

# 团 体 标 准

T/DGAG XXX—2026

## 政务数据人工智能安全运营总体要求

General requirements for AI security Operations in government data

XXXX—XX—XX 发布

XXXX—XX—XX 实施

广东省数字政务协会 发 布



目 次

前言 ..... II

1 范围 ..... 1

2 规范性引用文件 ..... 1

3 术语和定义 ..... 1

4 缩略语 ..... 2

5 总体要求 ..... 2

    5.1 总体框架 ..... 2

    5.2 总体原则 ..... 2

    5.3 运营组织 ..... 3

6 能力要求 ..... 3

    6.1 数据资产与标签底座 ..... 3

    6.2 运营模型的生命周期管理 ..... 3

    6.3 人机协同处置闭环 ..... 4

7 安全基线 ..... 4

8 运营场景 ..... 4

    8.1 采集汇聚与归集运营 ..... 4

    8.2 数据资源目录与分类分级运营 ..... 5

    8.3 内部域使用与运维操作运营 ..... 5

    8.4 跨部门共享交换运营 ..... 6

    8.5 公共数据对外开放运营 ..... 6

    8.6 授权运营加工与交付运营 ..... 6

    8.7 数据销毁/下线/退出运营 ..... 7

9 运营评价 ..... 7

    9.1 评价依据 ..... 7

    9.2 评价结果 ..... 7

附录 A （资料性） 政务数据安全运营场景 ..... 9

附录 B （资料性） 政策制度条款的对齐映射表 ..... 10

参考文献 ..... 11

## 前 言

本文件按照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。

本文件由广东省数字政务协会归口。

本文件起草单位：

本文件主要起草人：

# 政务数据人工智能安全运营总体要求

## 1 范围

本文件规定了各级政务部门对其持有、管理的政务数据开展安全运营时，引入人工智能技术增强运营能力的总体要求、能力要求、安全基线、运营场景和运营评价。

本文件适用于各级政务部门自身的安全运营团队、首席数据官工作体系，以及受委托的政务云平台运营方、安全运营中心（SOC）服务提供方、参与公共数据授权运营的运营机构，开展政务数据人工智能安全运营工作的规划、建设、实施和评价。

## 2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；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。

|                 |                        |
|-----------------|------------------------|
| GB/T 22080—2016 | 信息技术 安全技术 信息安全管理体系 要求  |
| GB/T 22239—2019 | 信息安全技术 网络安全等级保护基本要求    |
| GB/T 35273—2020 | 信息安全技术 个人信息安全规范        |
| GB/T 37988—2019 | 信息安全技术 数据安全能力成熟度模型     |
| GB/T 43697—2024 | 数据安全技术 数据分类分级规则        |
| GB/T 45654—2025 | 网络安全技术 生成式人工智能服务安全基本要求 |

## 3 术语和定义

下列术语和定义适用于本文件。

### 3.1

**政务数据** government data

公共管理和服务机构在履行职责过程中采集、制作、获取的，以电子化形式记录的数据资源。数据范围涵盖涉及自然人、法人和非法人组织的数据，以及政务部门履行职责过程中形成的政务运行数据。

### 3.2

**政务数据安全运营** government data security operations

以保障政务数据机密性、完整性、可用性和合规使用为目标的持续性活动集合，包括资产测绘、分类分级、风险监测、异常研判、处置闭环、审计追溯和持续改进。

### 3.3

**人工智能赋能** AI-enabled enhancement

应用机器学习、大模型、自然语言处理、知识图谱等人工智能技术，增强传统安全运营在数据识别精度、行为语义理解、多源关联分析、事件响应速度等方面的能力。AI 输出结果作为运营辅助参考，最终处置决策应由具备相应权限的人员作出。

