[11]在生成式人工智能嵌入数字政府过程中，算法黑箱使得数字政府的行政决策变得不可预测。公众只能看到最终结果，无法看到得出过程，决策程序完全处于闭环黑箱之中，造成“行为—结果”因果关系的断裂。实践中，政府部门常以“国家安全”“商业秘密”为由援引《数据安全法》第21条的规定以豁免披露关键参数。换言之，即便公布了算法程序，因其运行的高度专业性，公众也难以理解其逻辑架构，使行政公开制度沦为浮于表面的形式化流程。公开的法定程序转换到了技术后台，反而为权力和资本的结盟提供了温床，公众对强大技术的畏惧替代了政府信任[12]。这种信任危机一旦产生，将对政府的各项工作带来极大的掣肘，威胁到整个行政机关的公信力与权威性。

1.3 数据存储阶段的数据安全风险

数据存储作为数据生命周期的核心环节，既起到了先前数据收集、数据处理的基础性支撑作用，又为后面生成应用环节奠定基础。在数字政府建设中，因治理逻辑与智能时代数据特性的错位，本阶段的风险集中表现为数据存储泄露风险与存储责任界定风险。

1.3.1 数据自身特性引发存储泄露风险

在数字政府场景下，数据敏感性上升，存储环节的安全隐患已然从“潜在风险”演变为“现实危机”。一方面，存在技术承诺与现实的巨大鸿沟，生成式人工智能服务的提供者，一般会在软件程序中作出相关声明，同时于用户协议里设定前置性条款，其目的在于向用户承诺，将借助匿名化处理、数据加密等技术手段，对用户数据实施安全防护措施。但现状是，生成式人工智能模型的确存在存储数据泄露风险[13]。这些存储于平台的数据持续面临“数据投毒”“数据篡改”等风险，攻击者可通过植入恶意数据污染训练集，或篡改核心数据干扰政务决策，使数据存储系统沦为安全攻击的目标。另一方面，技术特性与政务安全原则存在根本性冲突。在生成式人工智能完成信息的处理并给出回应之后，信息的利用已然终止。然而，信息实际上依旧存在于智能系统的数据库内，不管是通过云存储还是分布式存储的方式，都能找到这些信息留存的迹象。生成式人工智能的持续学习特性要求存储系统保持“可写入”状态，与政务数据“归档即锁定”的安全原则产生根本冲突，这种由数据特性引发的矛盾为数据篡改、越权访问等攻击打开了“后门”，使得数据在存储阶段面临更高的安全风险。

1.3.2 多元数据主体导致存储责任风险

政府作为政务数据的法律所有权主体，在与服务商签订存储合同时，实际让渡了数据存储的技术控制权，而数据服务商实则掌握数据存储的物理基础设施、底层架构及运维权限。在数据所有权与技术控制权分离的背景下，政府因技术能力缺失难以有效监管，服务商则以协议条款规避主体责任，导致存储风险归责不清。

在搭建生成式政务服务平台过程中，庞大的数据库将被科技公司所掌控，敏感数据存储于私有云。与此同时，行政相对人在这一过程中处于信息严重不对称的弱势地位。行政相对人对于个人信息后续被存储在何处、由谁管理、采取了哪些安全措施等问题几乎一无所知[14]。在这种“政府—服务商—公众”的三角关系中，公众既无法参与数据存储的决策过程，也难以获得有效的权利救济，形成“数字弱势群体”。基于技术依赖与权责失衡的治理困局，本质上是数字时代公共治理能力与技术发展速度脱节的集中体现，“政府监管失能、服务商责任规避、公众权益受损”的恶性循环，表明传统二元治理结构已难以应对新型安全挑战。

1.4 数据输出阶段的数据安全风险

在生成式人工智能嵌入政务系统的最终输出环节，数据输出阶段的技术特性与公共治理的合法性、规范性要求之间存在显著张力，导致虚假内容对公共决策的威胁与输出内容引发的知识产权侵权隐患这两大风险。

