

团 体 标 准

T/BIOT XXXX—XXXX

矿山用人工智能算法模型评估方法

Evaluation methods for artificial intelligence algorithm models in mining

(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

目 次

前 言	II
引 言	III
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 评估原则	1
4.1 科学性	1
4.2 可操作性	1
4.3 公正性	2
5 总体要求	2
5.1 基本要求	2
5.2 安全要求	2
6 评价指标	2
6.1 准确性	2
6.2 稳定性	2
6.3 鲁棒性	3
6.4 效率	3
7 评分方式	3
7.1 评分标准	3
7.2 综合评分	4
8 综合评定	4
8.1 评定等级	4
8.2 评定结果	4
9 评价程序	4
9.1 评价流程	4
9.2 实施步骤	5
10 评价报告	5
附 录 A（资料性） 矿山用人工智能算法模型评估案例	6
参 考 文 献	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。

本文件由北京物联网智能技术应用协会提出并归口。

本文件起草单位：XXX。

本文件主要起草人：XXX。

本文件为首次发布。



引 言

随着人工智能技术在矿山领域的广泛应用，矿山智能化建设已成为行业转型升级的重要方向，而矿山生产环境的复杂性和特殊性，对人工智能算法模型提出了更高要求，亟须建立科学、统一的评估方法。当前矿山行业不仅在人工智能算法模型应用方面缺乏规范性指导，导致模型选型、部署和评价存在较大随意性，影响了智能化建设的质量和效率；同时，矿山领域尚未形成针对人工智能算法模型的专门评估标准，难以满足智能化矿山建设对算法模型评估的系统性需求，这也使得模型性能难以横向比较、安全性和可靠性无法得到有效保障，进而制约了人工智能技术在矿山领域的深度应用。

本文件参考 GB/T 45288.2、GB/T 46010 等标准的技术要求，结合矿山生产特点，建立了涵盖准确性、稳定性、鲁棒性和效率等维度的综合评价体系。文件规定了评估原则、指标体系和实施方法，为矿山用人工智能算法模型的选型、测试和验收提供技术依据。

本文件主要面向矿山企业、智能化系统集成商、算法开发单位和第三方检测机构等用户群体。通过实施本文件，用户可对矿山用人工智能算法模型进行客观评价，确保模型满足矿山安全生产和高效运营的要求。

本文件的实施将有效规范矿山人工智能算法模型的市场秩序，提升模型质量和应用效果。预计可降低模型选型成本，提高模型部署成功率，为矿山智能化建设提供可靠的技术支撑。同时，本文件可为监管部门提供技术依据，促进矿山人工智能应用的健康发展。

矿山用人工智能算法模型评估方法

1 范围

本文件确立了矿山用人工智能算法模型，规定了矿山用人工智能算法模型的总体要求、评价指标、评分方式、综合评定、评价程序及评价报告等内容。

本文件适用于矿山生产过程中使用的人工智能算法模型的性能评估，包括但不限于矿山安全监测、设备故障预测、资源优化配置等应用场景。

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 41479 信息安全技术 网络数据处理安全要求

GB/T 45288.2 人工智能 大模型 第2部分：评测指标与方法

GB/T 46010 信息技术 矿山大数据 技术要求

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 artificial intelligence

〈学科〉人工智能系统相关研制和应用的研究和开发。

[来源：GB/T 41867—2022，3.1.2]

3.2

准确性 accuracy

算法模型输出结果与真实值之间的符合程度。

3.3

稳定性 stability

算法模型在多次运行或不同环境条件下输出结果的一致性。

3.4

鲁棒性 robustness

〈人工智能〉系统在任何情况下都保持其性能水平的特性。

[来源：GB/T 41867—2022，3.4.9]

4 评估原则

4.1 科学性

评价过程遵循科学规律，采用经过验证的理论和方法。评价指标的设计基于矿山行业的实际需求和特点，确保评价结果能够真实反映算法模型的性能。

4.2 可操作性

评价方法在实际应用中便于实施和管理，评价流程清晰明确，评价指标易于理解和计算。评价过程中使用的工具和技术与矿山行业的现有技术条件相匹配，确保评价工作能够顺利开展。

