ICS 35.240.50
CCS D09

KA

中华人民共和国矿山安全行业标准

KA/T XXXXX—XXXX

矿山安全大模型通用技术要求

General technical requirements for mine safety large-scale models

（征求意见稿）

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

XXXX－XX－XX 发布　　XXXX－XX－XX 实施

国家矿山安全监察局　　发布

目　次

前言 II
1　范围 0
2　规范性引用文件 0
3　术语和定义 0
4　基本要求 0
5　能力要求 1
5.1　矿山行业知识理解与解读能力 1
5.2　矿山专业文档与合规内容生成能力 1
5.3　矿山安全风险推理与预警能力 1
5.4　矿山安全生产场景适配与应用能力 1
5.5　矿山安全生产智能辅助决策能力 1
6　安全要求 2
6.1　模型安全 2
6.2　内容安全 2
6.3　数据安全 3
6.4　服务安全 5
7　验证要求 5
7.1　测试题库建设 5
7.2　验证评估 6
7.3　模型能力验证 6
7.4　模型安全验证 6
7.5　内容安全验证 6
7.6　数据安全验证 7
7.7　服务安全验证 7

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。

本文件由国家矿山安全监察局提出。

本文件由矿山安全行业标准化技术委员会信息与智能化分技术委员会归口。

本文件起草单位：煤炭科学研究总院有限公司、中煤科工集团常州研究院有限公司、天地（常州）自动化股份有限公司、中国中煤能源集团有限公司、中煤科工西安研究院（集团）有限公司、应急管理部信息研究院、中国矿业大学、华为技术有限公司、深圳市中金岭南有色金属股份有限公司、长沙矿山研究院有限责任公司、中国恩菲工程技术有限公司。

本文件主要起草人：王海军、程健、骆意、杨培培、代芳琳、贺耀宜、赵奕、陈思宇、赵艳伟、毛开江、贾琨、雷超、韩安、王雅琨、雷晓荣、汪雪君、程前、赵小虎、杨嘉如、刘晓明、李龙、李贵红、陈长远、尹天明。

本文件为首次发布。

矿山安全大模型通用技术要求

1 范围

本文件规定了矿山安全大模型的基本要求、能力要求、安全要求，描述了模型的验证要求。

本文件适用于煤矿、金属非金属矿山等领域的矿山安全大模型提供方、需求方和独立评价方开展矿山安全大模型的开发建设、验证评估和产品选型等工作。矿山安全监管监察部门开展矿山安全大模型应用、监督管理、技术评价和辅助决策等工作时，可参考使用本文件。

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 信息安全技术 网络安全等级保护基本要求

GB/T 35273 信息安全技术 个人信息安全规范

GB/T 45288.1 人工智能 大模型 第1部分：通用要求

GB/T 45288.2 人工智能 大模型 第2部分：评测指标与方法

GB/T 45654 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

GB/T 45288.1界定的以及下列术语和定义适用于本文件。

3.1

矿山安全大模型 mine safety large-scale model

面向矿山安全，以矿山安全归集数据为训练与应用底座，融合矿山安全领域知识、业务规则、监测数据与应急处置逻辑的专业化基座模型。

3.2

矿山大模型安全 mine large-scale model safety

针对矿山安全大模型训练、部署、运行、退出全生命周期，以保障矿山安全生产为核心目标，覆盖模型安全、内容安全、数据安全、服务安全四大核心维度，融合矿山安全法规标准、行业业务规则与人工智能安全防护要求，涵盖安全防护、风险管控、合规管理与应急处置的专业化安全保障体系及相关技术与管理要求的总和。

3.3

归集数据 full-domain aggregated data

按应用场景对矿山安全领域内相关区域、层级、业务、模态的监测、管理、执法、应急等数据，进行统一采集、汇聚整合、规范治理、安全存储后形成的标准化、可共享、可复用的数据集合。

3.4

多模态融合 multi-modal fusion

对来自不同模态的数据（如文本、图像、传感器、音频、视频、三维地理空间数据等）进行联合表示与处理的过程。

4 基本要求

4.1 矿山安全大模型（以下简称“模型”）应符合 GB/T 45288.1 的要求。

4.2 模型应遵循矿山大模型安全的安全优先、依法合规、可控可信、公平适用的原则。

4.3 模型应支持配套构建多源知识库，知识库内容应包含事故案例、安全规程、地质信息、监测数据，宜涵盖企业私有安全知识库；知识库应具备在线增量更新能力，其中安全法规、规程标准等正式发布

