

KSSJ/YY24-2023

# 智能化矿山数据融合共享 AI 大规模预训练模型技术要求

Intelligent mine data fusion and sharing

Technical requirements for AI large-scale pre-training models

国家矿山安全监察局  
2023 年 6 月



## 目 次

前言.....	III
1. 范围.....	1
2. 规范性引用文件.....	1
3. 术语和定义.....	1
3.1. 大模型 Foundation Models.....	1
3.2. 人工智能 artificial intelligence.....	1
3.3. 人工智能系统 artificial intelligence system.....	1
3.4. 模型训练 model training.....	2
3.5. 推理 reasoning.....	2
3.6. 训练数据 training data.....	2
3.7. 自然语言处理 natural language processing;NLP.....	2
4. 缩略语.....	2
5. 大模型功能要求.....	3
5.1. 大模型通用要求.....	3
5.2. 矿山大模型功能要求.....	3
5.3. 算法兼容性要求.....	4
5.4. 兼容算法框架要求.....	4
6. 大模型接口技术要求.....	5
6.1. 矿山大模型与上层应用之间的接口.....	5
6.2. 矿山大模型与数据源之间的接口.....	5
6.3. 矿山大模型与人工智能系统使能平台之间的接口.....	5
7. 大模型对数据的技术要求.....	5
7.1. L0 通用预训练大模型数据要求.....	5
7.2. L1 行业预训练大模型数据要求.....	6
7.3. L2 部署模型数据要求.....	6
8. 大模型对硬件平台的要求.....	7
8.1. 计算能力.....	7

8.2. 软件能力.....	8
9. 大模型的部署规范.....	8
9.1. 大模型部署架构.....	8
9.2. 中心云部署要求.....	9
9.3. 边缘节点部署要求.....	10
附 录 A（资料性附录） 矿山大模型应用场景.....	11
附 录 B（资料性附录） 人工智能系统功能参考架构.....	14
参考文献.....	17

# 前 言

本文件参照 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这些专利的责任。

本文件起草单位：中国中煤能源集团有限公司、中煤信息技术（北京）有限公司、应急管理部信息研究院、山东能源集团有限公司、华为技术有限公司、国家电投集团科学技术研究院、国家电投集团内蒙古能源有限公司、中国煤矿机械装备有限责任公司、国家能源投资集团有限责任公司、中国华电集团有限公司、陕西煤业化工集团有限责任公司、中煤集团山西有限公司、中煤电气有限公司、国能神东煤炭集团有限责任公司、云鼎科技股份有限公司、陕煤集团神木张家峁矿业有限公司、国能数智科技开发（北京）有限公司、华电煤业集团有限公司、华电煤业集团数智技术有限公司、中国矿业大学、中国矿业大学（北京）、西安工程大学、中国工业互联网研究院、北京龙田华远科技有限公司、上海商汤智能科技有限公司、科大讯飞技术股份有限公司、杭州海康威视数字技术股份有限公司、新华三技术有限公司、阿里巴巴（中国）有限公司、精英数智科技股份有限公司、天津华宁电子有限公司、华洋通信科技股份有限公司、昆仑数智科技有限责任公司。

本文件技术指导：马世志、王立才、张忠温、王海春、胡而已、丁震、赵宇波、王致兵。

本文件主要起草人：蔡峰、王陈书略、肖圣龙、王前、黄韶杰、钱建生、张冬阳、付强、徐加利、赵文豪、高洪波、王孝丰、宋子涛、李雪健、董彦强、韩培强、郭星歌、蒋庆友、宋冰清、云涛、徐鹏、柴红强、杜博、张旭锋、刘波、赵金娥、关有利、潘涛、邓文革、郑耀涛、王瑞、徐金陵、黄金、陈帅领、郭一楠、李艳、马文静、李昕、宫韬、李艳民、宋广录、储汉卿、项疆腾、周佳奇、李飞、张海洋、王琳、侯宇辉、张远、顾军、申军军、王斌。



# 智能化矿山数据融合共享 AI 大规模预训练模型技术要求

## 1. 范围

本文件规定了 AI 大规模预训练模型功能要求、接口技术要求、对数据的技术要求、对软硬件平台的要求，以及部署规范。

本文件适用于基于工业互联网架构的矿山大模型人工智能系统的规划设计、工程建设、运营管理和运行维护。

## 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；凡是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 41867-2022 信息技术 人工智能 术语

## 3. 术语和定义

下列术语和定义适用于本文件。

### 3.1. 大模型 Foundation Models

大模型是预训练大模型的简称，指通过在大规模数据上进行预训练后能快速适应一系列下游任务的模型。大模型兼具“大规模”和“预训练”两种属性，在海量通用数据上进行预训练，大幅提升AI的泛化性、通用性、实用性。

### 3.2. 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。

[GB/T 41867-2022, 3.1.2]

### 3.3. 人工智能系统 artificial intelligence system

（学科）人工智能系统相关研制和应用的研究和开发。

针对人类定义的给定目标，产生诸如内容、预测、推荐或决策等输出的一类工程系统。

注1：该工程系统使用人工智能相关的多种技术和方法，开发表征数据、知识、过程等的模型，用于执行任务。

注2：人工智能系统具备不同的自动化级别。

[GB/T 41867-2022, 3.1.8]

#### 3.4. 模型训练 model training

利用训练数据，基于机器学习算法，确定或改进机器学习模型参数的过程。

[GB/T 41867-2022, 3.2.18]

