

# DeepSeek 等大模型崛起，数据安全面临的挑战与破局

李秋

国家广播电视总局二八一台 云南昆明 650000

**【摘要】**随着以DeepSeek为代表的大模型技术快速发展，单位部署内部AI系统时面临严峻的数据安全挑战。本文聚焦数据泄露风险，系统分析了数据采集、存储、训练及应用全生命周期的安全隐患，包括第三方数据污染、分布式存储漏洞、API接口攻击等新型威胁。研究提出“技术防护+管理优化”双轮驱动的破局路径：技术层面采用同态加密、联邦学习等前沿技术，构建零信任架构与AI增强的威胁检测体系；管理层面建立数据分类分级制度，完善合规审计机制，并通过“三员分立”制度强化人员管控。结合DeepSeek等标杆企业的实践案例，验证了该策略的有效性。研究结果表明，需通过动态自适应安全框架实现数据安全防护的闭环管理，为大模型时代的数据安全提供系统性解决方案。

**【关键词】**数据安全；大模型；数据泄露防范

DeepSeekWith the rise of large models, challenges and solutions to data security are faced

Li Qiu

State Administration of Radio, Film and Television 281 Kunming, Yunnan 650000

**【Abstract】**With the rapid development of large model technologies represented by DeepSeek, there is a severe data security challenge when deploying internal AI systems.This paper focuses on the risk of data leakage and systematically analyzes the security risks throughout the entire lifecycle of data collection, storage, training, and application, including third-party data contamination, distributed storage vulnerabilities, and new threats such as API interface attacks.The study proposes a dual-driven approach of "technical protection + management optimization": at the technical level, advanced technologies like homomorphic encryption and federated learning are adopted to build a zero-trust architecture and an AI-enhanced threat detection system; at the management level, a data classification and grading system is established, compliance audit mechanisms are improved, and personnel control is strengthened through a "three-person separation" system.By combining practical cases from benchmark companies like DeepSeek, the effectiveness of this strategy has been verified.The research findings indicate that a dynamic adaptive security framework is needed to achieve closed-loop management of data security protection, providing a systematic solution for data security in the era of large models.

**【Key words】**data security; large model; data leakage prevention

## 一、引言

随着人工智能技术的突破，以 DeepSeek 为代表的大模型在自然语言处理、代码生成等领域展现出强大能力。据《2025 全球 AI 大模型发展白皮书》显示，我国已建成超 200 个具有行业影响力的大模型，其中 67% 的企业选择部署私有云或混合云架构的内部大模型<sup>[1]</sup>。然而，模型训练所需的海量数据（平均单模型训练需 500TB 以上数据）与互联网连接的基础设施，使单位数据面临严峻的泄露风险。某金融机

构调研表明，73% 的 AI 相关数据泄露事件发生在模型训练阶段，传统安全防护体系已难以应对新型威胁<sup>[2]</sup>。本文以数据全生命周期为主线，从技术与管理双维度探讨防范策略，为企业构建安全可信的大模型应用体系提供参考。

## 二、大模型部署中的数据安全挑战

### 2.1 数据采集与预处理

数据采集阶段的合规性风险尤为突出。某车企在训练自

自动驾驶模型时，因违规采集道路监控视频中的人脸信息，被处以 2.3 亿元罚款。当前多数企业采用第三方数据供应商服务，但《网络安全法》第 37 条明确规定，数据跨境传输需通过安全评估。然而调研发现，61%的企业未对第三方数据进行溯源验证，导致训练数据中混入含恶意代码的污染数据<sup>[3]</sup>。在数据标注环节，某医疗 AI 公司因标注人员误将患者病历上传至公共云存储，造成 58 万条敏感信息泄露。

## 2.2 数据存储与管理

分布式存储架构带来新的风险敞口。某央企大模型平台采用 HDFS 分布式文件系统，因节点间数据同步机制存在漏洞，导致 23 个节点同时感染勒索病毒，3PB 训练数据遭加密勒索。云存储的暴露面管理问题同样严峻，某地方政府 AI 平台因未及时关闭开放 API 接口，导致内部政策文件被恶意爬取。更值得关注的是，模型参数本身可能成为攻击目标，2024 年某高校研究团队通过分析公开模型的梯度信息，成功还原出 72% 的原始训练数据特征。

