人工智能时代企事业单位网络安全管理的转型优化研究

曹姚
中国社会科学院图书馆 北京 100732

摘 要 随着人工智能技术与企事业单位业务的深度融合，网络安全管理的对象、方法与边界发生系统性变化，传统以边界防护和规则匹配为核心的模式难以应对高动态、强对抗、跨域化的威胁环境。本文围绕企事业单位网络安全管理的核心特征，提出以智能化威胁检测与防御体系、自动化安全运维与响应能力、数据驱动的安全决策支持系统为核心的转型方向，并从组织重塑、AI融合规范、全员安全意识与AI安全应急响应机制4个方面给出优化策略，以期构建自适应、协同化、可度量的安全治理框架，提升风险识别、处置与恢复的全链路能力，支撑企事业单位在数字化与智能化进程中的安全稳健运行。

关键词 人工智能；网络安全管理；威胁检测；自动化响应

Research on the Transformation and Optimization of Cyber Security Management in Enterprises and Institutions in the Era of Artificial Intelligence

Cao Yao
Chinese Academy of Social Sciences Library, Dongcheng District, Beijing 100732, China

Abstract With the deep integration of artificial intelligence (AI) technology into the business operations of enterprises and institutions, the objects, methods, and boundaries of cyber security management are undergoing systemic changes. Traditional models centered on perimeter protection and rule-based matching are increasingly inadequate against high-dynamic, highly adversarial, and cross-domain threat environments. This paper examines the core characteristics of cyber security management in enterprises and institutions, proposing transformation directions focused on an intelligent threat detection and defense system, automated security operation and response capabilities, and a data-driven security decision support system. It further provides optimization strategies across four aspects: organizational restructuring, AI integration regulations, all-staff security awareness, and AI-based security incident response mechanisms. The aim is to construct an adaptive, collaborative, and measurable security governance framework to enhance capabilities in risk identification, handling, and recovery throughout the entire security lifecycle, thereby supporting the secure and stable operation of enterprises and institutions in their digital and intelligent transformation.

Key words Artificial Intelligence; Cyber Security Management; Threat Detection; Automated Response

在数字经济与产业智能化加速推进的背景下，网络安全已成为企事业单位业务连续性与数据资产保护的关键支撑。以大模型为代表的人工智能技术带来效率与创新的同时，也引入深度伪造、提示注入、模型越狱、供应链依赖等新型风险，暴露出单点防御失效、告警过载、数据孤岛、攻防不对称等共性短板。面向人工智能时代，企事业单位需要从战略、组织与技术层面协同演进，系统推进AI原生的安全架构与运营体系，形成人机协同、联动防御、闭环运营的新型治理能力，以实现对复杂威胁的前置识别、快速处置与持续进化。

1 企事业单位网络安全管理的核心特征

面向人工智能深度融入业务的新环境，企事业单位网络安全管理呈现出以AI原生为牵引、体系化联动为支撑、数据驱动与自动化为特征、合规内生化为保障的综合形态。核心在于以智能底座整合多源安全数据，构建跨域联动防御与人机协同的运营体系，实现对威胁的持续检测、快速响应、闭环优化；同时将数据与模型纳入全生命周期治理，落实权限强管控、内容安全治理、零信任等机制，强化对模型安全、供应链安全与跨境数据的合规管理，并以实战化演练与能力评估检验策略有效性，形成弹性、可扩展、可度量的安全能力体系[1]。

2 人工智能赋能企事业单位网络安全管理的转型方向

2.1 智能化威胁检测与防御体系构建

面向AI驱动的复杂威胁，企事业单位需以AI原生理念重塑检测与防御，构建覆盖端、边、云、网的一体化XDR体系，打通EDR、NDR、SIEM、云WAF、邮件网关等数据孤岛，形成跨域关联与统一研判能力。以安全大模型与多模态分析提升对零日、APT、钓鱼、API滥用等攻击的发现率，结合行为基线、情报联动、沙箱与回溯取证实现检测、验证、处置的闭环。同步推进模型安全与内容安全治理，对提示注入、越狱、数据投毒等具备识别与拦截能力，实现红域隔离、零信任访问、最小权限加固模型。通过红蓝对抗、攻防演练与持续评估验证策略有效性，沉淀可解释与可审计机制，支撑稳态运营与弹性扩展的安全能力持续进化。

