ICS

T/GXDSL

团　　体　　标　　准

T/GXDSL —2026

网络数据资产安全事件 AI 应急处置技术规范

Technical Specification for AI Emergency Disposal of Network Data Asset Security Incidents

（工作组讨论稿）

（本草案完成时间：2026 - 5 - 6 ）

2026 - - 发布　　2026 - - 实施

广西电子商务企业联合会　发布

目　　次

前　言

本文件依据GB/T 1.1-2020 《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由广西产学研科学研究院提出。

本文件由广西电子商务企业联合会归口。

本文件起草单位：

本文件主要起草人：

本文件为首次发布。

网络数据资产安全事件 AI 应急处置技术规范

1 引言

我国数字化转型深入推进，网络数据成为国家基础性战略资源，而传统人工应急响应已无法适配网络数据安全威胁的智能化演变，难以满足国家网络安全战略与数字经济发展需求。为落实总体国家安全观与网络强国部署，依据相关法律法规及人工智能技术发展趋势，广西产学研科学研究院联合相关单位制定本规范，明确 AI 在安全事件全流程的技术角色与标准，提升应急处置效能，为国家数字安全提供支撑，本规范为团体标准，供社会自愿采用。

2 范围

规定了 AI 网络数据安全应急处置的核心要求，覆盖全流程，兼顾通用性与行业适配性。适用于我国所有涉及数据处理的主体，具体包括：相关国家机关、企事业单位及社会组织，用于规范 AI 应急处置能力建设；安全服务提供商，用于指导 AI 应急响应系统开发部署；第三方评估机构，作为应急处置能力评价依据；广西产学研科学研究院及相关主体，用于符合性判定与标准落地；科研机构，用于技术研发与成果转化。

3 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅该日期对应的版本适用于本文件；凡是不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件，契合我国网络安全标准体系建设和人工智能安全规范发展要求。

GB/T 20985.1—2017 信息技术安全技术信息安全事件管理第 1 部分：事件管理原理

GB/T 20986—2023 信息安全技术信息安全事件分类分级指南

GB/T 39204—2020 信息安全技术数据安全应急响应指南

GB/T 41817—2022 信息安全技术数据安全评估机构能力要求

GB/T 44462.1—2024 工业互联网企业网络安全第 1 部分：分级分类防护

JR/T 0223—2021 金融数据安全数据生命周期安全规范

《生成式人工智能服务管理暂行办法》（国家互联网信息办公室，2023 年）

《国家网络安全事件报告管理办法》（2025 年 9 月 15 日施行）

《网络数据安全管理条例》（国务院令 2025 年第 789 号）

《中华人民共和国网络安全法》（2025 年 10 月 28 日修正版）

4 术语和定义

下列术语和定义适用于本文件，兼顾技术准确性与国家政策导向，统一行业认知。

4.1 网络数据资产

指组织以电子形式存储、处理或传输的数据资源，是国家数据资源的重要组成部分，包括但不限于数据库资产、文件存储资产、大数据平台资产、API 接口资产、云端数据资产及各类业务系统生成、采集的结构化与非结构化数据，涵盖核心数据、重要数据和一般数据三个等级。

4.2 AI 应急处置

指依托机器学习、自然语言处理、自动化编排、生成式 AI 等人工智能技术，遵循合法合规、安全可控原则，对网络数据安全事件进行自动或半自动的识别、分析、决策、遏制、恢复及取证的系列活动，是“技防+智防”融合的核心体现，旨在提升应急处置的效率与精准度。

4.3 安全事件定级

依据 GB/T 20986—2023、《网络数据安全管理条例》及本文件第 7 章规定，综合考虑数据资产重要度、受影响范围、危害程度及对国家安全、公共利益的潜在影响，对网络数据安全事件划分等级（特别重大、重大、较大、一般），为分级处置、分级上报提供依据。