## 2.3 模型推理与应用

API 接口成为数据泄露的主要通道。OpenAI 的 GPT-4 API 曾因未限制调用频率，导致某用户单日下载 30 万份内部合同模板<sup>[4]</sup>。攻击手段也在不断升级，某银行智能客服被植入 prompt 注入攻击代码，将用户查询的账户余额信息转发至境外服务器。模型输出内容的合规性同样存在隐患，某法律咨询平台的 AI 生成法律文书中，意外包含客户未授权披露的商业信息。

# 三、数据安全破局的管理策略

## 4.1 数据分类分级制度的深化实践

数据分类分级体系的建设需要兼顾行业特性与技术发展。在“5 级 3 类”基础上，建议引入语义分析技术实现自动化分类。例如某金融机构通过自然语言处理模型，将客户信贷合同中的敏感字段识别准确率提升至 98.7%。动态分级机制方面，可建立数据价值评估模型，综合考虑数据更新频率、访问频次、业务关联度等 12 个维度。某电信运营商实践显示，该模型使数据保护成本降低 23%，关键数据响应速度提升 40%。

在实施路径上，建议采用“三步走”策略：首先完成资产普查建立数据清单，其次通过风险评估确定分级规则，最后开发自动化标签系统。某汽车制造企业通过该模式，将 2000TB 非结构化数据的分类周期从 6 个月缩短至 15 天。值

得注意的是，分级制度需与访问控制策略深度耦合，某军工企业采用 ABAC（属性基访问控制）技术，实现基于数据密级、用户角色、终端安全状态的三维权限管控。

## 4.2 合规与审计体系的创新构建

全生命周期审计需突破传统日志分析的局限性。某互联网公司部署的行为分析系统，通过建立用户操作基线模型，识别出异常数据导出行为的准确率达 92%。跨境数据传输方面，建议采用区块链存证技术，某跨国药企将临床试验数据出境审批时间从 30 天压缩至 7 天。数据安全能力成熟度模型（DSMM）的应用可显著提升合规水平，某央企通过 DSMM 三级认证后，数据安全事件处置效率提高 60%。

智能审计系统的发展呈现三个趋势：AI 驱动的行为异常行为检测、数字孪生技术构建审计沙盘、隐私计算实现审计过程的数据保护。某证券交易所的实践表明，结合知识图谱技术可将审计线索关联分析效率提升 8 倍。在跨境传输合规方面，需特别关注 GDPR、CCPA 等域外法规与国内《数据安全法》的衔接，某科技公司建立的合规决策树模型已成功处理 2000+跨境数据请求。

## 4.3 人员与组织管理的体系化建设

人员安全意识培养需构建长效机制。某银行开发的 VR 钓鱼模拟系统，使员工误点率从 43% 降至 12%。建议建立人员安全能力矩阵，将数据安全纳入绩效考核体系，某跨国公司实践显示，该措施使违规操作率下降 37%。在组织架构方面，数据安全委员会应具备跨部门协调权，某能源集团通过赋予委员会预算审批权，推动 32 个业务系统完成安全改造。

第三方管理需建立全链条管控机制。建议实施供应商分级管理，某零售企业将服务商分为战略级（S 级）、重要级（A 级）、普通级（B 级），分别对应不同的审计频率和保证金比例。在技术保障方面，可部署 API 安全网关实现第三方接口监控，某物流平台通过该措施拦截了 89% 的越权访问请求。特别需要关注云服务商的责任边界，某金融机构与云服务商签订的“双活备份+应急接管”协议，将数据恢复时间目标（RTO）缩短至 15 分钟。

# 四、案例分析：DeepSeek 的数据安全实践

## 5.1 DeepSeek 的安全架构演进

“三层同心圆”架构在实践中不断迭代升级。核心层采用 SM9 算法的多版本共存方案，支持国密/国际算法的平滑切

换。该方案通过动态密钥协商机制,在确保密钥更新过程中业务连续性的同时,实现了密钥生成效率提升40%。中间层联邦学习框架通过动态权重加密技术,使模型训练效率提升35%。其创新点在于将加密算法与梯度压缩技术结合,在保持模型精度的前提下,将通信开销降低了62%。外围层威胁检测系统引入因果推理模型,可提前72小时预测攻击路径。该系统基于多模态数据融合技术,整合了流量日志、设备指纹和行为画像等12类数据源,在某能源企业部署后成功拦截了一起APT攻击。沙箱隔离技术的最新突破是实现跨容器资源调度,某金融客户应用后,训练任务完成时间缩短28%。其底层采用基于硬件虚拟化的内存隔离技术,确保不同容器间数据零泄露。