2.2 自动化安全运维与响应能力提升

以SOAR与大模型智能体为核心，构建降噪、研判、处置、复盘的自动化运营链路，利用自然语言编排检测、调查、封禁、隔离、回滚等动作，显著降低告警过载与MTTD/MTTR。建设跨设备统一策略平台，实现异构防火墙、交换机等策略的集中采集、标准化、风险识别、自动下发与验证，终结策略漂移与冗余；结合零信任、SASE、WAAP、MSS等架构打造一体化主动防御，在近源侧实现拦截与响应。以边缘智能与全球节点缩短策略下发与处置时延，并通过实战化评估与闭环优化形成自演化运营体系，兼顾效率、可控与合规，支撑7×24 h稳定运行[2]。

2.3 数据驱动的安全决策支持系统发展

以数据治理为底座，构建指标、模型、知识、可视化的决策体系，汇聚资产、漏洞、配置、日志、威胁情报等多源数据，形成统一语义层与风险画像。同时，依托风险矩阵与因果图谱开展优先级排序、影响评估、资源优化与合规审计，以预测性分析与仿真推演支持策略制订与投资取舍。此外，建设安全数据湖与指标库，规范数据分级分域、质量度量、血缘与留痕，并以可解释AI与人机协同提升决策可信度。通过仪表盘、态势大屏与自助分析赋能管理层与一线，形成“数据—洞察—行动”闭环，推动从被动响应到主动治理的转型，并以持续评估与演练保障决策机制与业务目标同频共振。

3 人工智能时代企事业单位网络安全管理的转型优化策略

3.1 重塑网络安全管理组织架构

面向AI带来的高速度、高复杂度与高不确定性，企事业单位应从职能堆叠转向作战化、平台化、智能化的组织形态。可设立由主要领导机构牵头的AI与网络安全治理委员会，统筹信息安全、法务合规、数据治理、人力资源、业务线负责人等，形成“战略—执行—监督”三层治理闭环。在组织层级上，构建战略决策层、技术中台层、一线作战层的矩阵，战略层负责规划路线与资源，技术中台层负责沉淀可复用的AI安全能力组件，一线作战层负责以跨职能战队形态对接业务与威胁现场，实现从需求到处置的端到端闭环。同时，要配套建立AI安全机构与算法治理机制，将安全与合规要求内嵌到模型研发、数据治理、上线发布、运营监控全流程，确保可解释、可审计、可问责。此外，以“平台+战队”双轮驱动打破数据孤岛，建设统一的数据与模型资产台账、统一的安全运营中心与AI安全运营平台，并以三级联动形成“总部—区域—业务单元”的协同响应与资源调度能力，显著提升对AI驱动威胁的发现、研判与处置效率[3]。

在组织落地上，优先完成“三张清单+一套机制”，资产清单、风险清单、责任清单；机制上以“年度目标—季度OKR—月度演练”对齐安全与业务节奏，将红队评估、漏洞赏金、蓝军对抗常态化，并将安全KPI纳入业务与研发绩效。治理方式坚持包容审慎、风险导向、技管结合、开放合作，以风险管理为主线，动态跟踪AI技术演进与威胁态势，及时调整控制点与资源投入；同时，在组织文化上强化安全是业务能力的认知，推动安全前置与“DevSecOps、MLSecOps”落地，使安全成为AI产品与业务创新的加速器而非掣肘。

3.2 制定AI融合的安全管理规范

AI融合的安全规范应覆盖模型、数据、系统、内容、供应链的全链条与全生命周期。制度层面，围绕生成式AI建立安全评估、算法备案、内容标识、可追溯问责等刚性要求，明确生成合成服务提供者、内容传播服务提供者、应用分发平台等主体的法定义务[4]；标准层面，对标国家强制性标准与行业指南，细化显式标识、隐式标识的技术实现，确保跨平台可追溯与互操作；流程层面，将数据分级分类、脱敏与最小化采集、访问控制、模型可解释与公平性评测、上线前对抗测试、上线后持续监测与召回固化为标准作业；度量层面，建立覆盖率、误报、漏报率、MTTD/MTTR、策略命中率、合规通过率等指标，驱动持续改进。对跨境数据、个人信息、重要数据等敏感要素，明确采集、传输、存储、使用、共享、删除与出境的合规路径与审批机制，形成事前评估、事中监测、事后纠偏的闭环治理。

面向大模型与智能体应用，规范细化“接入放心用、权

限强管控、严把内容关、杜绝新隐患、运营可闭环”的操作要求：以安全红域对数据、算力、平台、应用、终端人员进行边界隔离与零信任访问；对提示词注入、越狱攻击、模型窃取与篡改、供应链投毒等高危场景设定专项控制；对生成合成内容实施显式/隐式双标识与来源可验，并在分发与审计环节校验标识完整性；在运营侧以AISOC实现告警降噪、人机协同研判、自动化处置与策略自优化，将威胁处置从小时级压缩至分钟级/秒级[5]。