### 3.4

**场景节点** scenario node

政务数据在其生命周期中穿越某一授权边界、网络域或责任主体边界的时刻或持续状态，是部署安全运营监测与管控措施的重点位置。

### 3.5

**运营资产图谱** operational asset graph

政务数据资产（包括数据库、数据表、文件、应用程序编程接口（API）等）及其承载系统、访问主体、授权关系、流转路径所构成的动态映射。

### 3.6

#### 目录漂移 catalog drift

业务系统数据结构（字段增删改、类型变化、敏感等级变化等）发生变化后，数据资源目录元数据未同步更新，导致目录声明与实际数据不一致的现象。

### 3.7

#### 重识别风险 re-identification risk

经去标识化或匿名化处理的数据，通过与其他数据关联、补充背景知识或借助算法推断等方式，重新识别出特定自然人或特定敏感信息的可能性。

## 4 缩略语

下列缩略语适用于本文件。

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

DLP：数据防泄露（Data Loss Prevention）

KMS：密钥管理系统（Key Management System）

MFA：多因素认证（Multi-Factor Authentication）

PII：个人身份信息（Personally Identifiable Information）

RAG：检索增强生成（Retrieval-Augmented Generation）

SIEM：安全信息和事件管理（Security Information and Event Management）

SOC：安全运营中心（Security Operations Center）

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

## 5 总体要求

### 5.1 总体框架

政务数据人工智能安全运营采用“四层架构”，总体框架自上而下包括：

- 数据资产与标签底座层：以动态维护的政务数据资产台账、分类分级标签、运营资产图谱为基础，为上层场景提供机器可读、可传播的数据安全上下文；
- AI 赋能能力层：提供数据语义理解、敏感信息智能识别、行为基线学习、异常风险关联推理、大模型辅助研判、自动化策略推荐等能力；
- 运营场景层：围绕政务数据生命周期的七大关键场景节点（采集汇聚、目录与分级、内部域使用、跨部门共享交换、公共数据对外开放、授权运营加工交付、销毁退出）部署监测与管控；
- 人机协同处置闭环层：通过告警分级、人工研判、分类处置、复盘标记、反馈优化形成持续改进闭环，并以全流程审计留痕实现可追溯。

### 5.2 总体原则

政务数据人工智能安全运营应遵循以下原则。

- 分类分级、按级管控：数据分类分级是安全运营的基础。已完成分类分级的数据进入 AI 运营处理链时，应按照对应的分级保护要求实施管控；未完成分类分级的数据，不应作为 AI 模型训练语料、知识库语料或自动化处置的输入，仅可在受控沙箱中由 AI 辅助识别和分级，定级完成前应采取最严格的隔离措施。
- 流转布控、节点管控：安全运营应围绕数据在跨主体、跨网络域、跨授权边界流转的关键节点部署监测与管控措施，而非仅关注静态存储防护。

- c) 最小授权、用途绑定：应核定具体业务场景、用数单位、数据范围和有效期限，实施最小必要授权、用途绑定和期限控制，杜绝超范围、超期限使用。
- d) 可追溯审计、按需共享：安全运营全过程应可审计、可追溯；运营报告的公开范围和内容，应按照国家及广东省政务数据共享开放相关规定执行，涉及敏感风险信息、漏洞细节、告警信息的内容，应脱敏后按权限共享。
- e) 人机协同、人负其责：AI 告警和建议应全过程留痕；涉及阻断、关停、降级、上报等处置决定，应由具备相应权限的人员按照制度要求复核并签字确认。AI 是运营辅助工具，不应作为免除人员责任的依据。
- f) 涉密隔离、红线管控：涉及国家秘密的数据和工作秘密数据，不应输入非涉密 AI 处理链；涉及个人隐私和重要数据的，应在符合法律法规要求的安全域内处理。

### 5.3 运营组织

政务数据人工智能安全运营应明确各方职责，建立协同工作机制，具体包括：

- a) 数源单位：负责对本单位提供数据分级结果的真实性、共享审批结论的合规性、数据销毁指令的有效性负责，并对数据归集质量承担主体责任；
- b) 主管单位：负责统筹制定政务数据 AI 安全运营规范、组织监督评估、推动跨部门监测能力贯通和数据共享，承担综合监管责任；
- c) 首席数据官（CDO）体系：各级首席数据官应统筹本地区、本部门政务数据安全运营的资源协调、重大事项决策和跨部门协同；
- d) 安全责任单位（数据安全管理机构）：负责运营策略卡的审批、高风险处置的最终签字确认、重大安全事件的应急指挥；
- e) 运营单位（含 SOC、政务云运营方）：负责日常告警研判与处置、资产台账维护、运营报告编制，以及 AI 模型与工具的日常运行管理；
- f) 授权运营机构：负责对授权范围内的数据加工、产品交付安全负责，接受主管单位和数源单位的监督，并按要求报送运营报告。