在密钥管理方面,DeepSeek独创的“密钥联邦”机制,将加密密钥分片存储在多个可信执行环境(TEE)中。该方案在保障安全性的同时,使密钥恢复效率提升50%。通过引入区块链存证技术,实现密钥生命周期的全流程可追溯。针对大模型训练中的数据污染问题,开发了基于对抗训练的清洗算法,可有效识别并过滤99.3%的恶意样本。该算法通过构建双鉴别器对抗网络,在检测准确率提升18%的同时,将误报率控制在0.07%以下。

### 5.2 行业标杆对比分析的深化

与OpenAI的对比显示,DeepSeek在数据主权保护方面具有显著优势。其自主研发的“数据水印”技术,可精确追踪数据泄露源头,某汽车客户应用后成功定位3起内部泄露事件。该技术基于DNA级特征编码,在1PB级数据中可实现秒级溯源。在合规性方面,DeepSeek的“数据护照”系统已支持23个国家的法规要求,相比华为盘古的“一云多芯”架构,其混合云部署方案更具灵活性。通过构建跨云身份联邦体系,实现了多云环境下的统一权限管理。

行业对比揭示三大发展趋势:零信任架构的广泛应用(如Anthropic采用持续验证机制)、量子加密技术的预研

(如IBM的量子密钥分发试点)、AI安全攻防的自动化(如C3.ai的自主响应系统)。DeepSeek的创新在于将联邦学习与区块链结合,构建数据交易可信环境,该模式已在医疗数据共享领域取得突破。其开发的智能合约自动审计系统,通过形式化验证技术,将合约漏洞发现率提升了73%。在量子安全领域,DeepSeek已完成基于格密码的密钥交换协议研发,性能相比传统RSA算法提升40%,为后量子时代的数据安全奠定基础。

### 5.3 经验总结与未来展望

DeepSeek的实践验证了“管理-技术-运营”三位一体的安全理念。在组织创新方面,建议建立数据安全官(CDSO)制度,某央企设置该职位后,跨部门协作效率提升45%。技术发展趋势方面,隐私计算与AI的融合将催生新范式,某科研机构的联合研究显示,结合联邦学习与大模型可使数据利用率提升60%。

未来挑战主要集中在三个方面:边缘计算场景下的轻量化加密技术、多模态数据的统一安全框架、生成式AI带来的新型攻击防护。DeepSeek已启动“边缘安全加速计划”,开发基于神经网络的轻量级加密算法。在标准化建设方面,主导制定的《人工智能数据安全技术指南》已被纳入国家标准草案。

## 五、结论

本文通过剖析大模型部署中的数据安全风险,提出了涵盖技术防护、管理优化和合规建设的综合解决方案。随着量子计算、联邦学习等技术的发展,数据安全体系将向“主动防御+智能响应”方向演进。建议企业构建动态自适应安全框架,通过持续监测、智能分析和自动响应,实现数据安全防护的闭环管理。在享受大模型技术红利的同时,必须坚守数据安全底线,为数字经济健康发展筑牢屏障。

## 参考文献

- [1]斯鑫,赵琼芳.大数据时代云存储平台隐私权保护的困境与破局[J].法制博览,2025(4).
  - [2]张利娟.“数据二十条”出炉,如何挖掘数据“石油”?[J].中国报道,2023(1):66-68.
  - [3]毛子骏,邹敏.数字经济与民生协调发展:数字政府的调节作用[J].贵州财经大学学报,2024(5):79-89.
  - [4]刘雯霏,马婷艳.从数字化车间建设场景“破局”,以先进技术深挖工业数据价值[J].智能制造,2022(3).
- 作者简介:李秋(1990.09),男,汉族,云南玉溪人,中级职称,本科,研究方向为广播电视监测监管、网络安全。