1.4.1 输出“虚假内容”威胁公共决策

生成式人工智能凭借其对大规模训练数据分布特征的学习与概率建模能力，展现出强大的内容创作效能，但在政务场景下存在“数据智能”向“数据失控”转变的隐患。当生成式人工智能面向用户提供所需信息时，其输出内容的客观真实性与逻辑自洽性无法得到根本保障。诸如ChatGPT之类的典型生成式人工智能，在应对中文知识、日常常识以及学术领域的问答任务时，时常会出现输出内容与事实不符的情形。虚假内容的持续输出会引发公众的信任危机，形成“真相稀释”的效应。例如某市“智慧信访”系统将方言表述“要个说法”错误关联至司法诉讼场景，生成建议信访人提起诉讼的错误指引。至于生成错误是否构成行政违法、是否威胁公

共决策、政府在此过程中应承担何种决策责任等问题，目前法律规范尚缺乏具体规定。

1.4.2 输出内容“知识产权”侵权风险

在数字政府的场景中，生成式人工智能输出内容的知识产权侵权风险，本质上是技术创新对传统知识产权法律框架的多维冲击。在我国现行知识产权法框架下，知识产权的主体通常为自然人或法人，而这一点是基于创作行为通常以人类的智力成果和独创性为前提[15]。当涉及生成式人工智能生成的传播内容时，内容的“独创性”认定存在争议。因为生成式人工智能凭借算法处理以及大数据分析手段产出作品的过程与传统人类创作模式显著不同，人类创造性要素的介入程度较低。若大模型平台利用政府资源训练生成式人工智能并生成政策分析报告、统计图表等，那么其法律属性应界定为“职务作品”还是“法人作品”？对此目前尚无法律结论，由此导致权利主体陷入“虚置”状态。

总的来说，本文重点就四个阶段的典型安全风险进行了分析，但是该阶段的典型安全风险往往并不为该阶段所独有，也可能存在于其他阶段，比如数据泄露风险可能同样存在于数据收集、处理、输出等各阶段。限于篇幅，本文仅就典型阶段的代表性风险进行分析，以便为相应的法律规制路径选择提供研讨基础。

2 生成式人工智能嵌入数字政府的域外立法及启示

2.1 欧盟：先监管后发展的稳妥构想

欧盟在人工智能监管与数据安全保护领域构建了全球领先的治理体系。2024 年正式生效的《人工智能法案》作为区域内首部全链条综合性立法，通过构建系统性规则框架，为欧盟人工智能产业的合规稳健发展提供了制度支撑。该法案针对人工智能的风险、责任、安全以及透明度等方面作出了详尽规定，并将人工智能应用划分为无风险、有限风险和高风险三个层级[16]。不同风险等级的人工智能应用对应差异化的规制策略：无风险应用仅需满足最低合规要求；有限风险场景下，服务提供者需履行基础透明度义务；而在关键基础设施、教育、就业等敏感领域的高风险技术系统，则需接受第三方合规评估、持续监测及全生命周期审查。

在立法过程中欧盟尽力摒弃传统回应型立法模式，构建以法律规范为核心、监管框架为支撑的风险防控体系。这与该区域一贯遵循的“先管控技术安全风险、再推动产业创新发展”的治理逻辑高度契合。这样的路径探索，既实现了对新兴技术潜在风险的前瞻性应对，又为产业发展预留了必要创新空间，为我国生成式人工智能赋能数字政府的数据风险法律治理提供了重要参考范式。

2.2 美国：数字政府与人工智能的融合推进

在人工智能监管策略层面，美国则独树一帜，以开放之姿推进治理。美国深知人工智能技术对于维持自身全球领先地位的关键意义，因而极力避免过度干预，力求为技术的蓬勃发展营造宽松自由的环境，鼓励科研创新与产业孵化。

2023 年 5 月，美国国会研究处发布《生成式人工智能和数据隐私：入门指南》，聚焦生成式人工智能的数据运用与潜在隐私风险剖析，针对性提出一系列对策。在数据收集环节，构建通知与披露体系，强制要求开发者在采集或使用个人数据前，务必征得数据主体的明示同意，并如实告知数据用途，确保数据主体知情权；关于数据存储，明晰删除与最小收集准则，赋予用户从当前数据集中彻底删除自身数据的绝对权利，同时限定数据保留的最短时限，最大程度保障个人数据安全，减少隐私泄露隐患；在数据生成方面，虽未出台类似欧盟那般详尽的专门法案，但通过版权法等现有法律体系，结合技术手段为生成内容添加数字水印等标识，以明确版权归属。