## 6 能力要求

### 6.1 数据资产与标签底座

政务数据资产台账应动态维护，分类分级标签应实现机器可读、可传播，并具备以下能力：

- a) 应支持对数据库、数据表、文件、API、应用系统的全量资产自动发现与测绘，形成覆盖采集、存储、使用、共享、开放、销毁全环节的资产视图；
- b) 应建立与 GB/T 43697—2024 及属地信息化主管部门的政务数据分类分级相关要求相一致的标签体系，并与数据资源目录联动；
- c) AI 识别结果应可解释，应输出命中规则或模型依据、置信度、样本摘要和判定过程，不应直接使用不可解释的黑盒标签驱动自动阻断；
- d) 标签变更、授权变更、资产上下线等关键操作应全量留痕，支持溯源查询。

### 6.2 运营模型的生命周期管理

运营所用的AI模型（包括行为基线模型、异常检测模型、敏感识别模型、大模型研判组件等）应实施全生命周期管理：

- a) 行为基线模型应支持周期性重训练和手动锁定机制；在重大活动保障、敏感时期等特殊阶段，应启动模型保守模式或冻结模型版本，避免模型漂移导致误判；
- b) 模型版本、训练数据窗口、特征集合、超参数配置、上线时间应全过程留痕，当告警结论被审计或申诉质疑时，应能够复现当时的判定依据；
- c) 应建立模型性能监测机制，对误报率、漏报率、准确率、漂移幅度等指标进行持续监测，当指标低于预设阈值时应触发模型重训练或下线评估；
- d) 模型上线前应通过测试数据集和对抗样本验证，评估其准确率、鲁棒性和安全性；模型退役或下线后应归档历史版本，确保历史告警可回溯。

### 6.3 人机协同处置闭环

应建立包含告警、分级、研判、处置、复盘、优化全流程的人机协同处置闭环：

- a) 应建立分级告警机制，告警级别、定义、处置时限、上报路径应符合附录 A 中表 A.2 的要求；
- b) 处置动作应至少包括放行、拦截、隔离、降级、关停、上报等类型；每一项处置动作均应录入处置理由、操作人、操作时间；
- c) 应建立复盘机制，对已处置告警定期标记真阳性、误报、漏报，并将标记结果反馈到模型训练和规则调优环节；
- d) 对超出运营单位处置权限或定级为 L3 的告警，应按照应急预案升级上报；涉及违法犯罪线索的，应按规定移送有关部门。

