大模型赋能政务治理：治理机遇与风险因应

建设数字政府是当前推进国家治理体系和治理能力现代化的内在要求与重要内容。随着生成式大模型产业化应用，尤其是国产大模型 DeepSeek 的成熟问世，政务数字化的技术壁垒取得显著性突破，垂直领域深度渗透，或能有效提升政务现代化水平。

文丨**王至辉 张亚鹏** 中国电子信息产业发展研究院新型工业化研究所

国产大模型突破与政务场景应用

DeepSeek 的突破和创新

人工智能技术预示着人类社会的一场革命。当前，DeepSeek 作为全球领先的 AI 大模型，其突破和优势体现在技术架构、成本控制、多模态能力及生态与本地化优势等多个维度。DeepSeek 技术不仅构成潜在生产力向现实生产力转换的驱动力，而且以前所未有的方式改变着政务治理。美国布莱恩·阿瑟（Brian Arthur）教授在《技术的本质》一书中指出，“技术以这样的方式不断地探索未知领域，不断地创造出进一步的解决方案和进一步的需求，因之而来的是持续不断的新颖性，整个过程就此呈现了某种有机性：新技术不断在旧技术之上衍生出来，其中创造和替换交叠着推进整个过程向前发展”。在技术架构创新方面，其混合专家模型（MoE）采用分层路由机制，显存占用空间不足传统模型的一成，显著提升了运行效率。他将一个超大的 AI 模型拆分成多个小模块，分布到多台计算机上同时处理或协同处理，单次任务处理速度比以前快 40%。采用边算边传，在处理数据的同时进行传输，避免等待时间，进一步提升了效率。其采用动态专家激活机制（Gating），根据输入特征动态选择专家模块，实现计算资源精准分配。在成本控制优势方面，其训练成本革命性降低，来自于 DeepSeek-V3 技术报告，其训练成本仅 557.6 万美元，是同类模型的 1/20（如 GPT-4 训练成本超 1 亿美元）。其推理成本极致优化，通过优化硬件使用时间（白天全力服务，晚上切换到训练研究）和缓存技术（56% 的输入数据直接从缓存读取），算力消耗降低了 37%。其定价仅为行业龙头的七分之一至二分之一，但效果相当。在多模态与效率突破方面，其多模态任务处理性能优越，支持文本、图像、语音的联合建模，例如医疗领域融合电子病历与影像数据生成诊断

建议。DeepSeek 具有高效推理引擎，文件处理效率提升 5 倍（如苏州占道经营识别），长沙市某医院信息部主任透露，截止 2025 年 2 月医疗诊断响应速度提高 40%。在长文本处理方面，通过双管道技术优化，复杂网页（含 PDF、手写体）语义提取效率提升 300%，支持政务网页秒级分类。在生态与本地化优势方面，其生态开源开放特质吸引了全球开发者持续参与优化。通过技术共享降低开发门槛，促进多方协作创新，加速技术迭代与应用落地。DeepSeek 适配多语言文化，如香港的 HKGAIV1 模型支持粤语、英语和普通话，注入本地知识库，成为首个面向海外华人的大模型。最重要的是其硬件兼容性突破，突破 CUDA 框架限制，实现英伟达和华为昇腾等多平台适配，为“算力封锁”提供了中国方案。

DeepSeek 的政务应用

DeepSeek 凭借其超高性能、超低成本、完全开源和适配全国产化硬件的特点，赋能政务治理全流程重塑，深刻重构传统政务治理范式，推动政府治理模式从传统科层制向智能化转型。信息技术的突破性跃进，使得 DeepSeek 得以嵌入政务系统，不仅能够增强决策对信息掌握的精准性和系统性，还能够通过深度分析政务数据，为政府精细治理提供更加有力的数据支撑。

治理精细化。北京市探索尝试为辖区企业提供全量模型服务，支持大模型开发及 API 调用。大模型整合多模态数据，提供智能决策支持和个性化公共服务，能够有效提升治理精准度。广州市将其应用于民生政策解读、12345 热线工单分派，并推出全国首个政务级安全算力 +DeepSeek 适配版，南沙区通过“南山通”小程序实现政务办事“语音智办”，覆盖近 1200 项公共服务。大模型通过全域数据感知和多情景策略创生能力，支持分散化、多中心的适应性治理结构。因此，深圳市拟在全市范围部署 DeepSeek 模型，实现政务应用一体化赋能升级，覆盖政务办公、城市治理等场景。在司法辅助方面，广东湛江司法局使用 DeepSeek 自动生成行政复议决定书，效率显著提升。天津静海区检察院借助 DeepSeek 辅助刑期研判，案件处理更精准。

