《政务数据人工智能安全运营总体要求》团体标准

编制说明

标准编制组

2026 年 8月

目录

一、工作简况

（一）任务来源

为贯彻落实《国务院关于深入实施“人工智能+”行动的意见》（国发〔2025〕11号）关于“提升安全能力水平。推动模型算法、数据资源、基础设施、应用系统等安全能力建设”的文件要求，通过标准化手段规范政务信息化人工智能应用安全治理能力建设，建立统一的安全能力框架与评价要求，提升政务数据人工智能安全运营水平，推动人工智能在政务领域安全、规范、可持续发展，特提出制定《政务数据人工智能安全运营总体要求》团体标准。

根据《关于〈政务数据接口安全风险人工智能识别技术规范〉等两项团体标准的立项公告》（粤政务协会〔2026〕12号）相关要求，该团体标准由广东省数字政务协会负责归口与发布，由领信数科信息技术有限公司牵头负责标准编制工作。

（二）起草单位

本标准的主要起草单位拟包括：广州市数字政府运营中心、江门市大数据管理中心、梅州市大数据管理中心、潮州市机构编制研究与事务中心、领信数科信息技术有限公司、广州云烽信息咨询有限公司、广东机场白云信息科技股份有限公司、广东交通职业技术学院、广州卫生职业技术学院共9家企事业单位。

二、目的和意义

随着生成式人工智能和大模型技术的快速发展，人工智能正加速在政务服务、城市治理、行政执法、公共服务等领域落地应用。智能

客服、政策问答、智能审批辅助、政务知识库问答、智能决策辅助等应用场景不断涌现，显著提升了政务服务效率和治理能力。

然而，在政务信息化人工智能应用快速发展的同时，其安全风险问题也逐渐显现。**一是**模型安全风险突出，大模型可能受到提示注入、越狱攻击及模型滥用等攻击，可能导致敏感政务数据泄露或产生错误决策建议；**二是**政务数据安全风险加大，人工智能应用涉及大量政务数据在采集、训练、推理和知识增强等环节的流转，存在数据泄露、越权访问及敏感数据外泄等隐患；**三是**现有政务信息化安全体系主要围绕网络安全和数据安全建设，对于人工智能模型调用、推理接口以及提示词输入输出等新型安全风险缺乏系统防护能力。

目前国内已发布的相关标准主要包括 T/ZGCMCA 023-2025《人工智能 智能体内生安全技术要求》、T/UNP 186-2026《生成式人工智能内容安全检测规范》、T/ZGCMCA 024-2025《人工智能 智能体互操作性接口规范》、T/ZGCMCA 006-2025《人工智能大模型 数据标注技术要求》、T/BJCSA 08-2025《人工智能应用安全监测与应急响应规范》等。这些标准主要从人工智能的内容检测、数据标注、安全监测等方面进行了规定，但**主要侧重于技术能力或应用形态，缺少对政务数据的人工智能安全运营规范。**

为贯彻落实《国务院关于深入实施“人工智能+”行动的意见》（国发〔2025〕11 号）关于“提升安全能力水平。推动模型算法、数据资源、基础设施、应用系统等安全能力建设”和《广东省加快数字政府领域通用人工智能应用工作方案》（2023 年 12 月 21 日发）

关于“增强数据安全保障能力”等政策文件精神，制定《政务数据人工智能安全运营总体要求》将填补政务数据人工智能安全运营能力方面的标准空白，为各级政务单位开展政务数据的人工智能安全运营提供统一技术依据。

三、编制原则

本标准制定遵循以下原则:

1. 协调性

本标准严格遵循国家、广东省及各地市各级法律法规与宏观政策导向，密切结合现代信息技术发展现状与广东省数字政府改革建设相关要求。在充分吸纳《广东省公共数据管理办法》（省政府令第290号）、《广东省公共数据资源授权运营管理办法》（粤政数〔2026〕9号）、《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）及《关于加快公共数据资源开发利用的实施意见》等政策文件精神的基础上，本标准紧扣政务信息化行业对政务数据安全运营提出的“高标准、高要求”，制定相关技术规范；并在编制过程中始终与国家现行政策体系、上位标准规范（如 GB/T 22080、GB/T 22239、GB/T 35273、GB/T 37988、GB/T 43697、GB/T 45654 等）及行业发展趋势保持协调统一。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关系见附录 B，以确保标准落地时有据可依、上下贯通。

2. 目标向导

本标准以“应用人工智能技术增强政务数据安全运营能力”为核

心目标，聚焦保障政务数据机密性、完整性、可用性和合规使用的根本宗旨。编制中坚持问题导向，围绕政务数据在采集汇聚、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、授权运营加工与交付、销毁退出等七大关键场景节点面临的实际风险（如敏感信息集中放大、目录漂移、合法账号越权、二次转发、关联推断泄露、重识别风险、影子副本残留等），明确总体框架、能力要求、安全基线与运营场景，确保标准条款有的放矢、直指痛点，避免泛化空泛。