## 7 安全基线

政务数据人工智能安全运营所使用的 AI 引擎、模型、工具和平台，自身应满足以下安全基线要求：

- a) 部署域隔离：AI 引擎应运行在政务专网或经批准的专用安全域内，不应直接暴露于公共互联网；与政务专网外部模型 API 之间的调用，应履行审批程序，并进行数据出域安全检查和敏感信息过滤；涉及数据出境的，应符合《数据出境安全评估办法》等国家有关规定；
- b) 训练与知识库数据安全：运营 AI 的知识库、提示词模板库、样本库和微调数据集，不应直接包含未经脱敏处理的真实政务敏感数据原文；确需使用政务数据进行模型训练、微调或提示词工程时，应采取去标识化、合成数据生成等安全措施，并经数据安全责任人审批；不应将涉密数据、工作秘密数据输入外部大模型服务；
- c) 提示词安全防护：面向运营人员、开发人员开放的大模型交互入口，应部署提示词注入防护、越权指令拦截、敏感提问识别等安全措施，防止通过提示词诱导模型泄露敏感信息或执行违规操作；
- d) 模型输出安全过滤：AI 输出内容应叠加内容安全过滤机制，对个人信息、敏感政务信息、涉密线索、有害内容进行识别和拦截；生成式 AI 服务还应符合 GB/T 45654—2025 的相关要求，按规定对生成内容进行标识；
- e) 对抗鲁棒性：AI 检测模型应具备基本的对抗样本防御能力，定期进行对抗测试，防止攻击者通过字符变形、编码绕过、注入干扰等方式规避检测；
- f) 模型供应链与版本管理：应建立模型来源登记机制，对第三方模型、开源模型、预训练权重的来源、版本、授权许可、安全检测结果进行登记；不应使用来源不明、未经安全评估的模型组件；
- g) 访问权限与审计：对 AI 运营工具的管理控制台、模型训练平台、知识库管理后台应实施基于角色的访问控制（RBAC）和 MFA 认证；所有 AI 辅助决策、模型调用、知识库查询、策略生成等操作，应全量留痕并可审计；
- h) API 调用安全：AI 引擎对外提供的 API 接口应实施身份认证、流量控制、调用频次限制、输入校验和异常调用监测，防止接口被滥用或作为数据渗漏通道；
- i) 模型下线与数据清零：模型退役、替换或运营项目终止时，应同步清理对应的训练数据副本、缓存、向量索引和临时文件，涉及敏感数据的应按销毁流程处置；
- j) 决策留痕与可追溯：所有 AI 辅助研判结论、策略推荐、自动化处置记录，应留存决策依据、输入摘要、输出版本、操作人员、时间戳等信息，审计日志保存期限应不少于 6 个月，法律法规另有规定的从其规定。

## 8 运营场景

### 8.1 采集汇聚与归集运营

#### 8.1.1 场景描述

政务部门生产系统向省、市政务大数据中心汇聚节点归集数据的场景。

主要风险在于数据源多、数据口径杂、结构变化频繁，隐藏的PII、重要数据和高敏字段可能随汇聚被集中放大；同时存在数据漏报、错报、未授权归集等问题。

### 8.1.2 工作要求

工作要求如下：

- 应采用 AI 辅助的全量资产自动发现与测绘，对汇聚通道流量、库表结构、文件落点、接口调用进行扫描，对未登记、无目录、无标签的数据应触发“待定级告警”，并将其置于隔离区直至完成定级；
- 应采用语义理解与规则混合的 AI 识别能力，对汇聚数据进行字段内容级的敏感信息识别，识别结果应形成机器可读标签并回写至数据资源目录或元数据库；
- 应建立汇聚数据量、结构、值域、来源的时序基线，当数据源推送异常字段、超比例 PII、与历史模式显著偏离或在非授权时间窗口推送数据时，AI 应输出风险等级和处置建议；
- 对跨地域、跨层级的归集通道，应实施传输加密、通道认证和完整性校验，AI 应辅助监测异常传输行为。

## 8.2 数据资源目录与分类分级运营

### 8.2.1 场景描述

数据资源目录注册、变更、发布，以及分类分级标识维护的场景。

目录的准确性直接影响授权边界——目录错误将导致授权错误和管控失效。

### 8.2.2 工作要求

工作要求如下：

- 应建立 AI 分级推荐引擎，以目录元数据和采样内容为输入，按照国家及属地信息化主管部门分类分级要求输出分级建议和置信度；低置信度结果应通过人工复核确认，高置信度结果可按制度规定进行批量标注并由人工抽检；
- 当业务系统字段增删改、数据类型变化、敏感属性变化导致目录声明与实际数据不一致时（即目录漂移），AI 应触发目录漂移告警，提示数源单位及时更新目录；
- 应定期对已发布目录与实际数据的一致性进行 AI 辅助核验，对长期未更新、长期无访问但存在敏感数据的目录条目，应触发复核流程；
- 分类分级标签应在数据共享、开放、授权运营等环节随数据一并流转，不应在流转过程中丢失或被篡改。

