ICS 35.240.99
CCS L67

团　　体　　标　　准

T/BYIDA 002—2026

人工智能教育平台安全可信建设规范

Specification for Secure and Trustworthy Construction of Artificial Intelligence Education Platforms

（征求意见稿）

2026 -XX - XX 发布　　2026 -XX -XX 实施

广州市白云区数智化发展协会　　发 布

目　　次

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这些专利的责任。

本文件由广州中长康达信息技术有限公司提出。

本文件由广州市白云区数智化发展协会归口。

本文件起草单位：广州中长康达信息技术有限公司、广州市云享数据科技有限公司、广州市电子行业协会、广州市白云区数智化发展协会、广州云之能科技有限公司。

本文件主要起草人：***,田传广,许鸿平,陈伟杰,严升兰。

本文件是首次发布。

引　言

人工智能教育平台作为融合人工智能技术与教育教学场景的新型基础设施，承载着教学资源管理、实训环境提供、学习行为分析等关键功能。平台涉及大量师生个人信息、教学核心数据及人工智能模型资产，其安全可信建设直接关系到教育数据安全、知识产权保护和教学秩序稳定。本文件针对人工智能教育平台的特殊应用场景，建立覆盖身份可信、数据可信、环境可信、运维可信的技术规范体系，为平台建设、运营、评估提供可量化、可验证、可落地的技术依据。

人工智能教育平台安全可信建设规范

1 范围

本文件规定了以人工智能为核心技术的教育系统（含教学、实训、资源管理、科研、赛事）在安全框架、数据可信、实训隔离、安全运维、应急响应等方面的技术与管理要求。

本文件适用于广州市中小学、中高职、本科及继续教育各类教育平台。

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 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 36627-2018 信息安全技术 可信计算规范
GB/T 37988-2019 信息安全技术 数据安全能力成熟度模型
GB/T 25070-2019 信息安全技术 云计算安全扩展要求
GB/T 36639-2018 信息安全技术 容器安全技术要求
GB/T 22240-2020 信息安全技术 人工智能安全准则（报批稿）
GB/T 39412-2020 信息安全技术 代码安全审计规范
GB/T 41819-2022 信息安全技术 可信执行环境服务规范

3 术语和定义

下列术语和定义适用于本文件。

3.1

人工智能教育平台 artificial intelligence education platform

基于人工智能技术，为教育场景提供教学支持、实训环境、能力评测等服务的软硬件系统。

3.2

安全可信框架 secure and trustworthy framework

涵盖身份可信、数据可信、环境可信、行为可信的技术与管理保障体系。

3.3

实训容器 practical training container

为人工智能教学实训提供隔离计算环境的虚拟化技术单元，包括Docker容器、Kata容器、gVisor等安全容器技术实现的运行实例。

3.4

数据分类分级 data classification and grading

依据数据敏感程度和一旦遭到破坏后的影响范围，对教学数据实施差异化安全保护的管理方法。

4 总体原则

平台应建立覆盖身份可信、数据可信、环境可信、运维可信四层的安全可信框架，并满足以下量化要求：

a）安全域划分不少于4个：公共接入域、应用服务域、数据存储域、运维管理域；

b）安全域间访问控制策略匹配度≥95%，默认拒绝所有跨域访问，仅开放白名单端口；

c）启动链度量值存储于符合36627-2018要求的可信计算模块，度量日志不可篡改。

5 数据可信要求

5.1 数据分类分级管理

数据分类分级应满足表1要求。

表1 数据分类分级量化标准

数据级别	定义标准	数据示例
核心级	泄露或篡改导致平台瘫痪、重大经济损失或严重社会影响	AI核心模型算法、平台根密钥、学生生物特征模板
敏感级	泄露或篡改导致个人权益严重受损或组织声誉损害	学生身份信息、学习行为轨迹、成绩数据、教师评价
重要级	泄露或篡改影响教学正常开展或模型性能下降	训练数据集、模型参数、课程资源
内部级	泄露或篡改对业务影响有限	系统日志、监控指标、配置参数