2.3 域外立法经验对完善我国法律规制的启示

我国现行的数据安全法律体系主要是在传统互联网的视角下构建的，随着人工智能技术的迅猛发展，数据保护面临着新的挑战。

在数据收集阶段，欧盟《通用数据保护条例》（GDPR）构建的“目的必要性原则”和“数据最小化 + 分类分级”双重机制，为破解数据过度收集的安全隐患提供解决方案。一方面通过“目的必要性原则”以法律强制力划定个人数据收集红线，坚决避免过度收集、随意扩大收集范围情况发生。另一方面，通过数据分级分类为生成式人工智能设立行为边界，平衡其与数字政府建设相结合过程中可能产生的价值冲突[17]。

在数据处理阶段，欧盟《数字服务法》（DSA）针对数据处理推荐算法构建了较高的透明度标准框架，重点强化用户权益保障与外部监督机制。法案赋予用户对平台内容审核决定的申诉权，同时允许合规研究者依法获取平台核心数据，为在线内容风险研究提供支持，有效抑制算法歧视现象，确保算法运行的公正性。同时，美国通过《算法问责法案》建立“算法审计师”认证体系，组织第三方专业研究者对高风险算法进行深度审查，包括审查的范围界定、方式选择及标准设定等。

在数据存储阶段，欧盟 GDPR 的数据存储标准要求对数据存储权限进行限定，一旦达成目的或存储不再必要，应当立即删除或匿名化处理数据。美国没有形成统

一的数据存储标准，而是通过分散式法规实现特定领域的重点监管。例如在跨境数据层面形成“受管控非密数据列表”，涉及国家经济、政府管理、敏感技术等数据，严格控制出境。在存储责任承担方面，美国曾推行服务商自我认证与政府监督相结合的公私合作模式，鼓励企业通过合规承诺履行存储责任；同时在部分行业法规中，明确要求企业作为数据存储主体，必须采取合理措施保障数据安全，承担因存储不当引发风险的法律后果。

数据生成阶段，欧盟《人工智能法案》等相关法律对数据真实性作出规定，确保数据准确、完整且及时更新，从源头上降低数据失真风险。美国则通过《深度伪造责任法案》，对利用生成式人工智能造成严重后果的，确立“生成即担责”原则。在知识产权保护领域，欧盟《人工智能法案》明确人工智能模型的提供者是著作权责任承担的主体。美国知识产权学界则提出扩大演绎作品、雇佣作品的认定范围等学说，以解决人工智能生成物版权归属问题。

通过借鉴欧盟的严格监管框架与美国的技术创新导向，我国可在保障政务生成式人工智能安全性的同时，培育具有中国特色的数字政府治理范式。

3　生成式人工智能嵌入数字政府的法律规制路径

本节针对风险样态中提出的四个问题予以回应规制，打造一个数据风险立体防控体系。

3.1　数据收集的安全风险规制

3.1.1　以“分类分级”实现数据安全保证

当前，《暂行办法》第3条确立了包容审慎监管原则与分类分级监管机制，为数字政府应用生成式人工智能提供了宏观指引，2024年10月，《数据安全技术数据分类分级规则》标准正式实施，其中对数据的分类分级已经有较为详细的规定，但针对生成式人工智能嵌入数字政府特殊场景的搭建还有所欠缺，因此，依据《数据安全法》和《个人信息保护法》框架建立政务数据的分类分级规则尤为重要。一方面，通过数据分类分级处理，可以为生成式人工智能模型输入高质量、标准化数据；另一方面，也能防止生成式人工智能借“政务服务”或“社会治理”之名行过度收集之实。为此，可以根据此类数据的特点，利用分类规则进行风险识别，并采取不同级别的应对措施，保证数据收集的合法、合规。同时，数字政府可以以数据分类分级为依托，制作数据收集的“目的清单”，确保每一次数据收集行为都有明确的目标指向，避免目的的模糊性和随意性。《浙江省公共数据条例》对此已有成功经验，其依托数据分级，建立“禁止采集负面清单”，该清单明确列出禁止收集的数据类型和范围，防止政府部门过度收集敏感数据，使生成式人工智能在政务领域的数据收集行为有法可依、规范有序。