4.3 公正性

评价过程对所有评价对象一视同仁，避免因主观因素导致评价结果出现偏差。评价标准的制定公开透明，评价过程接受监督，评价结论客观反映算法模型的真实性能，不受外部因素干扰。

5 总体要求

5.1 基本要求

- 5.1.1 评估对象应包括但不限于矿山设备故障预测模型、矿石品位识别模型、矿山安全监测模型等。
- 5.1.2 评估过程应建立完整的文档记录体系，包括算法模型的技术文档、测试数据集说明、评估环境配置等。
- 5.1.3 评估测试数据集应涵盖矿山生产过程中的常见工况数据、异常数据及边缘场景数据，数据规模需满足模型评估的统计学要求，并应符合 GB/T 46010 的相关规定。
- 5.1.4 所有参与评估的算法模型应提供完整的开发文档和训练数据来源说明。
- 5.1.5 评估机构应具备相应的技术能力和资质，评估人员应经过专业培训并熟悉矿山行业特点。评估过程中应建立有效的质量控制和监督机制。

5.2 安全要求

- 5.2.1 算法模型在数据处理过程中应符合 GB/T 41479 的数据安全要求，防止数据泄露。
- 5.2.2 算法模型应具备防篡改能力，模型文件应进行完整性校验。模型运行过程中应具备异常检测和处理机制，防止因模型故障导致的安全事故。
- 5.2.3 算法模型在矿山环境中的运行不应影响其他设备的正常工作。模型应具备资源占用监控功能，当资源占用超过阈值时应自动告警或降级运行。模型应定期进行安全评估和漏洞扫描，确保运行过程中的安全性。
- 5.2.4 算法模型的开发者应提供详细的安全说明书，说明模型的安全特性和使用注意事项。
- 5.2.5 模型使用者应按照说明书的要求正确配置和使用模型。对于涉及矿山安全生产的关键模型，应建立专门的安全评估和监控机制。

6 评价指标

6.1 准确性

准确性评价应基于矿山特定场景下根据模型在测试数据集上的预测结果与真实值的偏差程度确定。对于分类任务，可采用混淆矩阵计算精确率、召回率等指标，计算方法可参考 GB/T 45288.2 的规定；对于回归任务，可采用均方误差、平均绝对误差等指标。准确性评价指标分级见表 1。

表 1 准确性评价指标分级

等级	准确率范围	适用场景
一级	≥95%	安全关键系统
二级	90%~95%	生产调度系统
三级	85%~90%	辅助决策系统
四级	<85%	实验性系统

6.2 稳定性

6.2.1 稳定性评价应包括时间维度测试和压力测试。时间维度测试应连续监测模型 30 d 以上的性能指标波动率，压力测试则应模拟极端工况下的数据输入。

6.2.2 稳定性评价应采用变异系数 (CV) 作为主要量化指标，见公式 (1)。

$$CV = \frac{\sigma}{\mu} \times 100\% \quad \dots\dots\dots (1)$$

式中：

σ——标准差；

μ——均值。

6.2.3 矿山用算法模型的 CV 值应控制在 5%以内。对于涉及人员安全的模型，CV 值要求不超过 3%。

6.3 鲁棒性

6.3.1 鲁棒性评分应采用性能衰减率计算，见公式（2）。

$$D = 1 - \frac{P_a}{P_n} \times 100\% \quad \dots\dots\dots (2)$$

式中：

D——鲁棒性评分；

P_n ——正常工况下的性能指标；

P_a ——异常工况下的性能指标。

6.3.2 一级鲁棒性要求衰减率不超过5%，二级不超过10%，三级不超过20%。对于涉及重大安全风险的模型，应达到一级鲁棒性标准。

6.4 效率

计算效率通过单次推理耗时和吞吐量衡量，资源效率则关注内存占用率和 GPU 利用率。在矿山实时控制场景中，算法模型的单次推理耗时应小于 100 ms，批处理吞吐量应达到每秒 1000 条以上。效率评价参数阈值见表 2。

表 2 效率评价参数阈值

参数类型	一级标准	二级标准	三级标准
单次推理耗时	<50 ms	50 ms~100 ms	100 ms~200 ms
内存占用率	<30%	30%~50%	50%~70%
GPU 利用率	>80%	60%~80%	<60%