## 8.3 内部域使用与运维操作运营

### 8.3.1 场景描述

政务云、政务大数据平台内的分析查询、报表生成、批量处理作业、开发测试、运维人员访问等内部数据使用场景。

此类场景中，合法账号下的越权查询、批量导出、违规下载、运维侧数据渗漏最为隐蔽。

### 8.3.2 工作要求

工作要求如下：

- 应以无监督或弱监督学习构建“用户—角色—数据对象”多维行为基线，覆盖查询量、访问表集合、结果集规模、操作时间窗、来源 IP 和终端等维度；对普通经办人突然扫描全表、夜间批量导出、只读账号异常路径访问、离职前集中下载等行为，应输出高置信度告警；
- AI 应对高风险 SQL 和数据操作模式（包括但不限于 UNION 拼接注入、绕过视图的直表访问、无 WHERE 条件的全表查询、布尔盲注、时间盲注、跨库 JOIN 敏感表、存储过程滥用等）进行语义级分析，并结合被访问对象的敏感级别进行综合风险评级；
- 对 DBA、云平台运维、批量任务账号、应急账号的操作应实施全量审计；AI 应用于压缩海量日志中的正常操作噪音，将分析资源聚焦于偏离基线的异常操作序列；

- d) 对批量导出、跨安全域数据复制、敏感表 JOIN 操作等高风险行为，应结合工单系统、审批系统数据进行交叉校验；无对应审批记录的应触发告警；
- e) 开发测试环境使用生产数据的，应经过脱敏审批；AI 应辅助扫描开发测试环境中是否存在未脱敏的真实敏感数据。

## 8.4 跨部门共享交换运营

### 8.4.1 场景描述

数据需求申请、数源单位审批、平台传输、使用方消费的跨部门共享交换全链路场景。

核心难点在于跨主体信任边界管控，条件共享的审批结论与实际使用量、使用范围之间可能存在不一致，二次转发风险较高。

### 8.4.2 工作要求

工作要求如下：

- a) AI 应将共享申请单中核定的用途、字段范围、数据量、调用频次和有效期限，与实际传输量、接收方查询行为、访问 IP 等日志进行对齐校验；发现字段范围与审批不符、数据量或调用频次超出审批核定阈值、授权期限届满后继续使用等情形时，应触发告警并按级别处置；
- b) 应通过 AI 补全多源日志（含平台日志、网络流量日志、数据库审计日志、API 网关日志），还原“数源→传输通道→接收方→再分发”的全链路数据流转路径；
- c) 对接收方出口流量应进行抽样语义监测；当出现“刚从共享平台获取的高敏数据样本向外网、第三方平台或非授权系统转发”的信号时，应发出高等级告警；
- d) 对接入共享平台的部门节点，应建立节点信用评分模型；对多次出现超范围使用、二次转发的接收方，AI 应建议提升其审批严格度或限制共享方式。

## 8.5 公共数据对外开放运营

### 8.5.1 场景描述

公共数据开放平台、开放超市向社会公众开放数据API和数据集下载的场景。

面向社会面开放，存在爬虫滥用、批量拉取重建数据集、通过多数据集关联推断泄露个人信息或敏感信息、智能问答接口被诱导输出敏感内容等风险。

### 8.5.2 工作要求

工作要求如下：

- a) 对开放 API 调用应建立调用者画像，包括调用者标识、UA、IP 分布、调用频次、结果集体积、时间序列、参数分布等特征；
- b) AI 应检测爬虫式批量拉取、凭据共用、异常聚合查询、分布式协同抓取等行为，并按开放数据的分级属性采取限速、熔断、临时封禁或升级核验等措施；
- c) 对已开放数据集的组合关联性应定期（不少于每半年一次）进行 AI 辅助风险评估；
- d) 当外部信息环境变化导致关联推断风险升高时，应触发降敏处理、撤回下架或升级授权访问机制；
- e) 当开放门户提供智能搜索、智能问答等生成式交互服务时，应在 AI 输出侧叠加内容安全过滤机制，过滤日志应予留存；不应通过智能交互接口输出未开放的敏感数据；
- f) 对无条件开放类数据的下载行为应做基础监测；
- g) 对有条件开放类数据（受控开放）的访问应进行严格身份核验、用途登记和使用行为审计。