3.1.2　以“源头审查”化解数据质量困境

我国在《促进新一代人工智能产业发展三年行动计划（2018—2020年）》中明确指出要加强高质量的标准测试数据集和训练数据集建设，这为数字政府数据治理锚定了方向[18]。针对收集数据过程中耗费大量资源校验清洗和数据污染等风险持续威胁数据安全的问题，需要改变事后被动应对的做法，将数据质量审查防线向源头迁移，实施数据风险的“敏捷治理”。

通过源头审查机制，实现治理成本的显著降低与数据应用效益的最大化提升。基于此，可以专门搭建数据质量管理平台，对外部收集数据实施严格的准入管控，从源头上阻断数据质量的冲突隐患。通过建立数据质量核验任务，自动完成数据质量规范性、一致性、准确性和完整性的检查[19]。通过建立全流程的闭环化监管体系，定位、分析、跟踪及解决数据质量问题，形成数据质量闭环化的处理机制，保证数据质量持续稳定可靠，有效遏制数据污染、数据投毒，保障数据来源的合法性与安全性。

3.2　数据处理的安全风险规制

3.2.1　以“伦理审查”构建算法公平标准

“伦理责任的实现不仅依靠内在的道德力量，而且必须将这种力量外化为一种制度安排，这样才能使道德之花结出丰硕的技术‘善’果。”[20]数字时代的行政法治，本质是技术治理与法律治理的深度耦合。为了解决算法歧视导致的决策“不公平”，需要借鉴美国《算法问责法案》中的“算法审计”构建具有本土特色的算法伦理审查。伦理审查可以作为二者的转换器，通过制度化价值校准机制，建立以平等原则为核心的价值评价体系。一是技术维度，将伦理设计嵌入针对数字政府视角下的生成式人工智能平台。一项数字技术是否符合特定的国家、社会所要求的伦理和价值观，是推广和应用该技术的重要标准。我国的民族伦理集中体现于社会主义核心价值观，这一价值体系为生成式人工智能算法在行政领域的应用划定了伦理红线。在具体实践中，严格对标社会主义核心价值观，进行伦理审查。明确禁止将性别、种族等受保护特征作为算法决策变量，对地域、收入水平等衍生性敏感特征设置比例原则审查机制，要求算法证明这些特征与决策目标之间存在实质关联性，且对特定群体的影响控制在合理范围内。

二是制度维度，设立“算法公平审查委员会”，成员

汇集法学专家、计算机专家和公众代表，搭建起预防算法歧视的坚实制度防线。首先，由法学专家从法律合规性角度审查算法是否违反现有法律法规，确保算法决策全程符合法定框架；其次，依托计算机专家的技术专长对算法的设计逻辑、运行机制进行分析，判断算法是否存在技术层面导致的歧视风险；最后，从社会公众利益出发，反映不同群体的诉求和关切，保障算法决策的公平性能够切实回应社会需求。

3.2.2 以“清单管理”重塑算法透明度

我国《暂行办法》引入算法备案制度，明确要求涉及舆论属性导向或社会动员能力的服务主体履行备案手续，通过备案信息公示和风险评估强化监管。这一制度为人工智能监管提供了重要抓手，但具体运用于数字政府场景，在透明度和可解释性方面仍缺乏明确具体的指引，而“清单管理”作为一种系统化、精细化的治理工具，通过明确算法边界、规范操作流程，将算法决策从隐秘的技术后台拉至公开透明的法治轨道，为破解行政透明度难题提供了创新路径。

从制度衔接角度看，编制“算法透明度标准清单”能够有效弥补《暂行办法》在数字政府应用中的不足。一方面，通过“清单”列举明确规定行政机关在使用生成式人工智能时必须公开算法的核心原理、决策逻辑，避免算法成为脱离监管的“法外之地”。同时，缩小“国家安全”“商业秘密”等豁免条款的适用范围，建立严格的豁免审批程序。