4.4 AI 应急响应基线与阈值

指通过历史数据训练、行业最佳实践提炼或运维人员预设的，用于度量网络流量、用户行为、数据访问模式异常程度的参数集合，是 AI 实现精准监测、预警和处置的基础，需结合业务场景动态优化，适配新型攻击手法和业务变更需求。

5 缩略语

AI：人工智能 Artificial Intelligence

SOAR：安全编排自动化与响应 Security Orchestration, Automation and Response

UEBA：用户与实体行为分析 User and Entity Behavior Analytics

NTA：网络流量分析 Network Traffic Analysis

SLA：服务等级协议 Service Level Agreement

API：应用程序编程接口 Application Programming Interface

6 总体原则

本规范遵循“安全优先、合规引领、创新适配、持续提升”的核心思路，立足国家数据安全保障需求，结合人工智能技术特点，确立以下总体原则，契合我国生成式人工智能服务管理和人工智能安全治理要求。

6.1 合法合规与数据最小化

AI 应急处置系统在采集网络流量、日志及用户行为数据进行模型训练和推理时，应严格符合《中华人民共和国网络安全法》《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》及《生成式人工智能服务管理暂行办法》要求，坚持数据合法来源，对涉及个人信息的数据进行匿名化或脱敏处理，仅采集应急处置必要的最小数据集，不得收集非必要个人信息，严禁非法留存、泄露用户相关数据，确保数据处理全流程合规。

6.2 可解释性与人工最终决策

AI 模型的判定结果应当具有可解释性，能输出置信度评分、关键特征贡献度及决策依据，提升 AI 应急处置的透明度和可信度，契合人工智能安全可解释性要求。涉及数据销毁、大规模隔离、关键业务下线等高风险动作时，必须设置“人工确认”环节，实行“AI 辅助、人工终审”，防范 AI 决策失误引发的安全风险，保障处置行为的合理性与可控性。

6.3 弹性与对抗稳健性

AI 应急模型应具备较强的弹性与对抗稳健性，能够有效抵御数据投毒、模型后门、对抗性攻击等新型威胁，契合人工智能安全防护要求。当攻击者试图通过构造微小扰动绕过 AI 检测时，模型识别逃逸攻击的成功率不低于 85%（指标依据来源：GB/T 41817—2022 相关评估要求）；同时，模型应能适配业务系统变更、网络环境调整等场景，确保应急处置能力持续有效。

6.4 持续演进与自主创新

AI 模型应建立在线学习或定期重训机制，周期不得超过 90 个自然日，及时吸收新型攻击手法、业务变更及行业最佳实践，持续优化模型性能。同时，鼓励依托我国自主创新的人工智能技术、芯片及软件平台，构建安全可信的 AI 应急处置体系，推动 AI 应急处置技术自主可控发展，助力我国人工智能安全标准体系完善，提升国家网络数据安全自主防护能力。

7 AI 应急处置技术参考架构

为实现网络数据安全事件 AI 应急处置的自动化、智能化、规范化，组织应构建“分层协同、闭环管控、安全可信”的四级功能架构，衔接国家网络安全等级保护要求，确保各层协同联动、数据互通，提升整体应急处置效能，契合“全链条协同”的网络安全防护理念。

7.1 数据源与采集层

作为 AI 应急处置的基础，需实现全维度、高精度、实时化的数据采集，覆盖网络、主机、应用、数据等全场景，为 AI 模型分析提供高质量数据支撑，同时符合数据存储合规要求：

7.1.1 实时采集数据源：网络边界流量（包括加密流量元数据）、主机审计日志、API 访问日志、数据库审计日志、威胁情报源、业务系统操作日志及终端行为数据；

7.1.2 存储周期：原始日志至少存储 6 个月，符合《中华人民共和国网络安全法》相关要求；安全事件告警及 AI 判定结果存储周期不少于 2 年，确保事件可追溯、可复盘；