文件应在其实施之日起30日内完成知识库更新，生产事故案例、现场隐患数据、地质与监测动态数据等业务信息宜实时在线增量更新。

4.4 模型应建立全生命周期迭代管理机制，明确模型参数与核心能力的迭代要求：定期优化迭代周期不宜超过6个月；当矿山安全监管规则、核心业务场景发生重大变更时，应及时启动专项迭代，迭代完成后需经安全符合性与业务性能验证，合格后方可上线运行。

4.5 模型应具备与矿山数据基础设施高效交互能力，可实现数据的快速检索、调用及协同处理。

4.6 模型应具备满足业务需求的实时处理能力，如矿山安全实时监测与应急响应场景。

4.7 模型识别结果应满足矿山生产与安全管理应用场景的技术要求。

4.8 模型应具备多模态融合推理能力，应至少支持两类及以上模态的融合推理、关联分析与联合决策，模态包括文本、图像、传感器、音频、视频、三维地理空间数据。

4.9 模型宜优先基于新增与存量数据开展增量训练、微调优化。

4.10 模型宜支持验证，验证工作可由系统建设运营单位组织实施，也可委托具备相应检测、评估或认证能力的第三方机构实施。

5 能力要求

5.1 矿山行业知识理解与解读能力

5.1.1 模型应具备矿山安全规程、企业安全管理制度等文件的适用场景解析能力，以及地质数据核心参数、设备运行状态、人员行为、环境监测数据的异常识别能力；模型相关输出和建议参考GB 16423等适用标准。

5.1.2 模型应具备多源数据关联分析能力，能够识别通风量、顶板位移与地质构造等关联关系。

5.1.3 模型应支持基于矿山行业多类型数据形成专属知识库，深度解析归集数据，形成可复用、可推理的行业知识资产，并支持外挂知识库的接入与调用，支撑信息分析、内容提取概括、安全判别等应用。

5.2 矿山专业文档与合规内容生成能力

5.2.1 模型应支持基于矿山法律法规、行业标准、作业规程、应急预案、专利文献、事故报告、监测数据等资源，围绕采煤、掘进、机电、运输、通风等专业技术维度，以及安全管理等综合管理维度构建知识库，并融合矿山安全风险库、隐患库等专属知识资源。

5.2.2 模型应支持生成覆盖采煤、掘进、机电、运输、通风等专业技术维度，以及安全管理等综合管理维度的设备维护方案、隐患整改方案、应急流程示意图等行业专属知识，且生成内容应可对接矿山安全风险库、隐患库开展关联应用。

5.2.3 生成内容涉及标准规程条款解读、阈值数据引用等事实性输出时，应支持明确标注数据来源、依据条款及适用范围，保障可追溯性；对于综合研判、报告撰写等非确定性生成内容，宜建立内容审核机制（见6.2.3）。

5.3 矿山安全风险推理与预警能力

5.3.1 模型应具备逻辑推理与模式识别能力，可基于矿山行业数据开展深度分析与辅助指导。

5.3.2 模型应支持推理过程及决策依据的可说明性，完整展示结论对应的支撑数据、研判依据与分析思路；宜提供可解释性报告或可视化分析结果。

5.3.3 模型应支持文本规程、图像监控、传感器数据等多模态融合推理，具备基于事故案例与实时监测数据预测潜在安全风险的能力。

5.4 矿山安全生产场景适配与应用能力

5.4.1 模型应支持将行业知识有效应用于安全生产、安全监控、隐患排查、应急处置、安全培训等实际场景。

5.4.2 模型典型应用场景宜支持涵盖瓦斯监测预警、地压监测预警、设备故障风险预警、隐患智能识别、人员不安全行为识别、应急救援辅助决策等核心安全业务。

5.5 矿山安全生产智能辅助决策能力

5.5.1 模型应支持基于大数据分析提供科学决策支持，辅助矿山企业管理层开展科学决策，实现资源配置优化。
5.5.2 模型决策应严格遵循矿山安全优先原则，当生产效率与安全要求存在冲突时，决策建议应符合适用的法律法规、规章和强制性标准要求。