另一方面，“清单”起到限权的关键作用。在强调算法公开的同时，明晰算法边界，合理界定需要披露的内容。设法避免不必要的算法披露，不盲目扩大算法的披露内容，防止法律滞后对创新活力的窒息，十分契合我国对于人工智能“包容审慎”的治理态度。此外，应该对数字政府下行政机关增设算法解释义务，行政机关在依据算法作出决策后，向行政相对人公开说明算法决策的依据以及考虑的因素，确保行政相对人能够理解决策的形成机制，增强决策的可接受性的同时也维护了行政机关的公信力。

3.3 数据存储的安全风险规制

3.3.1 形成“软法”和“硬法”协同的存储标准

鉴于由数据敏感易泄露导致的安全风险不断升级，尝试以技术“软法”与制度“硬法”二者协同互补，构建高效数据存储安全标准。

一方面，通过“硬法”划定数据存储安全的刚性边界。《数据安全法》第 27 条明确了数据处理者应建立健全全流程数据安全管理制度，因此，在平台搭建伊始，各类平台应以“硬法”标准在技术承诺部分对取得许可的核心数据、重要数据和一般化数据的差异化存储进行管控。同时，国家网信、公安、国安等部门被赋予明确的监管职权，通过定期检查、专项整治等手段，对数据存储主体保障数据安全状态进行监督，确保数据存储安全责任的落实。

另一方面，以“软法”填补数据存储治理的柔性空间。“软法”涵盖行业标准、技术指南、自律公约等非强制性规范，以灵活的方式填补“硬法”在适应性和精细化方面的不足。具体而言，根据《数据安全法》第 21 条规定的数据分类分级制度，应针对政务数据存储场景制定差异化的技术规范，制定严格且统一的政务数据存储技术安全标准，涵盖数据加密、访问控制、存储系统架构等关键方面；规定数据存储必须采用高强度加密算法，且加密密钥的管理需遵循严格的规范，定期更新密钥以提升数据保密性。与此同时，在存储系统管理中，构建清晰可溯的访问控制规则体系，实现数据访问权限的精细化管控。及时发现并阻止非法访问行为，对违反技术安全标准的企业或机构，监管部门可先依据“软法”要求其限期整改，若拒不执行则启动“硬法”处罚程序。

3.3.2 构建“公私合作型”存储责任承担模式

在生成式人工智能嵌入数字政府的框架下，数据存储的安全风险不再是服务商一方或者政府一方的责任，而是形成了“责任共同体”，公共部门与私人部门应合作应对各类风险[21]。为此，针对传统政企二元对立结构的局限性，需要在数据存储阶段构建公私合作型的责任承担模式，将“风险共担、利益共享、权责对等”作为核心理念，并通过合同契约、合作治理机制等方式，明确多方在数据存储全生命周期中的责任或义务，形成权责清晰、协同高效的治理体系。

首先，政府责任应实现由“技术弱势者”到“战略主导者”的转变。一方面，要采用“技术控制力 + 数据利益”的双重标准划分存储责任，依据《网络安全法》第 37 条规定，政府保有核心数据控制权，确保数据主权不受侵蚀。另一方面，要重构公私部门的责任分配框架，政府部门作为政务数据的所有者，在生成式人工智能运行过程中对数据存储安全负有担保责任，参照《民法典》第 1194 条有关网络服务提供者责任的规定，由企业承担数据存储运维责任。

其次，私主体责任应完成由“责任规避者”到“责任履行者”的转变。为避免生成式人工智能服务提供商以协议条款规避责任，公私合作型存储责任承担模式除了私主体承担前述数据存储运维责任之外，还需建立明确的私主体责任追溯机制，在合同中应细化服务商的违约责任，对因技术失误、管理不善导致的数据泄露等事

故，设定赔偿标准，并要求服务商购买数据安全责任保险，增强风险应对能力。

最后，公众责任应完成从“被动受害者”到“权力参与者”的转变。公私合作型模式致力于打破公众在数据存储中的信息不对称局面，政府可以通过官方网站、政务 APP 等渠道，向公众公开数据存储的位置、管理主体、安全措施等关键信息。这样既能拓宽公众参与的渠道，优化公众参与的形式，提高公众参与的意识，也能让公众更好地了解生成式人工智能技术在政府治理中的应用，切实增强数字政府治理的透明度与公信力。