7.1.3 数据标准化：统一数据格式契合标准 Schema，时间戳同步至毫秒级，实现不同数据源的数据互通、融合分析，提升数据利用效率。

7.2 AI 模型与检测分析层

作为 AI 应急处置的核心，需部署高性能、高可靠的 AI 模型，实现安全事件的精准识别、智能研判，兼顾检测精度与响应速度，契合人工智能安全能力成熟度要求：

7.2.1 部署基于深度学习（如 Transformer、LSTM）的 UEBA 模型，用于识别偏离基线的异常数据访问行为、特权账号滥用等潜在风险；

7.2.2 部署自然语言处理模型，用于实时解析威胁情报、违规数据出境内容识别及安全日志的智能化分析，提升威胁识别的全面性；

7.2.3 部署生成式 AI 模型，辅助安全事件复盘、处置方案优化，提升应急决策的科学性；

7.2.4 模型性能指标：误报率应控制在 5%以内；漏报率应控制在 1%以内（针对已训练过的已知

攻击类型）；对新变种攻击的检出率不低于 75%，确保模型能够有效应对新型安全威胁。

7.3 编排与处置执行层

作为 AI 应急处置的执行核心，需集成 SOAR 平台，实现处置流程的自动化编排、快速执行，确保安全事件得到及时遏制、有效处置，提升处置效率：

7.3.1 预置处置剧本不少于以下 8 类，覆盖各类典型网络数据安全事件，可根据行业特点灵活扩展：勒索软件应急处置剧本、数据泄露遏制剧本、API 滥用封禁剧本、权限异常回收剧本、数据库拖库阻断剧本、内部违规拷贝告警剧本、供应链数据泄露剧本、业务可用性故障切换剧本；

7.3.2 动作执行时延：从 AI 发出告警到执行自动化脚本（如封禁 IP、终止进程、挂起账号）的平均时间应小于 60 秒，其中高置信度攻击的处置时延不超过 30 秒，实现“秒级响应、快速遏制”。

7.4 可视化管理层

作为 AI 应急处置的管控中枢，需提供全面、直观的可视化管控能力，支撑应急决策、过程监控和效果评估，提升应急处置的可控性：提供 3D 拓扑或动态仪表盘，实时展示数据资产分布、流动风险、安全事件态势及 AI 应急处置进度，实现“一屏统览、动态管控”；支持自然语言交互查询（如输入“查询最近 24 小时针对数据库的高危数据导出事件”），AI 自动生成安全报告，为应急决策提供数据支撑；具备应急处置过程追溯、日志留存、报表生成功能，满足合规审计和复盘改进需求，契合国家网络安全事件报告管理要求。

8 数据资产分类分级与事件定级

数据资产分类分级是 AI 应急处置分级管控、精准施策的基础，需严格依据国家法律法规和行业规范，明确分类分级标准及事件定级映射关系，确保与国家数据安全分级分类管理要求保持一致。

8.1 数据资产分级

组织应在前置步骤中，依据《网络数据安全管理条例》《中华人民共和国数据安全法》及行业规范（如 JR/T 0223—2021），结合自身业务特点，将数据资产分为以下三级，明确各级数据的防护要求：

8.1.1 核心数据：可能危害国家安全、国民经济命脉、重大公共利益的数据，是国家数据安全保护的核心对象，需实施最高级别防护；

8.1.2 重要数据：可能对政治、国土、军事、经济、文化、社会等造成损害但不触及国家秘密的数据，需实施重点防护；

8.1.3 一般数据：仅涉及组织内部管理或公开数据，实施常规防护即可，兼顾防护效能与成本控制。

8.2 安全事件分级映射

AI 系统应根据数据资产分级、事件危害程度，将事件定级自动映射至对应的处置策略，实现“分级处置、分级上报”，契合国家网络安全事件应急响应要求：

8.2.1 特别重大/重大事件（涉及核心数据泄露、大规模重要数据丢失，或可能危害国家安全、重大公共利益）：触发最高应急级别，系统自动上报至国家网信部门及公安机关，根据《国家网络安全事件报告管理办法》（2025 年）要求，2 小时内完成初步报告；AI 执行全链路追踪取证并实施网络隔离，防止风险扩散，同时启动最高级别应急响应预案；