6 安全要求

6.1 模型安全

6.1.1 训练安全

6.1.1.1 模型训练应支持分布式并行计算架构，具备数据并行、模型并行等分布式训练能力。
6.1.1.2 训练数据应经过安全校验、合规审核、质量清洗，符合 GB/T 45654 的规定，应具备防数据投毒、防后门攻击能力，采用数据溯源与校验机制，确保训练数据来源合规、真实可靠。
6.1.1.3 训练过程应支持加密存储，训练参数、中间结果等敏感数据严禁泄露，训练环境应进行隔离防护，防止未授权访问。
6.1.1.4 训练过程应具备实时监控能力，监测算力状态、数据质量及训练过程异常，并能根据异常等级自动触发告警、训练暂停或训练中断等处置动作。对可能危及矿山安全生产的高风险异常，应自动触发训练中断。

6.1.2 部署安全

6.1.2.1 模型应支持身份认证与访问控制，支持角色分级授权，未授权用户严禁访问、修改模型参数及部署配置。
6.1.2.2 模型部署文件应支持完整性校验，防止模型文件被篡改，校验失败时应拒绝部署并触发告警。
6.1.2.3 模型应支持建立完善的审计日志机制，对部署全过程进行记录。审计日志应防篡改、可追溯，支持与安全审计接口（见 6.4.4）对接，日志保存期限不应低于 2 年。
6.1.2.4 模型宜支持部署边缘节点并配置边缘算力，具备离线独立运行能力。

6.1.3 运行安全

6.1.3.1 应支持运行状态实时监测，包括算力占用、内存使用率、推理时延、准确率等，异常自动告警。
6.1.3.2 宜具备故障自愈、主备切换能力，对于未造成业务中断或数据丢失的轻微故障，应能自动完成恢复；对于导致核心服务中断、数据丢失的严重故障，应能通过主备切换机制恢复业务。
6.1.3.3 应抵御对抗样本攻击、模型窃取攻击、数据投毒攻击，保障模型稳定运行。
6.1.3.4 模型迭代更新前应开展全面的更新影响分析，评估模型更新对矿山安全生产业务、模型运行性能、安全防护能力及数据合规性的影响，排查更新风险；模型更新完成后必须开展合规性审查，核查更新数据、算法逻辑、输出内容及安全机制的合规性，验证无安全风险、业务适配正常、符合矿山安全法规标准及本文件要求后方可上线运行，更新全过程资料应完整留存、可追溯。

6.1.4 退出安全

6.1.4.1 应支持彻底销毁训练数据、模型参数、运行日志等敏感数据，数据应无法恢复，销毁方式应符合 GB/T 22239 的销毁要求。
6.1.4.2 应支持退出过程全程记录、留痕可查，形成退出报告，包括退出时间、数据销毁清单、设备处置情况等，报告保存期限不低于 3 年。

6.2 内容安全

6.2.1 输出准确性

6.2.1.1 模型输出结果应支持保存完整推理过程，具备输出结果回溯查验能力。
6.2.1.2 涉及应急处置建议、避灾路线规划等关键安全决策的模型输出，仅作为辅助参考，须经授权安全管理人员确认后方可执行，模型不应自主触发执行任何关键安全操作。



6.2.1.3 涉及安全规程解读、避灾路线规划、应急处置建议等关键决策类输出，应严格遵循有据可依原则，依托专属可信知识库生成内容；引用规程条款、具体数值时应标注原文内容及条款编号，开展多源数据综合分析时应说明信息来源范围并展示核心特征数据；应确保输出内容真实可溯源，无事实性幻觉、无编造信息；无法匹配可靠依据的安全决策类需求，应明确提示依据不足，不应输出确定性结论。

6.2.1.4 模型宜支持对所有生成内容统一提供溯源支撑，明确标注所引用知识库原文片段、外部数据来源等溯源信息；无有效可信支撑依据时，应拒绝生成确定性输出内容，或明确提示“无可靠依据，仅供参考”。

6.2.2 合规性评估

6.2.2.1 模型输出内容不应包含违法违规内容，不应违背矿山安全管理规范。

6.2.2.2 输出的应急处置建议应符合矿山应急救援相关标准，具备科学性、可操作性，严禁输出误导性、危害性建议。

6.2.3 内容审核机制

6.2.3.1 模型应支持建立定期与实时相结合的内容审核机制，高风险业务输出开展实时抽检，一般业务输出开展定期抽检。

6.2.3.2 模型应支持对错误输出、违规内容和疑似幻觉信息进行标记、拦截、人工复核、纠正和闭环处置，并将相关样本纳入模型迭代优化流程，错误信息保存期限不低于 1 年。