7 评分方式

7.1 评分标准

7.1.1 准确性

准确性指标的评分标准为：

——误差 <5%得 100 分；

——误差 5%~10%得 80 分；

——误差 10%~15%得 60 分；

——误差 15%~20%得 40 分；

——误差 >20%得 20 分。

7.1.2 稳定性

稳定性指标的评分标准为：

——标准差 <2%得 100 分；

——标准差 2%~5%得 80 分；

——标准差 5%~8%得 60 分；

——标准差 8%~12%得 40 分；

——标准差 >12%得 20 分。

7.1.3 鲁棒性

鲁棒性指标评分标准见表 3。

表 3 鲁棒性指标评分标准

干扰类型	干扰强度	性能衰减率	分值（分）
数据缺失	10%	≤5%	100
	20%	≤10%	80

干扰类型	干扰强度	性能衰减率	分值（分）
噪声干扰	30%	≤15%	60
	10 dB	≤5%	100
	20 dB	≤10%	80
	30 dB	≤15%	60

7.1.4 效率

效率指标的评分标准为：

- CPU 利用率 < 50% 且响应时间 < 1 s 得 100 分；
- CPU 利用率 50%~70% 或响应时间 1 s~2 s 得 80 分；
- CPU 利用率 70%~90% 或响应时间 2 s~3 s 得 60 分；
- CPU 利用率 > 90% 或响应时间 > 3 s 得 40 分。

7.2 综合评分

7.2.1 权重分配应经专家评审确认，并在评估报告中明确说明。各指标权重占比宜按如下分配：准确性占 40%，稳定性占 25%，鲁棒性占 20%，效率占 15%。

7.2.2 矿山用人工智能算法模型的评分方法应采用各级指标加权求和的方式计算得出综合评分，见公式（3）。

$$X = A \times 0.4 + B \times 0.25 + C \times 0.2 + D \times 0.15 \quad \dots\dots\dots (3)$$

式中：

- X——综合评分；
- A——准确性得分；
- B——稳定性得分；
- C——鲁棒性得分；
- D——效率得分；

8 综合评定

8.1 评定等级

矿山用人工智能算法模型的综合评定等级分为五个级别，具体划分见表 4。

表 4 矿山用人工智能算法模型评定等级划分

等级	综合得分（分）	描述
A 级	90~100	优秀，完全满足高标准需求
B 级	80~89	良好，个别指标存在轻微不足
C 级	70~79	合格，某些关键指标需改进
D 级	60~69	基本合格，多个指标存在明显缺陷
E 级	0~59	不合格，无法满足基本要求

8.2 评定结果

评定结果应包含算法模型的综合得分、评定等级以及各评价指标的详细得分。评定结果的生成应基于评价指标的加权计算。评定结果的呈现应清晰明了，便于使用者快速理解算法模型的整体性能。

9 评价程序

9.1 评价流程

9.1.1 评价程序的具体流程应包括：

- 准备阶段：应明确评价目标、组建评价团队、制定评价方案；
- 实施阶段：包括数据采集、模型测试、指标计算等核心工作；
- 报告阶段：应对评价结果进行综合分析并形成正式报告。

- 9.1.2 流程设计应考虑评价对象的技术特性、应用场景以及评价资源的可获得性。
- 9.1.3 评价流程的每个环节都应建立明确的质量控制点，对关键步骤进行复核验证。
- 9.1.4 评价程序主要阶段及工作内容见表 5。

表 5 评价程序主要阶段及工作内容

阶段	主要工作	输出成果
准备阶段	确定评价目标、组建评价团队、制定评价方案	评价方案文档
实施阶段	数据采集、模型测试、指标计算	测试数据、指标结果
报告阶段	结果分析、报告编制、评审确认	评价报告

9.2 实施步骤

- 9.2.1 评价实施步骤应严格按照评价方案执行，确保评价过程的规范性和一致性。实施步骤主要包括：
- 数据准备：收集矿山生产历史数据和实时数据，数据量应满足统计显著性要求；
 - 模型部署：按照生产环境规格配置计算资源，确保模型运行环境的一致性；
 - 测试执行：按照测试方案开展功能测试、性能测试和异常场景测试；
 - 结果记录：详细记录测试过程中的各项指标数据和异常情况；
 - 初步分析：对测试数据进行统计分析，识别模型的性能特征和潜在问题。
- 9.2.2 对于复杂场景下的测试结果，应进行多维度交叉验证，排除偶然因素的影响。
- 9.2.3 评价实施过程中测试数据应进行脱敏处理，防止敏感信息泄露；模型参数和结构信息应严格保密。
- 9.2.4 评价团队应建立完善的数据管理制度。对于评价过程中发现的技术问题，应及时记录并分析原因。矿山用人工智能算法模型评估案例见附录 A。