8.2.2 较大事件（涉及重要数据被未授权访问但可控，未造成大规模泄露或严重危害）：启动组织级应急预案，自动限制涉事人员权限、封锁涉事数据接口，向属地公安机关及行业监管部门报告时限为 24 小时内，同步开展事件处置与溯源；

8.2.3 一般事件（涉及一般数据且未造成实质性损失）：AI 自动记录事件详情、生成整改工单，纳入月报汇总，定期开展复盘优化，实现闭环管理。

9 AI 监测与预警要求

AI 监测与预警是实现网络数据安全事件“早发现、早预警、早处置”的关键，需建立科学的基线体系、完善的监测指标和高效的预警推送机制，提升预警的精准度和时效性，契合“主动免疫”的网络安全防护理念。

9.1 异常行为基线建立

基线建立应结合组织业务特点、历史数据及行业最佳实践，确保基线的科学性、适用性，同时支持动态调整：

9.1.1 基线周期：系统上线后需经过至少 30 天的学习期，全面采集正常业务场景下的网络流量、用户行为、数据访问模式等数据，建立初始行为基线；

9.1.2 动态调整：系统应每周更新基线模型，适配“节假日模式”“业务高峰期模式”“业务系统升级”等场景变化，同时结合新型攻击手法，持续优化基线参数，避免误报、漏报。

9.2 关键监测指标

AI 系统至少监测以下关键指标，并设置动态阈值，实现对各类异常行为的精准识别，覆盖数据访问、账号安全、接口调用等核心场景：

9.2.1 数据冗余度：单位时间数据导出量超过历史均值 300%时触发预警，重点监测核心数据、重

要数据的异常导出行为；

9.2.2 地理位移异常：账号在不可达时间内异地登录并下载数据时，立即封锁账号并触发告警，防范账号被盗引发的数据泄露风险；

9.2.3 特权账号滥用：数据库管理员、系统管理员等特权账号，在非维护时间执行数据导出、权限变更等敏感操作时触发告警；

9.2.4 API 异常调用：单个令牌在 10 分钟内调用敏感数据接口超过阈值 50 次，触发限流与熔断，防范 API 滥用导致的数据泄露；

9.2.5 补充监测指标：模型异常输入、训练数据异常篡改（防范数据投毒攻击）、违规数据出境行为，确保监测覆盖 AI 自身安全与数据安全。

9.3 预警推送

预警信息应实现快速推送、精准触达，确保相关责任人及时响应，同时明确处置指引：

9.3.1 推送渠道：通过即时通讯工具、短信或 API 推送到对应责任人，确保预警信息不遗漏；

9.3.2 推送内容：包含置信度评分、受影响数据资产标签（如“核心客户库”“重要业务数据”）、推荐处置剧本 ID、预计影响范围及初步处置建议，为责任人快速处置提供支撑；

9.3.3 预警分级：根据事件等级，将预警分为紧急、高危、中危、低危四级，对应不同的响应时限和处置优先级，提升处置效率。

10 AI 自动化处置流程

AI 自动化处置应遵循“快速遏制、精准取证、及时恢复、彻底溯源”的原则，明确各阶段的处置目标、操作标准和时间要求，实现应急处置全流程闭环，确保处置过程规范、高效、可控，契合国家网络数据安全应急响应流程要求。