3.4 数据输出阶段的安全风险规制

3.4.1 聚焦输出内容“监管沙箱”的引入

如前所述，由生成式人工智能导致的虚假内容的输出，已经引发人们对公共决策合法性的担忧。为此，应引入“监管沙箱”机制，通过场景化测试、动态化监管，最终实现法律规则的适配。即通过立法授权创设“规制实验区”，允许政府在可控范围内测试生成式人工智能的政务应用，同步完善法律规则。

在《暂行办法》框架下，可将信访、政策解读、公共安全预警等高风险场景纳入沙箱监管，要求服务商与政府部门联合申报实验方案，明确测试目标、风险预案及法律责任等事项。通过沙箱机制的事先运行，将法律模糊地带转化为规则生成场景，为技术缺陷设定合法容错空间，确保生成的内容符合有关法律规范对“真实、准确”的要求，防止风险外溢至公共决策核心领域。

通过沙箱运行发现输出阶段的共性问题，可以为法律、法规、规章等的修改提供实践依据。如在《数据安全法》中对政务数据进一步细化规定，增设政务应用特别条款，明确生成式人工智能输出内容的行政合法性标准；在《政府信息公开条例》中增加“生成式人工智能生成信息标识义务”等条款，规定所有政务场景输出内容需标注“生成主体”“校验状态”等信息，增强信息的透明度、真实性以及可追溯性。监管沙箱的引入既为生成式人工智能在政务场景中的创新应用保留了试错空间，又通过场景化规则生成填补了“数据失控”的治理真空。

3.4.2 针对输出内容“数据确权”的展开

由于生成式人工智能输出内容存在权利主体“虚置”的现象，政府在数据资产利用、公共服务供给中面临侵权隐患与合规困境的安全风险。对此，应构建以“数据确权”为核心的法律规制体系，为平衡技术创新与权利保护探索有益的路径。数据确权的实质在于穿透生成式人工智能创作的技术迷雾，明确“数据控制权”与“创作主导权”的归属。当生成式人工智能嵌入数字政府时，训练数据、算法参数与输出内容本质上是政务数据资产的智能化转化，应当可以明确政府作为“数据控制者”的主体地位。在数据权属层面，修订《著作权法》或《著作权法实施条例》，增设“算法生成作品”条款，并明确在政务场景下若政府在生成式人工智能训练与内容生成中发挥实质性主导作用，且生成成果用于公共服务，应认定为“法人作品”；若公职人员仅辅助性使用生成式人工智能工具，可参照“职务作品”处理。通过数据确权的法律规制，最终实现政务生成内容从模糊到明晰的转变。

4 结束语

在数字政府转型建设进程中，生成式人工智能是一把双刃剑，既能作为驱动政府治理创新的核心赋能工具，成为智能行政的重要助推力量，又带来了诸多数据安全风险与法律规制挑战。本文着眼于数据四大阶段的典型安全风险样态，提出分阶段治理框架，但是需要指出的是，各阶段典型安全风险与法律规制路径并非绝对的一一对应关系，而是“一把钥匙开几把锁的关系”。例如，在数据存储阶段构建“公私合作型”责任承担模式，同样可以将此方法推广应用至风险样态的其他三个阶段。

数字政府建设是一个长期且复杂的过程，生成式人工智能技术也在不断迭代更新。未来，需要持续关注技术发展的动态，不断完善法律规制的体系，强化多方合作治理，在保障数据安全的基础上，充分释放生成式人工智能在数字政府建设中的巨大潜能，实现科技赋能与风险防控的良性互动，助力国家治理体系和治理能力现代化建设。

参考文献

[1] 程圆圆. 生成式人工智能嵌入数字政府的技术路径、潜在风险与制度规制 [J]. 昆明理工大学学报（社会科学版），2024，24（6）：19－28.

[2] 关鑫，魏琨翔，姜莹. 生成式人工智能嵌入数字政府建设：内在机理与路径优化 [J]. 辽宁行政学院学报，2024（5）：44－49.