6.2.3.3 模型应建立用户反馈纠错通道，快速响应内容安全问题。

6.3 数据安全

6.3.1 数据安全全生命周期管控

6.3.1.1 数据采集安全

采集设备应具备安全防护能力，防止采集数据被篡改、窃取；采集过程应支持加密传输方式，所用加密技术及产品应符合国家密码管理部门认证核准要求。

6.3.1.2 数据存储安全

数据储存应符合以下要求：

a) 矿山归集数据应支持加密存储，敏感数据采用加密算法，存储设备应具备防篡改、防丢失能力。

b) 数据存储应支持备份，采用本地备份与异地备份相结合的方式，备份频率每日至少 1 次，备份数据保留期限不低于 1 年，备份数据恢复成功率应符合强制标准及辖管地法律法规要求。

6.3.1.3 数据处理安全

数据处理应采用安全的处理环境，防止数据在处理过程中泄露、篡改，处理完成后的数据应支持完整性校验。

6.3.1.4 数据销毁安全

数据销毁应符合以下要求：

a) 模型不继续使用的已读取及产生的数据，应支持采用符合信息安全要求的销毁方式（如物理销毁、多次覆写），保证数据无法恢复。

b) 数据销毁应支持全程记录，形成销毁报告，报告包含销毁数据清单、销毁方式、销毁时间、经办人等信息，报告保存期限不低于 3 年。

6.3.2 敏感数据隐私保护与脱敏技术

敏感数据应符合以下要求：

a) 个人信息合规保护：涉及从业人员个人信息的数据采集，应在数据采集前，以书面、电子等可留存、可追溯且清晰易懂的形式，获得从业人员的单独明示同意授权；明确告知采集目的、

范围、使用方式、存储期限、数据安全保障措施，符合 GB/T 35273 的规定；对含人员身份信息、人脸信息、工牌信息、定位信息的各类模态数据，应采取人脸模糊、工牌遮挡等隐私保护措施，严禁非法采集、泄露、滥用个人信息；

b) 全场景数据脱敏处理：应对文本、三维地理空间等数据中的矿山核心敏感信息，包括关键巷道坐标、炸药存储量与存储位置、商业秘密、涉密地理信息等，支持自动化脱敏处理，可采用坐标偏移、模糊化、匿名化等适配方式，宜支持自定义脱敏规则适配不同应用场景；脱敏处理后的数据应无法关联到具体主体及精准敏感信息；

c) 分级权限管控：应对矿山敏感数据、关键设施图像、三维地理空间数据等，依据岗位职责实施分级访问权限管控，一线作业人员仅限查询所属区域简化数据，管理岗位按授权获取对应数据，严禁非授权人员查看、下载、篡改敏感数据；

d) 全流程合规管控：隐私保护与数据脱敏应贯穿数据全生命周期，相关处理过程应全程留痕，确保敏感数据与隐私信息全流程可控，防止涉密及隐私信息泄露。

6.3.3 典型多模态数据安全管控

6.3.3.1 文本模态数据安全

文本模态数据应符合以下要求：

a) 数据来源合规：模型应用的矿山安全规程、事故案例、隐患报告等文本数据，应确保来源合法有效，优先采用企业备案文档、官方标准及权威机构发布资料，严禁使用盗版或涉密文本；

b) 内容一致性校验：生成的安全指引、整改建议等成果文件，应与原始数据源核心条款一致，必要时应支持关联引用原文片段。

6.3.3.2 图像模态数据安全

图像模态数据应符合以下要求：

a) 图像真实性保障：矿山监控图像、设备巡检照片投入应用前，应支持通过哈希校验、数字水印等技术验证完整性，防止数据篡改；

b) 场景适配验证：图像识别结果（如设备故障判定、隐患等级划分等）宜支持结合井下低光照、高粉尘等实际作业场景进行校验，建立专属识别阈值，结果准确性应符合实际矿山生产应用要求。