10.1 遏制阶段（T0—T1，即事发时刻至 1 小时内）

10.1.1 核心目标：快速遏制风险扩散，防止数据进一步泄露、业务进一步受损，实现“秒级响应、快速控险”：

10.1.2 秒级封堵：AI 判定为高置信度（>95%）攻击时，自动调用防火墙或软件定义边界控制器封锁源 IP、攻击端口，时延（从检测到下发策略）需≤30 秒；

10.1.3 进程降维打击：检测到勒索病毒加密行为、恶意进程等威胁时，AI 调用端点检测与响应接口，立即隔离涉事主机、杀掉关联进程，阻断威胁传播；

10.1.4 凭证熔断：疑似凭证泄露、账号被盗时，自动重置用户密码或 Access Key，强制所有会话注销，防止攻击者进一步利用被盗凭证获取敏感数据；

10.1.5 数据隔离：涉及核心数据、重要数据的安全事件，自动隔离涉事数据资产，禁止非授权访问，防范数据泄露扩大。

10.2 取证与分析阶段（T1—T2，即 1 小时至 8 小时内）

10.2.1 核心目标：全面采集取证数据，精准分析攻击路径、事件原因，为后续恢复、溯源和追责提供依据：

10.2.2 日志筛选：AI 自动筛选事发前后（即事件发生前 15 分钟和发生后 5 分钟）的全部相关日志、网络流量数据、用户操作记录，确保取证数据的完整性；

10.2.3 攻击路径分析：利用图计算技术描绘攻击路径，生成“杀伤链”可视化图谱，明确攻击来源、攻击手段、受影响范围；

10.2.4 数据取证：对涉事数据、恶意文件、攻击指令等进行固定、备份，确保取证数据的合法性、完整性，为后续追责提供支撑；

10.2.5 初步研判：AI 结合威胁情报，对事件类型、危害程度、攻击意图进行初步研判，生成研判报告，为后续恢复决策提供依据。

10.3 恢复阶段（T2—T24，即 8 小时至 24 小时内）

10.3.1 核心目标：在确保安全的前提下，快速恢复业务正常运行、数据完整性，降低业务中断造成的损失：

10.3.2 数据恢复：若数据完整性受损，AI 应验证备份数据的有效性，排除备份数据被篡改风险，并协调自动化运维平台从最近（恢复点目标 ≤15 分钟）的备份点进行数据回滚，优先恢复核心业务数据；

10.3.3 业务恢复：恢复完成后，AI 运行全量连通性测试、安全检测，确认业务服务等级协议（SLA）恢复至 99.9%以上，确保业务正常运行且无残留安全风险；

10.3.4 安全加固：AI 自动识别涉事系统、数据资产的安全漏洞，推送加固建议，协助运维人员完成安全加固，防范同类事件再次发生。

10.4 根因分析与复盘

10.4.1 核心目标：彻底查明事件根因，总结处置经验，优化 AI 模型和应急处置流程，提升后续应急处置能力：

10.4.2 根因追溯：调用历史数据资产访问模式、系统日志、攻击路径图谱，利用大语言模型生成

自然语言格式的安全事件复盘报告，明确事件根因、处置过程中的不足；

10.4.3 专家复核：复盘报告由安全专家复核确认后归档，确保根因分析的准确性；

10.4.4 改进建议：AI 结合复盘结果，推送 AI 模型优化、处置剧本完善、安全防护加固等改进建议，为后续持续优化提供支撑。

11 协同联动与报告机制

构建“人机协同、内外联动、分级上报”的协同联动与报告机制，强化各主体、各环节的协同配合，确保应急处置高效推进，同时满足国家监管要求，提升整体应急处置能力。

11.1 人机协同

坚持“AI 主战、人工终审、协同高效”的原则，明确人机分工，提升应急处置的效率与准确性，契合“技防+智防”融合的防护理念：

11.1.1 AI 主战，人审核：对于 90%的低危事件（一般事件），由 AI 闭环处置，处置完成后生成报告供人工复核，提升处置效率；

11.1.2 人主战，AI 辅助：对于高危事件（特别重大/重大事件），AI 提供情报建议、攻击路径分析、处置后果模拟推演，由应急指挥长决策，确保处置决策的科学性、合理性；