决策高效化。南京部署全国首个应急管理大模型“宁安晴”，响应效率明显提升。深圳福田“安全生产助手”生成演练脚本，效率提升百倍。江苏无锡市部署 DeepSeek-R1-671B 全尺寸模型，其在交通领域构建智慧交通 AI 平台，融合多源数据提升安全治理效率。江西省赣州市完成全省首个地级市 DeepSeek 部署，公文处理效率相较之前显著提升。苏州市运用图像识别技术自动识别占道经营。在防返贫监测领域，广西玉林部署 DeepSeek 通过动态分析脱贫户数据，精准识别潜在返贫风险家庭，并自动生成参考性的帮扶建议，促进决策时效显著增强。

服务智能化。在城市治理领域，深圳政府在线 2025 年 2 月报道“一句话找人 / 找视频”功能基于多模态大模型，结合 23 万路视频监控，已成功找回走失人员

300 余次。在医疗健康方面，沧州市卫生健康委将 DeepSeek 与全民健康信息平台融合，优化就诊流程并制定健康管理方案。在民意速办方面，民生诉求智能匹配功能分析诉求情感倾向，实现工单自动分类分拨，提高工单处理精准度，减少重复工单。在融资增信方面，临沂市大数据局将“沂蒙慧眼”系统融合应用政府的公共数据为企业精准画像，赋能企业融资增信，有效破解企业“融资难、融资贵”问题。截止 2025 年 2 月已助力企业融资增信超过 33 亿元。

DeepSeek 接入政务场景的挑战

数据安全与隐私风险

目前大模型仍然还有很多亟待完全解决的风险点。大模型训练需依赖海量数据，若数据脱敏不彻底或访问权限失控，敏感信息可能通过模型输出泄露，或涉及公民隐私和国家安全。尽管大模型可以采用本地化部署，但存在数据被恶意攻击的潜在风险。另外，攻击者可通过“数据投毒”手段在训练阶段植入恶意样本，或造成政务数据污染，进而影响政务治理的有序开展。新近研究表明，研究人员张恩、高鹏程在《科学决策》发表题为《国家意识形态安全视域下生成式人工智能治理论析》指出，“生成式人工智能技术创新对现行治理规则的先行性突破，必然引发社会政治场域中多元主体在规制‘真空’中的策略性行动，这对意识形态安全带来危与机。”

算法偏见与决策偏差

生成式 AI 的“幻觉”问题可能导致错误信息输出。国外生成式人工智能产品平台 Vectara 使用 Hughes 幻觉评估模型（HHEM）对全球主流大语言模型进行了评估，结果表明 DeepSeek-R1 幻觉率达到 14.3%，高于阿里通义千问的 2.5%，GPT-4 的 1.4%。现实表明，金融领域的“大数据杀熟”或构成隐性价格歧视。

政务决策过度依赖人工智能可能引发责任界定模糊。算法决策缺乏透明度时，公众可能质疑其公正性，社交媒体内容审核规则的不透明加剧了舆论分裂。比如某地城市规划预测误差引发问责争议，不仅影响了政府公信力，而且引发了“算法歧视”的舆情危机。更需要警惕的是，学者张龙辉撰写的《总体国家安全观视域下的算法安全及其风险治理》一文指出，“算法、数据以及算力向平台企业的集中使平台型企业可能进入国家主权相关领域，给国家安全带来挑战，引发国家安全诸领域、各环节的算法安全问题。”

法律法规适配滞后

大模型的迭代周期以月，甚至更短的时间为单位计算，而法律体系的更新往往需要数年立法周期。这种技术迭代速度差与法律框架响应效率失衡导致既有法律法规难以有效覆盖大模型研发、训练、应用与监管等全生命周期，或导致出现“算法黑箱越复杂，法律空白越明显”的困境。如现行《中华人民共和国网络安全法》等法规主要针对传统网络数据，对政务 AI 接入缺乏明确要求。

传统政府管理体制与政务数智化之间存在一定的不适应。政务信息资源采

集、共享、协同和开发利用是一项庞大的系统工程。虽然我国已制定并发布了相关共享管理办法，但执行起来仍会遇到部门利益带来的阻力，导致信息孤岛、信息烟囱问题的发生。以部门为单元按照传统方式各自为政的政务建设，造成政务信息难以充分共享开放，为今后的数据整合再开发利用留下隐患。

区域应用不平衡与伦理冲突

全国数字基础设施存在地区性差异导致数字鸿沟加剧，影响了政务数智化的均衡性和普及性。经济发达地区凭借资金和技术优势，更易推动大模型等人工智能技术的应用，欠发达地区政府因领导治理理念滞后、机关财政不足、办公人员数字技能薄弱，易导致大模型政务应用的“低端锁定”，进而加剧区域政务治理水平和能力的不平衡状况。同时，城乡信息化基础设施落差是基层政务数字化落后的瓶颈，导致乡镇政务数智化普遍滞后城市机关。