[3] 刘霜，张潇月. 生成式人工智能数据风险的法律保护与规制研究——以 ChatGPT 潜在数据风险为例 [J]. 贵州大学学报（社会科学版），2023，41（5）：87－97.

[4] 林伟. 人工智能数据安全风险及应对 [J]. 情报杂志，2022，41（10）：105－111，88.

[5] 谷飞，傅建平. 生成式人工智能嵌入数字政府的风险结构及其回应——以风险社会理论为视角 [J]. 中共福建省委党校（福建行政学院）学报，2024（4）：77－93.

[6] 李振林，潘鑫媛. 生成式人工智能背景下数据安全的刑法保护困境与应对——以 ChatGPT 为视角的展开 [J]. 犯罪研究，2023（2）：25－33.

[7] 卢荣婕. 生成式人工智能赋能政务服务智能化建设的价值、困境与治理 [J]. 西华师范大学学报（哲学社会科学版），2025（3）：51－61.
[8] 刘辉，雷崎山. 生成式人工智能的数据风险及其法律规制 [J]. 重庆邮电大学学报（社会科学版），2024（4）：40－51.
[9] 江必新，王鑫. 数字行政行为算法歧视的法律规制 [J]. 学术论坛，2024（6）：49－63.
[10] 刘银喜，吴京阳. 生成式人工智能嵌入政府治理的应用前景、潜在风险和防范机制 [J]. 北京航空航天大学学报（社会科学版），2025，38（1）：103－112.
[11] 雷刚. 数字政府时代算法行政的程序法治研究 [D]. 重庆：西南政法大学，2022.
[12] 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.
[13] 苏子龙. 生成式人工智能的数据安全风险防控与法律规制研究 [J]. 通信与信息技术，2024（5）：95－98，110.
[14] 张佳琳. ChatGPT 模型辅助数字政府建设的风险及其法律规制 [J]. 内蒙古社会科学，2024（1）：65－75.
[15] 刘艳红. 生成式人工智能的三大安全风险及法律规制——以 ChatGPT 为例 [J]. 东方法学，2023（4）：29－43.
[16] 李芳，刘鑫怡. 欧盟人工智能立法最新动向 [J]. 科技中国，2021（6）：35－38.
[17] 傅建平. 生成式人工智能引入数字政府建设的规制路径——基于风险社会的视阈 [J]. 湖南大学学报（社会科学版），2025，39（1）：124－132.
[18] 何振，彭海艳. 人工智能背景下政府数据治理新挑战、新特征与新路径 [J]. 湘潭大学学报（哲学社会科学版），2021，45（6）：82－88.
[19] 林伟，周耀铭. 国内外数据治理研究述评 [J]. 数字图书馆论坛，2022（6）：65－72.
[20] 谭九生，杨建武. 人工智能嵌入政府治理的伦理风险及其防控 [J]. 探索，2021（2）：126－138.
[21] 邹焕聪. 公私协力法律问题研究 [D]. 南京：南京大学，2011.

（收稿日期：2025－02－22）

作者简介：

邹焕聪（1974－），男，博士，教授，硕士生导师，主要研究方向：行政法学，经济行政法、数字行政法等。

王庆（1999－），男，硕士研究生，主要研究方向：刑法学、数字法治。

（上接第 35 页）

[10] 李点横. 基于物联网的教育资产管理系统设计与实现 [D]. 大连：大连交通大学，2021.

（收稿日期：2025－04－15）

作者简介：

安宁（1990－），男，硕士，高级工程师，高级经济师，主要研究方向：软件工程、云计算、人工智能、大数据、信息安全、数字经济等。

许文静（1990－），男，博士，讲师，主要研究方向：人工智能、大数据智能、数字教育、信息安全等。

刘珠慧（1992－），女，硕士，讲师，主要研究方向：信息化项目管理、信息安全等。

版权声明

凡《网络安全与数据治理》录用的文章，如作者没有关于汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权等版权的特殊声明，即视作该文章署名作者同意将该文章的汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权授予本刊，本刊有权授权本刊合作数据库、合作媒体等合作伙伴使用。同时，本刊支付的稿酬已包含上述使用的费用，特此声明。

《网络安全与数据治理》编辑部