5.2 数据全生命周期保护

5.2.1 采集

数据采集应满足以下要求：

a）个人信息采集前明示告知率100%，敏感个人信息单独同意率100%；

b）采集字段与业务功能必要性匹配度100%；

c）学生行为数据采集频率≤1次/min，实时行为分析场景除外。

5.2.2 传输

数据传输应满足以下要求：

a）互联网传输强制TLS 1.3，国密场景优先GM/T 0024；

b）商密算法：SM2密钥交换、SM3完整性校验、SM4数据加密；国际算法：ECDHE P-256、AES-256-GCM、SHA-384；

c）数字证书有效期≤397d，到期前30d自动预警，到期前7d自动轮换；

d）消息认证码或数字签名覆盖率100%，重放攻击防护窗口≥60s。

5.2.3 存储

数据存储应满足以下要求：

a）核心级、敏感级数据静态加密覆盖率100%，加密密钥与数据物理分离；

b）密钥轮换周期≤90 d，密钥销毁后数据不可恢复；

c）核心数据存储持久性≥99.999999999%（11个9）；

d）敏感数据删除后，逻辑删除复写≥3次，物理介质粉碎处理。

5.2.4 使用

数据使用应满足以下要求：

a）敏感数据展示场景动态脱敏率100%，脱敏规则：身份证号显示前3位后4位，手机号显示前3位后4位，姓名显示首字；

b）生产数据禁止直接进入开发测试环境，测试数据与生产数据相似度≤80%；

c）批量导出（>100条记录）需双人审批+数字水印（包含用户ID、时间戳）。

5.2.5 销毁

数据销毁应满足以下要求：

a） 敏感数据到达保留期限后，自动触发销毁流程，延迟≤24 h；

b） 提供销毁证明，包含数据标识、销毁时间、执行人、验证哈希值；

c） 存储介质报废处置过程录像留存≥90d。

5.3 数据备份与恢复

数据备份与恢复应满足表2要求。

表2数据备份与恢复量化指标

指标项	核心级数据	敏感级数据	重要级数据	内部级数据
RTO	≤4 h	≤8 h	≤24 h	≤72 h
RPO	≤15 min	≤1 h	≤4 h	≤24 h
备份频率	实时同步+每 4h 增量+每日全量	每 6h 增量+每日全量	每日增量+每周全量	每周全量

指标项	核心级数据	敏感级数据	重要级数据	内部级数据
副本数量	本地 2 份+异地 3 份	本地 2 份+异地 2 份	本地 1 份+异地 1 份	本地 1 份
恢复演练	每月 1 次	每季度 1 次	每半年 1 次	每年 1 次
演练成功率	≥99%	≥95%	≥90%	≥85%