3.3 构建全员网络安全意识培训体系

在AI场景中，员工既是最后一道防线也是最大变量，培训体系需从通用合规宣贯升级为岗位化、实战化、持续化的能力工程。治理层面，设立全员网络安全责任制与岗位安全KPI，将AI风险识别、数据分级与脱敏、钓鱼与社会工程学防范、生成内容核验、合规使用AI工具纳入入职、在岗与专项培训；内容层面，围绕AI时代的新型风险开发情景化课程等，以真实案例与攻防演练提升识别与处置能力；运营层面，采用“线上平台+线下演练+竞赛激励”的组合，如知识竞答、红蓝对抗、钓鱼演练、应急演练等，形成“学—练—赛—评—改”的闭环，并将培训覆盖率、演练通过率、违规事件率纳入部门与个人考核，推动安全意识从被动接受走向主动运用。

针对高管与关键岗位，培训应聚焦负责任AI、算法偏见与可解释、隐私与知识产权、供应链安全、安全开发生命周期等主题，结合数据出境评估、算法备案、内容标识等合规实操，强化安全即能力、安全即业务价值的认知；同时，建立常态化演练与复盘机制，如年度/季度桌面推演、专项红蓝对抗、钓鱼演练与应急处置复盘，并将演练结果固化为更新后的SOP与培训教材。

3.4 建立AI安全应急响应管理机制

AI驱动的攻击呈现自动化、隐蔽化、链式化特征，传统单点响应、人工处置的模式难以满足时效与规模要求，需建设全天候、全域、全链路的应急响应体系。机制设计上，以事件响应预案为总纲，明确分级分类标准、指挥链路、角色职责、沟通与通报、证据保全与合规要求，并针对模型投毒、提示词越狱、数据泄露、深度伪造引发的社会工程学攻击、供应链入侵等AI特有场景制订专项SOP；技术支撑上，建设AISOC与跨域联动能力，打通网络、应用、终端、云、模型的数据与告警，利用大模型驱动的事件处置智能体实现告警降噪、根因定位、策略编排与自动化修复，将策略下发与阻断从小时级缩短至秒级；外部协同上，预置与监管、公安、云与运营商、行业组织的联络与协同流程，建立7 × 24 h值守与专家库，确保重大事件响应迅速、处置得当、证据完备、对外透明[6]。

运行层面，以平战结合为原则，常态化开展红蓝对抗、桌面推演与专项演练，覆盖AI生成内容引发的舆情与业务连续性风险，验证预案有效性与资源充足性；在事件处置中，严格遵循发现、研判、隔离、根除、恢复、复盘流程，落实取证与链上溯源，并在24 h内完成初报、72 h内完成根因分析与整改计划；事后通过威胁情报共享、漏洞修复与配置基线加固、策略与模型再训练实现闭环改进，将经验沉淀为可复用剧本与知识库，并纳入年度审计与合规评估。

4 结束语

综上，人工智能正在重塑企事业单位网络安全的对象、方法与边界，唯有以AI原生理念重构治理体系，才能在高动态、强对抗、跨域化的威胁环境中保持稳健。面向实践，需以组织重塑为牵引、AI融合规范为准绳、全员安全意识为底座、AI安全应急响应为关键抓手，构建自适应、协同化、可度量的安全治理框架，持续把风险识别、快速处置、闭环优化做深做实。展望未来，应坚持“人机协同、联动防御、数据驱动”的路径，推进XDR/SOAR/大模型智能体与安全运营的深度融合，沉淀可解释、可审计、可问责的机制；同时以合规内生与供应链安全为边界条件，兼顾效率、可控与合规。

参考文献

[1]吴锐锋.企业网络信息安全管理体系建设分析[J].中关村,2025(6):174-176.

[2]姚柔娜.数字化转型背景下企业网络安全管理现状及优化路径[J].中国宽带,2024,20(12):58-60.

[3]彭浩鸣,王安蒙.数智化转型企业网络信息安全管理体系的建设策略[J].智慧中国,2024(10):71-72.

[4]中央网络安全和信息化委员会办公室.网络安全技术人工智能生成合成内容标识方法：GB 45438—2025[S].北京:中国标准出版社,2025.

[5]陈小辛.数字化转型下企业网络安全管理战略调整路径[J].金融科技时代,2024(10):52-56.

[6]药丹.数字化转型背景下企业网络安全管理现状及优化路径[J].中国管理信息化,2024,27(17):114-116.