6.3.3.3 传感器时序数据安全

传感器时序数据应符合以下要求：

a) 采集可靠性：对瓦斯浓度、顶板压力、电流、电压、温度等传感器数据，宜建立数据来源可信保障机制，实时监测数据传输丢包、异常值，应支持自动补传及告警功能；

b) 实时性保障：冲击地压监测等关键安全指标数据，宜采用低延迟传输协议，确保数据与现场实际状态的时差控制在安全阈值范围内，防止预警失效；

c) 阈值联动校验：模型生成的预警信息应支持同时匹配行业标准阈值及企业分级标准，结合监管预警平台历史数据优化阈值，保障预警信息的合规性与精准性。

6.3.3.4 音频模态数据安全

音频模态数据应符合以下要求：

a) 音频完整性保护：井下通讯、设备异响等音频数据，应支持完整性校验，防止因片段截取导致语义失真；应采用加密存储方式，防范数据非法窃取；

b) 语义识别准确：对紧急指令、故障报警等关键音频信息的识别，应支持结合语境校验，避免将日常交流误判为安全事件；

c) 噪声干扰适配：应支持部署噪声过滤模块，高噪声环境（如风机、破碎机运行场景）下音频关键信息提取与识别准确性应满足业务要求。

6.3.3.5 视频模态数据安全

视频模态数据应符合以下要求：

a) 采集可靠性：对井下工作面、变电所、炸药库、提升机房、主运输巷等关键区域的视频设备，应适配低光照、高粉尘、强电磁干扰井下工况环境，实现视频画面稳定、清晰采集。
b) 数据源可信校验：宜建立视频数据来源可信与完整性校验机制，对视频数据流来源合法性、数据完整性进行常态化校验。

6.3.3.6 三维地理空间数据安全

三维地理空间数据应符合以下要求：
a) 精度与保密：三维地理空间数据应符合相关精度标准，严格遵守地理信息保密规定；对涉密敏感的数据，应支持坐标偏移处理，防止涉密信息泄露；
b) 关联验证：基于三维数据生成的工程规划、隐患定位结果，应支持与实际地质条件、工程布局关联验证，规避安全风险；
c) 多模态时空同步：应支持建立“空间位置+传感器监测+视频监控”的时空同步机制，确保同一时刻不同模态数据的时空基准统一。

6.4 服务安全

6.4.1 接口通用安全

6.4.1.1 所有应用程序接口均应支持采用超文本传输安全协议进行数据传输，加密算法采用传输层密码协议 1.2 及以上版本。
6.4.1.2 应支持严格的身份认证与鉴权，支持密钥认证、令牌认证等多种认证方式，未通过认证的请求应拒绝响应，认证失败处理应符合安全规范。
6.4.1.3 应具备防过载、防攻击能力，支持按应用场景、用户角色、接口类型配置请求频率限制和限流阈值，对异常请求（如恶意请求、超大流量请求）进行拦截与告警。

6.4.2 系统应用接口

6.4.2.1 面向第三方应用接口，宜支持采用表征状态转移风格架构设计，接口文档规范完整。
6.4.2.2 应支持输出结果的数字签名，以确保输出内容未被篡改，签名验证失败时，第三方应用应拒绝接收数据。

6.4.3 数据源接口

6.4.3.1 数据可由矿山业务系统、数据平台、知识库或经授权的第三方应用提供，并覆盖多种场景，接入通用平台等数据源的接口，应支持数据校验功能，对无效数据、异常数据进行过滤与告警。
6.4.3.2 批量数据上传接口应支持断点续传功能，以确保数据传输的完整性，传输失败时应支持自动重试，重试次数不超过 3 次。

6.4.4 安全审计接口

6.4.4.1 应支持提供安全审计接口，支持系统推送模型训练、运行、调用、输出等全生命周期的审计日志。
6.4.4.2 审计日志应支持包含操作时间、操作主体、操作内容、操作结果等关键信息，日志格式规范。

6.4.5 软件通用安全

6.4.5.1 应支持采用安全的开发框架与编程语言。
6.4.5.2 应支持访问控制、权限管理功能，支持角色分级授权，不同角色仅能访问其权限范围内的功能与数据。
6.4.5.3 应支持安全审计功能，对所有操作行为进行记录，审计日志保留期限不低于 1 年，支持日志查询、统计与分析。

7 验证要求

7.1 测试题库建设

7.1.1 测试题库应支持围绕模型核心应用场景与安全风险，按“通用基础+场景专项”分级分类构建，覆盖数据安全、内容安全、模型安全、应用安全维度；测评采用 GB/T 45288.2 规定的理解能力、生成能力双主线评测框架，测试题库对应的数据集应符合 GB/T 45288.2 中关于时效性、可用性、完备性等维度的规定。
7.1.2 生成内容测试题库应支持按共性场景、专项场景分类构建。
7.1.3 拒答测试题库应支持按模型应拒答的问题、模型不应拒答的问题分类构建应拒答测试题库、非拒答测试题库。
7.1.4 针对井下特殊环境与高风险场景，专项测试题库应支持按井下边缘计算场景、多模态融合场景、应急响应场景分类构建。