## 8.6 授权运营加工与交付运营

### 8.6.1 场景描述

数据提供单位通过授权运营平台向授权运营机构提供数据，由运营机构加工形成数据产品和服务并对外交付的场景。

核心要求是原始数据不出域、脱敏或匿名化到位、防止超范围加工使用、防范数据产品的重识别风险。

### 8.6.2 工作要求

工作要求如下：

- a) 应持续监测授权运营加工环境中是否存在原始表被直接导出、脱敏环节被绕过进行直表访问、通过调试端口或临时挂载通道渗漏数据等情况；AI 应将分散在数据库审计、容器日志、网络流量、文件操作等多源的审计线索聚合为“出域风险事件”；
- b) 授权运营机构对数据的每次加工任务应建立任务标签，标签内容至少包含授权范围 ID、数据提供单位、起止时间、加工算法、产品交付物清单；AI 应进行“授权范围—实际用量—交付物清单”三者的一致性校验；
- c) 应定期（产品交付前必做、加工期间定期抽检）开展 AI 驱动的重识别风险评估，对脱敏或匿名化后的数据样本在受控环境下进行链接攻击、差分攻击等模拟测试；评估结果作为是否允许产品对外交付的重要依据，评估方法和通过阈值应符合国家相关标准要求；
- d) AI 可辅助授权运营机构按授权运营管理办法要求，对年度报告所需的结构化证据进行收集和初稿生成；AI 生成的报告初稿应经人工审核确认后方可作为正式报告提交，最终报告由法定代表人或授权签字人签字负责；
- e) 对授权运营数据产品的对外服务接口，应参照 7.5 的对外开放运营要求实施调用监测、限流和异常行为检测。

## 8.7 数据销毁/下线/退出运营

### 8.7.1 场景描述

过期数据、项目下线、授权到期、业务终止等情况下的数据副本清理和日志留存边界场景。

常见风险是数据残留于快照、备份、临时目录、测试环境克隆、向量索引、RAG 缓存、AI 模型权重等位置。

### 8.7.2 工作要求

工作要求如下：

- a) AI 应辅助扫描快照、备份、临时目录、交换文件、回收站、测试环境克隆、RAG 向量索引、缓存、AI 模型微调权重等位置，识别应销毁的高敏数据残留痕迹，并生成残留清单；
- b) 销毁动作（包括逻辑删除、介质覆写、消磁、密钥销毁等）、销毁时间戳、操作人、销毁方式、加密密钥销毁证明材料，应纳入不可篡改的审计链；销毁完成后应进行残留数据验证，出具销毁验证报告；
- c) AI 应检测异常模式，包括但不限于同一工单多次出现“销毁后又重现”、销毁范围与授权范围不一致、销毁前有异常批量复制行为等；
- d) 法律法规要求留存的日志、审计记录，在数据销毁后仍应按规定期限保存，不应随数据一并销毁；数据销毁记录本身应长期归档。

## 9 运营评价

### 9.1 评价依据

评价工作应依据 GB/T 37988—2019 数据安全能力成熟度模型的方法论，结合本文件规定的总体原则、能力要求和场景工作要求，对运营单位的 AI 安全运营能力进行综合评估。

### 9.2 评价结果

评价结果可作为以下工作的重要依据：

- a) 政务数据安全运营工作考核和资金预算安排；
- b) 数据共享、开放、授权运营资格审核；
- c) 运营单位（含 SOC、政务云运营方、授权运营机构）服务能力评估和续聘参考；