10 评价报告

评价报告应包含但不限于以下内容：

- 评价对象的基本信息：应详细记录算法模型的名称、版本、开发单位、应用场景等关键信息。应说明算法模型的主要功能和技术特点，以及与矿山生产环境的适配性。
- 评价依据：应明确列出评价所采用的标准规范和技术要求，包括本文件规定的评价指标体系、评分方法以及相关行业标准。
- 评价过程：应详细记录评价的实施步骤、测试环境、数据来源以及评价方法。
- 评价结果及其计算方法。
- 结论建议：应基于评价结果给出明确的结论，并对算法模型的改进和优化提出建设性意见。

附录 A (资料性)

矿山用人工智能算法模型评估案例

A.1 矿山运输调度模型评估案例

本案例展示了某大型露天煤矿运输调度 AI 模型的评估过程。该模型对矿区 50 台自卸卡车、12 台挖掘机进行实时路径优化，评估周期为连续 30 个工作日。模型采用强化学习算法，通过动态调整运输路线实现燃油消耗降低 15.7%，设备利用率提升 22.3%。

评估过程严格遵循本文件第 6 章规定的指标体系，其中准确性指标通过运输任务完成率（98.4%）和计划偏离度（2.1%）双重验证；稳定性指标响应时间标准差为 0.87 s。

运输调度模型性能评估参数见表 A.1。

表 A.1 运输调度模型性能评估参数

评估指标	测量值	单位	达标阈值
路径优化准确率	96.5	%	≥90%
系统响应时间	3.2	秒	≤5 秒
燃油节约率	15.7	%	≥10%
设备利用率	78.3	%	≥70%
故障预警准确率	89.2		≥85%

A.2 矿井安全监测模型评估案例

该案例针对某金属矿山的巷道安全监测 AI 系统进行评估。系统部署于地下 800 m 作业面，集成 12 类传感器数据，实现顶板压力、瓦斯浓度等 16 项参数的实时预测。评估采用本文件第 9 章规定的程序流程，包括 72 h 连续压力测试和 3 次人工异常注入测试。

模型在鲁棒性测试中表现突出，当输入数据缺失 30% 时仍能保持 83.4% 的预测准确率。效率指标方面，单次推理耗时仅 46 ms，满足井下实时监测要求。评估发现的主要改进点是异常事件误报率（7.2%）略高于行业平均水平，安全监测模型测试结果见表 A.2。

表 A.2 安全监测模型测试结果

测试项目	正常工况	数据缺失 30%	噪声干扰 20dB
顶板位移预测误差 (mm)	3.2	5.7	4.9
瓦斯浓度报警准确率 (%)	97.1	89.3	92.8
系统响应延迟 (ms)	46	52	49
CPU 峰值占用率 (%)	68	73	71

A.3 矿石分选模型评估案例

本案例评估某铁矿应用的 X 射线智能分选模型。系统处理量达 300 吨/h，采用深度学习算法实现矿石品位在线识别。评估参照本文件第 8 章综合评定方法，对连续处理的 5000 吨原矿进行抽样检测，模型在三个关键指标上表现如下：

- a) 品位识别准确率达到 94.7%，超过人工分选基准（88.3%）6.4 个百分点；
- b) 废石剔除率提升至 91.2%，同时有效矿石误弃率控制在 2.1%；
- c) 设备能耗较传统分选方式降低 23.6%。

该案例特别展示了模型在矿石表面附着泥土（覆盖率约 15%~20%）工况下的稳定表现，其识别准确率波动范围控制在±1.2%以内。

参 考 文 献

- [1] GB/T 33936—2017 矿井提升机 回收评估规范
- [2] GB/T 41867—2022 信息技术 人工智能 术语
- [3] [3] GB/T 42249—2022 矿产资源综合利用技术指标及其计算方法

BIOT