7.2 验证评估

7.2.1 自评估测试责任主体为系统建设运营单位，具体测试工作可由责任主体自有具备相应技术能力的团队实施，或委托具备相应检测、评估或认证能力的第三方机构实施。
7.2.2 评估实施前应明确模型的应用场景，依据场景建立生成内容测试题库、安全拒答测试题库及专项测试题库。
7.2.3 评估活动应遵守国家法律法规及标准规范，结合应用场景制定自评估方案与流程，完成测试并出具自评估报告；自评估报告应包含矿山大模型安全保障体系的有效性评价内容，并由责任主体归档保存；需备案或提交的，按主管部门、委托方或相关管理要求执行。

7.3 模型能力验证

7.3.1 矿山行业知识理解与解读能力验证：模型的矿山规程制度解析、矿山各类数据异常识别、多源数据关联分析、矿山专属知识库形成及外挂知识库调用功能有效。
7.3.2 矿山专业文档与合规内容生成能力验证：模型的矿山专业知识库构建、矿山行业专属内容生成、生成内容来源标注及内容审核相关功能有效。
7.3.3 矿山安全风险推理与预警能力验证：模型的矿山业务逻辑推理模式识别、推理决策过程可说明、多模态融合风险推理预测能力符合要求。
7.3.4 矿山安全生产场景适配与应用能力验证：模型的矿山各类安全生产实际场景适配、核心安全业务场景覆盖能力符合要求。
7.3.5 矿山安全生产智能辅助决策能力验证：模型的矿山安全辅助决策支撑、安全优先决策原则遵从能力符合要求。

7.4 模型安全验证

7.4.1 数据投毒防御验证：应支持构建包含正常矿山数据与典型数据投毒样本的测试集，投毒样本应覆盖后门攻击、样本污染等典型类型，可采用 BadNets 等基准攻击方法生成。训练数据的采集、清洗、标注和使用应符合 GB/T 45654 的规定，建立数据来源追溯、训练数据完整性校验及异常样本检测能力。数据投毒防御有效性，覆盖不同投毒比例与强度场景，明确通过/失败判定准则。将测试集输入模型进行验证，检验模型对投毒数据的识别与过滤能力，防御成功率及性能衰减指标应符合相关管理办法。
7.4.2 对抗性攻击防御验证：采用常见对抗攻击算法构造测试数据输入运行中的模型，验证其输出准确性；模型在对抗样本测试集上的鲁棒准确率相较于干净测试集上的原始准确率，下降幅度不应超过 5 个百分点，同时应能抵御已知对抗攻击，建立持续的安全监测与更新机制以应对包括自适应攻击在内的新型威胁，当检测到模型性能因潜在攻击而异常下降时应支持触发告警。
7.4.3 运行状态监测验证：模拟模型运行异常场景，验证模型告警机制的有效性与响应时间，应符合 6.1.3 的要求。

7.5 内容安全验证

7.5.1 内容生成输出准确性验证：构建矿山安全场景测试数据（涵盖危险源识别、人员安全管控、设备安全监测），输入模型，统计输出结果的准确性，应符合 6.2.1 的要求。
7.5.2 内容审核机制验证：模拟模型输出错误、违规内容场景，验证内容审核机制的过滤、修正能力，应符合 6.2.3 的要求。
7.5.3 应急响应支持功能验证：模拟矿山安全事故场景，验证模型事故分析时间、应急处置方案生成

的科学性与可操作性。

7.6 数据安全验证

7.6.1 危险源识别验证：选取 10 种典型矿山危险源场景，模拟实时监测数据输入模型，验证危险源识别准确性与预警提前量。

7.6.2 数据加密验证：对存储的敏感数据进行解密测试，验证加密算法的有效性；对传输数据进行抓包分析，验证传输加密的安全性。

7.6.3 数据备份与恢复验证：触发数据恢复流程，验证备份数据的完整性与恢复成功率，应符合 6.3.1.2 的要求。

7.7 服务安全验证

7.7.1 接口验证：认证、加密、限流、审计功能有效。

7.7.2 平台验证：漏洞、抗攻击、可靠性、自愈能力符合要求。