技术过度依赖容易消解人类主体性。AI 应用过度或导致政务人员职能弱化。如某县强制公务员日均使用 AI 超 3 小时，不仅没有简化工作量，反而增加了事务的复杂性。另外，AI 缺乏情感交互能力，难以替代政务服务的社会价值，出现“技术替代民主”的深层危机。例如智能客服系统替代传统政民互动渠道，以及技术官僚主义对民主协商机制的冲击。

大模型政务数智治理策略

强化技术应对，完善数据安全防线

采用本地化部署与数据脱敏。治理逻辑与算法逻辑的融合正在遭遇数据泄露带来的阻力。为此，要加快推动政务信创环境部署（如龙岗区全尺寸模型部署），确保敏感数据不出域。对非公开数据强制实施端到端加密，禁止接口直连调用，并采用水印等加密技术追踪数据流向。只有这样，才能确保政务治理的自主权界限。

推进联邦学习与区块链协同。作为一种分布式机器学习框架，联邦学习在保护数据隐私的前提下能够实现跨机构协作建模，提升异构数据整合和边缘计算的效率。探索在城市治理等领域开展示范应用，逐步建立安全可信的政务联邦生态体系，待条件成熟时及时推广。通过联邦学习实现跨部门数据联合建模，仅共享模型参数而非原始数据。借鉴学习马斯克主持的创新经验，在区块链上记录政务数据使用日志，实现操作可追溯、防篡改、可校验。

提升数据纠偏能力，强化数据协同监测

数据纠偏，构建公平数据集，通过合成数据补充少数群体样本，确保训练数据地区、性别比例均衡，防止算法学习隐性偏见。建立“多样性训练数据”机制，要求标注团队多元化背景以减少标签偏差。降低因数据偏见产生的政策判断偏见，比如某市大模型出现对低收入群体保障性住房入住条件的偏见判断。

协同监测，在模型层面加入公平性约束设计，在训练中加入公平性正则化项，使模型在优化准确率的同时保持群体间预测一致性。使用对抗性去偏见技术分

离敏感特征与任务无关特征，保证人工智能判别更加符合社会公序良俗。在实时监测方面，采用多样化的指标量化公平性，同时建立实时偏见检测系统，对可能出现算法误判的领域进行重点检测。

完善法律法规，健全技术治理体系

面对大模型技术的指数级发展，法律体系需要从“追赶式修补”转向“前瞻性构建”。尽快制定《人工智能法》，对可能出现的安全、伦理、法律等风险进行及时防控。加强对大模型等人工智能系统的合法审慎监管，确保其合法性、可靠性和稳定性。

加快制定大模型等进入政务领域的准入标准和管理规范，建立大模型产品和服务的认证制度，确保其质量；加强应用过程中的数据管理，保障数据的合法收集、使用和共享。推动《生成式AI政务服务应用安全指南》立法，明确算法审计、人工复核等要求，建立政务AI接入“白名单”制度，通过安全评估后方可上线。明确责任界定与伦理审查，实施“双岗制”：AI初审+人工复核，重大决策需保留人工干预通道，为迈向数智政务奠定“安全阀”基础。

推动设施均衡，增强技术伦理建设

统筹新型基础设施建设，设立数字政务发展基金，引导社会资本参与新型基础设施建设。优化算力网络布局，重点推进中西部地区5G基站、数据中心等设施建设。建立健全国家算力枢纽节点间的动态调配机制，将东部冗余算力向西部转移，缩小区域数字鸿沟。加强AI政务下沉基层社区，通过智能终端延伸服务触角。鼓励各地因地制宜，探索实施“数字政务官”培育，开展基层公务员数字素养轮训。提升基层公务员大模型应用能力，实现政务服务从“能办”到“智办”的质变。

打造“政府+行业+公众”的立体伦理网络，实现技术赋能与人文关怀的有机统一。鼓励各地适时成立智能政务伦理督查专班，探索制定政务领域的“智能伦理指南”。在公务员培训中增设“技术伦理”模块，培养既懂政务又通技术的管理人才。建立算法伦理审查委员会，定期评估模型伦理风险。引导行业协会制定“AI开发自律公约”。通过网页端、移动端、热线端等多渠道集成，降低公众参与政务门槛。

结束语

随着新一轮科技革命和产业革命的推进，我国应抓住时代机遇实施创新驱动发展战略，适应经济高质量发展的要求，促进产业体系向技术与资本密集型产业升级发展。

在技术赋能背景下，国产大模型实现突破性发展，催生构建更具韧性、响应力和包容性的治理新范式。顺应技术变革推动政务数字化治理的时代进程，技术不仅提升了治理效能，而且推动政务治理结构的迭代。然而，需要警惕“技术乌托邦”陷阱带来的风险与挑战。这表明，技术变革既是治理范式变革的驱动力，也需要制度投放和伦理建设以实现工具理性与价值理性的平衡。

责任编辑：金烨 投稿邮箱 zhouhl@staff.ccidnet.com