6 实训环境安全要求

6.1 容器隔离与资源控制

容器隔离与资源控制应满足以下要求：

a）采用Kata Containers、gVisor或等效安全容器技术，提供轻量级虚拟机级隔离；

b）PID、Network、Mount、IPC、UTS、User命名空间隔离率100%；

c）CPU限制误差≤5%，内存硬限制准确率100%，存储配额超配禁止；

d）默认资源上限：CPU≤2核，内存≤4 GB，存储≤20 GB，网络带宽≤100 Mbps，PIDS≤100；

e）触发限制后，容器优雅终止时间≤30 s，强制终止≤10 s。

6.2 网络隔离

网络隔离应满足以下要求：

a）实训网络与其他网络VLAN隔离，ACL策略匹配率100%；

b）实训容器默认DROP所有出站连接，白名单开放率≤5%；

c）禁止访问财务、人事、OA等敏感系统，IP黑名单覆盖率100%；

d）实训环境强制使用平台内建DNS，禁止自定义DNS服务器，DNS查询日志留存≥6个月；

e）实训容器间流量默认禁止，经审批后开放特定端口，流量镜像至入侵检测系统。

6.3 实训实例生命周期管理

实训实例生命周期管理应满足表3要求。

表3 实训实例生命周期量化管理

阶段	时间阈值	技术要求
创建	≤60 s	从请求提交到环境就绪，包含镜像拉取、配置注入、网络绑定
初始化	≤30 s	安全基线检查通过后方可开放用户访问
运行	按教学计划	资源使用率监控周期≤30 s，异常行为检测实时告警
空闲回收	空闲≥30 min	自动暂停，释放计算资源；恢复时间≤20 s

阶段	时间阈值	技术要求
到期销毁	教学结束+24h 宽限期	强制销毁，数据卷自动清理，日志归档保留 6 个月
紧急销毁	检测到高危威胁时	单实例≤10s，同批次≤60s

6.4 日志审计

日志审计应满足以下要求：

a）实训操作指令、文件访问、网络连接、系统调用日志采集率100%，丢包率≤0.1%；

b）日志包含用户ID、实例ID、时间戳（毫秒级）、操作类型、操作对象、操作结果、来源IP；

c）日志传输采用TLS 1.3或Syslog over TLS，端到端加密覆盖率100%；

d）在线存储≥6个月，归档存储≥3年，核心安全事件永久保存；

e）日志写入WORM存储或区块链存证，篡改检测率100%；

f）最近7天数据查询响应时间≤3 s，跨月复杂查询≤30 s。

7 安全技术要求

7.1 端口与服务管理

端口与服务管理应满足以下要求：

a） 开放端口清单化，未登记端口默认阻断；

b） 135~139、445、3389 等高危端口互联网暴露率 0%；

c） 对外服务隐藏版本信息，指纹识别准确率降低至≤30%；

d） 管理后台限制 IP 白名单≤20 个段，禁止公网直接访问。

7.2 漏洞管理

漏洞管理应满足以下要求：

a） 自动化漏洞扫描每周 1 次，渗透测试每季度 1 次，代码审计每半年 1 次；

b） 严重漏洞（CVSS 9.0~10.0）≤24h 修复，高危（7.0~8.9）≤7d，中危（4.0~6.9）≤30d，低危（0.1~3.9）≤90d；

c） 修复后复测覆盖率 100%，误报率≤5%，漏报率≤1%；

d） 关键漏 4h 内预警，24h 内排查。

7.3 基础安全设施部署

基础安全设施应满足表 4 要求。

表 4 基础安全设施量化部署指标

设施类型	性能/功能指标	部署位置

设施类型	性能/功能指标	部署位置
下一代防火墙	吞吐量≥10Gbps，并发连接≥10 万，新建连接≥10 万/s	互联网边界、安全域边界
Web 应用防火墙	防护规则库≥5000 条，0Day 规则更新≤4h，误拦截率≤0.1%	Web 服务集群前端
入侵检测/防御系统	检测规则≥10000 条，检测率≥99%，漏报率≤1%，阻断延迟≤100μs	核心交换节点、实训网络出口
全流量审计	流量存储≥90d，会话还原率 100%，支持 PCAP 回溯	关键网络节点
防病毒系统	终端覆盖率 100%，服务器覆盖率 100%，病毒库更新≤4 h	全终端、全服务器
堡垒机	支持 SSH/RDP/VNC，会话录像≥6 个月，命令审计覆盖率 100%	运维管理域入口