- d) 重大风险隐患通报、整改约谈和责任追究。

附录 A  
(资料性)  
政务数据安全运营场景

A.1 场景-风险-AI 赋能点速查

政务数据安全运营七类场景的主要风险与 AI 赋能点速查见表A.1。

表 A.1 政务数据安全运营场景—风险-AI 赋能点速查矩阵

| 场景编号 | 场景名称        | 主要风险                    | AI赋能点                              |
|------|-------------|-------------------------|------------------------------------|
| S-1  | 采集汇聚与归集     | 隐藏 PII 放大、口径杂乱、未授权归集    | 全量资产测绘、敏感字段智能识别、汇聚基线异常检测           |
| S-2  | 数据资源目录与分类分级 | 目录漂移、目录错导致授权错           | AI 分级推荐、目录漂移监测、目录一致性核验             |
| S-3  | 内部域使用与运维操作  | 合法账号越权操作、运维渗漏、开发测试环境未脱敏 | UEBA 行为基线、SQL 语义风险识别、日志噪音压缩、多源交叉校验 |
| S-4  | 跨部门共享交换     | 超范围使用、到期继续使用、二次转发       | 申请—使用对账、全链路还原、出口流量监测、节点信用评分        |
| S-5  | 公共数据对外开放    | 爬虫滥用、关联推断泄露、智能接口输出失控    | 调用者画像、组合风险分析、输出安全过滤、分级访问控制         |
| S-6  | 授权运营加工与交付   | 原始数据出域、超范围加工、重识别风险      | 出域风险聚合、用量一致性校验、重识别风险评估、产品交付核验      |
| S-7  | 数据销毁/下线/退出  | 数据残留于快照/缓存/向量索引         | 影子副本扫描、销毁证明链异常检测、残留验证辅助            |

A.2 AI 安全运营告警处置 SOP

AI 安全运营告警分级与处置时限参考见表A.2。

表 A.2 AI 安全运营告警分级处置 SOP

| 告警级别    | 定义                         | 处置时限 | 处置方式                | 上报路径                |
|---------|----------------------------|------|---------------------|---------------------|
| L1—自动处置 | 确定性规则触发的低风险告警，无人工研判必要      | 实时   | 自动阻断/放行，全量记录日志      | 记录日志备查              |
| L2—人工研判 | 行为偏离基线、中等置信度告警、需人工确认       | ≤4小时 | 安全运营团队人工研判，按结论处置    | 安全运营团队              |
| L3—应急关停 | 高敏数据外传、疑似数据泄露、大规模异常访问等高危信号 | ≤1小时 | 立即启动应急预案，必要时先关停后补流程 | 按预案上报安全责任人和公共数据主管部门 |

**附 录 B**  
**(资料性)**  
**政策制度条款的对齐映射表**

**B.1 政策制度条款对齐映射**

本标准主要条款与广东省现行公共数据管理制度重点条款的对齐映射关系见表B.1，供实施参考。

**表 B.1 政策制度条款对齐映射表**

| 序号 | 广东省制度条款                               | 本标准对应条款                                                    |
|----|---------------------------------------|------------------------------------------------------------|
| 1  | 《广东省公共数据管理办法》（省政府令第290号）第44—47条（安全管理） | 第5.2条（总体原则）、第5.3条（运营组织）、第8章（运营评价）                          |
| 2  | 《广东省公共数据资源授权运营管理办法》第二十条（加工安全与产品交付）    | 第8.6条（授权运营加工与交付运营）                                         |
| 3  | 《广东省公共数据资源登记管理实施细则（试行）》               | 第6.1条（数据资产与标签底座）、第7.2条（数据资源目录与分类分级运营）                      |
| 4  | 《关于加快公共数据资源开发利用的实施意见》                 | 第5.2 d) 条（可追溯审计、按需共享）、第7.5条（公共数据对外开放运营）、第8.6条（授权运营加工与交付运营） |
|    | ...                                   | ...                                                        |

## 参 考 文 献

- [1] 《广东省公共数据管理办法》（广东省人民政府令第290号，2021年发布）
  - [2] 《广东省公共数据资源登记管理实施细则（试行）》（粤政数〔2025〕21号）
  - [3] 《广东省公共数据资源授权运营管理办法》（粤政数〔2026〕9号）
  - [4] 中共广东省委办公厅、广东省人民政府办公厅关于加快公共数据资源开发利用的实施意见
  - [5] 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等七部门令第15号，2023年）
  - [6] 数据出境安全评估办法（国家互联网信息办公室令第11号，2022年）
-