11.1.3 协同机制：建立 AI 与人工的实时交互渠道，人工可干预 AI 处置过程、调整处置策略，AI 可实时推送处置进度、预警信息，实现人机协同闭环。

11.2 对外报告

严格依据《国家网络安全事件报告管理办法》（2025 年）、《网络数据安全管理条例》等法律法规，建立规范的对外报告机制，确保报告及时、准确、合规：

11.2.1 报告模板：AI 系统应预设符合国家监管要求的报告模板，涵盖事件发生时间、受影响 IP、数据量、事件类型、处置进度等关键字段；

11.2.2 自动填充与人工确认：发现事件后，AI 自动填充报告关键字段，经人工确认无误后，通过加密渠道上报至对应主管部门，确保报告数据的准确性；

11.2.3 报告时限：严格遵循分级上报要求，特别重大/重大事件 2 小时内完成初步报告，较大事件 24 小时内完成报告，一般事件纳入月报汇总上报；

11.2.4 报告更新：事件处置过程中，AI 实时更新报告内容，及时上报处置进展、后续措施，确保主管部门全面掌握事件情况。

11.3 内部协同

建立组织内部安全、运维、业务等部门的协同联动机制，AI 系统实时推送事件信息至相关部门，明确各部门职责，确保处置过程协同高效；同时，加强与广西产学研科学研究院及相关单位的技术协同，共享威胁情报、处置经验，提升整体应急处置能力。

12 应急评估与改进要求

建立常态化的应急评估与改进机制，通过应急演练、模型优化、合规评估等方式，持续提升 AI 应急处置能力，确保规范的落地执行，契合我国人工智能安全标准体系建设和网络安全持续改进要求。

12.1 应急演练

定期开展 AI 应急处置能力红蓝对抗演练，模拟各类典型网络数据安全事件，检验 AI 模型性能、处置流程、协同机制的有效性：

12.1.1 演练频率：每半年至少开展 1 次针对 AI 应急处置能力的红蓝对抗演练，每年开展 1 次综合性应急演练，覆盖各类事件场景；

12.1.2 评估指标：明确演练评估指标，确保演练效果可量化、可评估，核心指标包括：攻击发现时间应小于 5 分钟；平均响应时间应小于 30 分钟；AI 自动化处置准确率应大于 90%；事件处置完成率应达到 100%；

12.1.3 演练总结：每次演练结束后，及时开展总结复盘，分析存在的问题，优化 AI 模型、处置剧本和协同机制。

12.2 模型投毒防范

结合人工智能安全防护要求，将模型投毒攻击防范纳入应急演练核心内容，验证训练数据隔离保护机制、模型异常检测机制的有效性：

12.2.1 演练内容：模拟训练数据篡改、恶意数据注入等投毒攻击场景，检验 AI 模型对投毒攻击的识别能力和抵御能力；

12.2.2 防护优化：根据演练结果，完善训练数据隔离保护措施、模型异常检测指标，提升 AI 模型的抗投毒能力，防范 AI 模型被攻击导致的应急处置失效。

12.3 持续优化

建立“事件处置—复盘分析—模型优化—能力提升”的闭环改进机制，持续提升 AI 应急处置能力：

12.3.1 模型优化：每次演练或真实事件后，应在 15 个工作日内完成 AI 模型的增量训练，优化误

报、漏报模型，提升模型识别精度；

12.3.2 流程优化：结合复盘结果，完善 AI 自动化处置流程、处置剧本，优化协同联动机制，提升处置效率；

12.3.3 技术升级：跟踪人工智能、网络安全领域的最新技术发展和攻击手法，及时升级 AI 应急处置系统，融入新型防护技术，确保系统的先进性、安全性。

12.4 符合性范围与推广应用

本规范由广西产学研科学研究院提出并归口，严格遵循国家网络安全、数据安全相关法律法规和人工智能安全标准体系要求，具有广泛的适用性和可操作性。执行本规范的单位，可依据本规范第 12 章的内容开展自我声明或第三方评估，以证明符合本文件要求；同时，鼓励各单位结合自身行业特点，在本规范基础上细化实施细则，推动规范落地执行。广西产学研科学研究院及相关单位应加强规范的推广应用，开展技术培训、案例分享等活动，推动 AI 应急处置技术的普及，助力提升我国网络数据安全应急处置的整体水平，为国家数字安全、数字经济高质量发展提供支撑。