8 安全运维与应急响应

8.1 安全日志管理

安全日志管理应满足以下要求：

a） 网络设备、安全设备、服务器、应用系统、数据库日志采集率 100%，未采集告警延迟≤5 min；

b） 统一日志格式（JSON 或 CEF），字段映射准确率 100%，时间戳同步误差≤1s；

c） 支持实时关联分析（窗口≤5 min），异常行为模型≥50 个，误报率≤10%；

d） 热数据（7d）SSD 存储，温数据（90d）SATA 存储，冷数据（3a）对象存储；

e） 权限变更、数据批量导出（>1000 条）、配置修改、登录异常实时告警，延迟≤30s。

8.2 安全事件响应

8.2.1 事件分级与响应时效

安全事件分级与响应时效应满足表 5 要求。

表 5 安全事件分级量化响应指标

事件级别	判定标准	MTTD	MTTR	处置要求
特别重大	平台全面瘫痪、>10 万用户数据泄露、核心模型被窃取、监管通	≤5min	≤15min	启动最高级应急预案，2h 内书面报告教育主管部门和网信部门，24 h

事件级别	判定标准	MTTD	MTTR	处置要求
	报			内提交初步调查报告
重大	核心功能受损>4h、1 万~10 万用户数据泄露、教师/管理员账户大规模失陷	≤15min	≤30min	启动应急预案，4h 内报告，48h 内完成处置并提交报告
较大	非核心功能异常>8 h、<1 万用户数据泄露、单一实训环境被攻破	≤30min	≤2h	专项处置小组响应，24h 内报告，72h 内闭环
一般	单一模块异常、未遂攻击、低危漏洞利用尝试	≤2h	≤4h	常规工单处置，72h 内记录归档

8.2.2 应急响应能力

应急响应能力应满足以下要求：

a） 预案覆盖数据泄露、勒索软件、DDoS 攻击、内部威胁、供应链攻击 5 类场景，可执行率 100%；

b） 桌面演练每季度 1 次，实战演练每半年 1 次，覆盖全部特别重大和重大场景；

c） 演练目标达成率≥90%，改进项关闭率 100%；

d） 应急工具包更新周期≤30d，可用性 100%；

e） 与 CNCERT、公安机关、供应链厂商应急联络响应时间≤30 min。

8.3 安全审查机制

安全审查机制应满足表 6 要求。

表 6 审查机制量化要求

审查类型	审查时机	通过标准	审查周期
上线前审查	版本发布前	代码审计：高危漏洞 0 个，中危≤5 个；渗透测试：高危风险 0 个；配置基线：合规率≥95%；开源组件：无高风险许可证冲突	每次发布
周期性评估	每季度	漏洞扫描：新增高危 0 个；基线核查：合规率≥95%；权限复核：闲置权限清理率 100%；日志审计：覆盖率 100%	每季度
等保测评	每年	符合 GB/T 22239 第三级要求，测评分数≥80 分，无高危风险项	每年

审查类型	审查时机	通过标准	审查周期
第三方测评	每两年	聘请具备 CNAS 资质的机构全方位测评，无重大风险	每两年
变更审查	重大架构调整、核心算法更新、数据流转变更时	风险可控，方案可执行	按需

附录 A 安全可信建设检查清单

（规范性）

安全可信建设检查清单

章节	检查项	检查方法	合格标准	检查频率
5.2	静态加密覆盖率	存储扫描	核心级+敏感级 100%	每半年
5.3	备份 RTO/RPO 达标率	演练验证	100%达标	每季度
6.1	容器隔离有效性	渗透测试	逃逸测试失败	每半年
6.2	实训网络外联阻断	流量分析	未授权外联 0 次	每月
7.3	漏洞修复时效	工单审计	严重 24h，高危 7d	每月
8.2	应急演练成功率	演练评估	≥90%	每半年

附录 B 安全可信建设检查清单
（资料性）

英文索引

中文术语	英文对照
安全可信框架	secure and trustworthy framework
数据分类分级	data classification and grading
实训容器	practical training container
最小权限原则	principle of least privilege；PoLP
恢复时间目标	recovery time objective；RTO
恢复点目标	recovery point objective；RPO
平均检测时间	mean time to detect；MTTD
平均响应时间	mean time to respond；MTTR
多因素认证	multi-factor authentication；MFA
零信任	zero trust