3. 实用性

本标准在编制过程中，广泛结合广东省数字政府建设发展实际情况，以及现有政务云平台运营、安全运营中心（SOC）建设、政务数据分类分级与共享开放、公共数据授权运营等已开展的实践基础，从总体框架、总体原则、能力要求、安全基线、运营场景、运营评价等方面进行规范。编制组进行了充分的调查研究与必要验证，吸收一线运营单位、首席数据官（CDO）体系及授权运营机构的实践经验，保障标准条款设置合理可行、切实可用，具有良好的实用性和可操作性。

4. 可行性

本标准注重落地实施的可行性。条款设置遵循“人机协同、人负其责”的务实路径，区分 L1 自动处置、L2 人工研判、L3 应急关停的分级处置机制，既鼓励 AI 赋能提效，又明确最终处置决策由具备相应权限的人员作出，避免脱离实际的无条件自动化。对模型上线前的测试验证、上线后的性能监测与重训、退役后的版本归档等均给出可执行的闭环要求，并以 GB/T 37988—2019《数据安全能力成熟

度模型》的方法论作为运营评价依据，使标准可度量、可考核、可改进。

5. 规范性

本标准严格按照 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》确立的规则进行编写。编制组在起草阶段即对标该导则，对标准文本的要素构成、编排格式及表述逻辑进行规范，确保本标准在结构与形式上与国家标准化文件的基本规则保持一致，术语和定义、缩略语、引用文件等要素齐备、表述严谨。

6. 公开透明

本标准的编制遵循团体标准公开、透明的工作程序，编制过程向社会相关方公开征求意见，充分吸纳政务部门、运营单位、行业专家与社会各界的建议。同时，标准内容本身强调运营全过程“可审计、可追溯、留痕可查”，并要求运营报告的公开范围与内容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、告警信息的内容脱敏后按权限共享，在保障安全的前提下提升运营的透明度与可监督性。

7. 可持续发展

本标准将政务数据安全运营视为持续性活动，注重长效演进与持续改进。标准确立了“告警—研判—处置—复盘—优化”的人机协同处置闭环与 AI 模型全生命周期管理机制（含周期重训练、版本冻结、

退役归档），并以运营评价结果驱动政务数据安全运营考核、预算安排与资格审核，推动运营能力在成熟度模型框架下持续螺旋上升，适应人工智能技术与数据安全风险态势的不断演化。

四、编制过程

1. 标准立项：2026 年 4 月 9 日，经协会审批，该标准项目正式立项，《关于〈政务数据接口安全风险人工智能识别技术规范〉等两项团体标准的立项公告》（粤政务协会〔2026〕12 号）经协会官网和全国团体标准信息平台公开发布。

2. 成立标准编制组：2026 年 5 月 6 日，协会正式发布公告公开征集参编单位，吸纳了广州市数字政府运营中心、江门市大数据管理中心、梅州市大数据管理中心、潮州市机构编制研究与事务中心、领信数科信息技术有限公司、广州云烽信息咨询有限公司、广东机场白云信息科技股份有限公司、广东交通职业技术学院、广州卫生职业技术学院等单位的专业技术骨干和相关人员组成标准编制组，主要负责标准文本的起草编制任务。在前期资料搜集和初步调研的基础上，形成标准制定工作方案，包括阶段进度安排、关键时间节点、人员工作分工等，并确立了内容大纲。

3. 编制初稿：2026 年 6 月 1 日—2026 年 8 月 21 日，标准编制组根据计划开展标准初稿编制工作，系统查阅了《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）《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）《广东省公共数据资源授权运营管理办法》（粤政数〔2026〕9 号）

等政策文件、《信息安全技术 数据安全能力成熟度模型》（GB/T 37988—2019）《数据安全技术 数据分类分级规则》（GB/T 43697—2024）等相关资料，在充分收集整理基础资料和开展调研论证的基础上，对标准框架和相关技术内容反复进行了修改，形成了标准（草案）。

4. 技术研讨:

2026 年 8 月 24 日，标准编制组在广东省广州市天河区黄埔大道西平云路 163 号广电平云广场广电科技大厦 5 楼会议室组织召开标准编制内部研讨会，邀请领信数科信息技术有限公司、广东机场白云信息科技股份有限公司、广东交通职业技术学院、广州卫生职业技术学院等起草代表参与讨论。参会代表基于各自的行业经验，提出了修改完善建议:

一是总体框架从一段文字扩充为“四层架构”（资产底座层→AI 能力层→场景层→处置闭环层）的结构化描述。

二是运营组织补充首席数据官（CDO）体系、授权运营机构、政务云平台方的职责，原 4 类角色扩充为 6 类。

三是 AI 运营工具自身安全基线从原稿 3 句话单薄内容扩充为 10 项具体要求（部署域隔离、训练数据安全、提示词防护、输出过滤、对抗鲁棒性、模型供应链、权限审计、API 安全、模型下线清零、决策留痕）。

5. 形成征求意见稿:

2026 年 9 月 2 日，标准编制组在完成标准初稿的基础上，通过内部研讨、组内审稿及定向咨询等方式，针对标准草案中少数存在异

议的技术条款与表述进行反复讨论，在充分吸纳合理意见并对文本进行修改完善后，正式形成了标准（征求意见稿）及其编制说明。

五、标准的主要内容

（一）技术内容说明

本标准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框架、总体原则、运营组织、能力要求、安全基线、运营场景和运营评价。本标准拟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。具体如下所示：

（1）总体要求：描述政务数据人工智能安全运营的总体框架、总体原则和运营组织内容。

（2）能力要求：描述数据资产与标签底座、运营模型的生命周期管理和人机协同处置闭环能力。

（3）安全基线：描述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，自身应满足部署域隔离、训练与知识库数据安全、提示词安全防护、模型输出安全过滤、对抗鲁棒性、模型供应链与版本管理、访问权限与审计、API 调用安全、模型下线与数据清零和安全基线要求。

（4）运营场景：描述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）的运营工作要求。

（5）运营评价：描述评价依据和评价结果应用。

（二）编制总体思路

标准中所有核心量化指标均基于行业实践、国标对标与技术可行性综合确定，确保指标“可落地、可验证、无超出现有技术水平的过高要求”，核心指标设定依据如下:

《政务数据人工智能安全运营总体要求》整个架构的设计逻辑可以归纳为“一个中心、两条主线、七类场景、三层闭环”：

1. 一个中心:

始终围绕“政务数据”这一防护对象，而不是围绕网络域、系统域或技术产品。这意味着标准回答的是“数据在哪里、被谁、以什么方式使用才安全”，而不是“买了什么安全设备”。

2. 两条主线:

横向主线——政务数据的生命周期与流转边界(采集汇聚 → 目录分级 → 内部使用 → 共享交换 → 对外开放 → 授权运营 → 销毁退出）

纵向主线——AI 赋能安全运营的通用能力（资产标签底座、模型生命周期、人机协同处置、AI 工具自身安全）

3. 七类场景:

把横向主线拆解为 7 个关键场景节点，每个场景独立成章，分别规定风险点和 AI 赋能要求。这样做的考虑是——政务数据安全风险高度依赖上下文，脱离场景谈“AI 赋能”会变成技术堆砌，只有

落到具体场景才能给出可执行的要求。

4. 三层闭环:

运营闭环: 告警 → 研判 → 处置 → 反馈 → 模型优化

责任闭环: 数源部门、公共数据主管部门、首席数据官、运营执行团队各司其职。

评价闭环: 对标 GB/T 37988 成熟度模型开展能力评估。

六、知识产权情况说明

本标准不涉及知识产权。

七、与现行相关法律、法规和强制性国家标准的关系

本标准与现行法律、法规和强制性国家标准均不存在冲突。

本标准符合《中华人民共和国个人信息保护法》《中华人民共和国数据安全法》《中华人民共和国标准化法》等法律法规，与《新一代人工智能发展规划》《广东省公共数据管理办法》等政策文件精神相契合,响应了人工智能发展和数字政府改革建设的政策部署与实践需求。

八、相关国内标准情况简要说明

与本标准较为相关的标准规范包括: 《人工智能 智能体内生安全技术要求》（T/ZGCMCA 023-2025）、《生成式人工智能内容安全检测规范》（T/UNP 186-2026）、《人工智能 智能体互操作性接口规范》（T/ZGCMCA 024-2025）、《人工智能大模型 数据标注技

术要求》（T/ZGCMCA 006-2025）、《人工智能应用安全监测与应急响应规范》（T/BJCSA 08-2025）等。这些标准主要从人工智能的内容检测、数据标注、安全监测等方面进行了规定，但主要侧重于技术能力或应用形态，缺少对于政务数据的人工智能安全运营规范。

总体来看，现有标准体系尚未形成针对“政务数据+人工智能+安全运营”场景的安全治理能力框架，尤其在大模型安全防护、政务数据安全保护、AI 接口安全及 AI 安全运营等方面缺乏统一规范。本标准的制定将填补政务数据人工智能安全运营能力方面的标准空白，为各级政务单位开展政务数据的人工智能安全运营提供统一技术依据。

九、重大分歧意见的处理经过和依据

本标准未产生重大分歧意见。

十、标准性质的建议

本标准作为推荐性标准，属于团体标准，供协会会员和社会自愿使用。

十一、标准实施建议

本标准批准发布实施后，尽快将本标准的批准信息通告编制小组成员单位及其他有关单位。积极组织本标准的宣传和培训，使涉及政务数据人工智能安全运营的相关单位与管理人员熟悉了解本标准的

内容、特点，保证标准的顺利实施，使标准的应用真正落到实处。

为全面掌握本标准的执行情况，鼓励使用单位和相关单位将本标准的执行情况以及所发现的问题及时反馈到归口团体或本标准的起草单位，为标准完善工作做好准备，不断提高标准的科学性、合理性、协调性和可操作性。

十二、其他应说明的事项

无。