#### 3.5. 推理 reasoning

从给定的前提进行论证并得出结论。

[GB/T 41867-2022, 3.2.30]

#### 3.6. 训练数据 training data

用于训练机器学习模型的输入数据样本子集。

[GB/T 41867-2022, 3.2.35]

#### 3.7. 自然语言处理 natural language processing;NLP

（系统）基于自然语言理解和自然语言生成的信息处理。

[GB/T 41867-2022, 3.3.16]

### 4. 缩略语

下列缩略语适用于本文件。

AI：人工智能（Artificial Intelligence）

IaaS：基础设施即服务（Infrastructure as a Service）

NLP：自然语言处理（Natural Language Processing）

CV：计算机视觉（Computer Vision）

RTSP：实时流传输协议（Real Time Streaming Protocol）



## 5. 大模型功能要求

### 5.1. 大模型通用要求

应包括数据收集和索引、大模型预训练、大模型微调和部署、大模型迭代四方面，符合下列规定：

- a) 应支持图像、文本、语音等多模态数据；
- b) 宜支持自动化基础 10 亿量级超大规模架构设计，可灵活扩展至百亿量级，满足多样化硬件速度需求；
- c) 应支持超大批次训练优化，保持收敛过程稳定；
- d) 应支持多样化训练方式，包括无监督训练与监督训练；
- e) 应支持高效微调与知识提取，能够根据用户的不同需求，快速进行部署模型的高精度生产；
- f) 应具备自动化数据标注、挖掘与样本发现，增量训练功能，为开发者提供长期、稳定的后续服务，使得模型能够在更长的时间内发挥更大的价值。

### 5.2. 矿山大模型功能要求

应支持用户只需提供少量标注数据，即可以自动化地通过 workflow 产生部署小模型，对应称之为 L2 部署模型，可参考附录 10.2.2 图 3。能够实现不同矿山场景数据，任务的高效复用，应符合下列规定：

- a) 应支持海量无标注数据构建行业预训练模型，通过行业数据预训练，大模型应拥有更强的针对行业数据的表征能力以及对不同场景的泛化能力，通过一个模型沉淀行业知识，更高效的使用 AI 技术解决行业业务问题；
- b) 应支持结合 workflow 的低门槛 AI 开发，应提供自动化 workflow，自动数据处理，自动化调参，自动化生成模型，减少对 AI 开发工程师专业依赖。针对井下数据受光照、粉尘、拍摄角度、设备新旧程度、污渍等因素造成图像过曝、欠曝、运动模糊、图像损毁等原因影响，workflow 提供图像质量自动评估及图像增强算法，通过亮度调节、自动图像平移/旋转/缩放，图像加噪声，图像仿真等算法对图像进行预处理，提升图像识别精度。同时可以根据更新数据快速进行模型迭代，做到“边学边用”。针对不

同区域各地方言情况，做到通过自动化训练文本语言模型小包的方式，实现大模型复用，语言模型小包方言个性化部署，达到区域场景定制效果；

- c) 涵盖百万参数量级到十亿参数量级的 3 个数量级规模模型，支持按需灵活部署；
- d) 应支持边用边学，自我进化能力，支持终身学习机制，确保在模型持续演进过程中对模型先前的用例样本其识别性能不下降，克服在模型进化中的遗忘机制，实现模型的持续优化；
- e) 应支持解决 AI “小数据” 问题：结合行业预训练模型，实现零训练样本下缺陷样本识别，应较传统方法发现率提升 2~3 倍。同时更好应对缺陷样本中长尾分布等 “小数据” 问题；
- f) 宜支持行业定制，结合行业先进知识开发定制化解决方案。

### 5.3. 算法兼容性要求

算法兼容性应满足如下要求：

- a) 算法应支持物体检测、图像分类、字符识别（OCR）、人体动作识别等任务场景；
- b) 算法应支持不同的系统平台，如 Windows、Linux、Android、MacOS 等；
- c) 算法应支持不同的硬件产品，如 Intel CPU、NVIDIA GPU、Jetson、飞腾、昆仑芯、昇腾、ARM CPU 等，部分硬件包含 Python 部署和 C++ 部署；
- d) 算法应支持不同量级的数据、不同类型数据进行模型训练和模型推理。

### 5.4. 兼容算法框架要求

在兼容算法基础上，应兼容多种开发框架，便于模型更好的建立，宜支持飞桨(PaddlePaddle)、Keras、Tensorflow、Pytorch 等不同的算法框架，实现模型的简单、灵活易用、高效内存使用、动态计算等功能。

硬件系统应支持建立机器学习模型，满足深度学习模型训练模型预测的基本环境。

矿山大模型应具备直接调用多种模型的能力，可调用的文件格式应包括：

\*.ckpt (TensorFlow 模型)、\*.h5(Keras 模型)、\*.pth(Pytorch 模型)等。

## 6. 大模型接口技术要求

### 6.1. 矿山大模型与上层应用之间的接口

大模型提供行业能力可被第三方应用调用,为了使用接口更加安全,应符合以下规范:

- a) 采用通过安全通道的 HTTPS 协议;
- b) 接口需要进行认证鉴权;
- c) HTTP 协议风格宜采用 `restful` 接口,便于理解使用。

### 6.2. 矿山大模型与数据源之间的接口

数据源包括矿山边缘节点、应用使能模块、数据使能模块等,提供训练数据给大模型用于预训练。数据源上传训练数据,包括提供单个数据上传 API 接口和大量的文件存储系统上传接口两部分功能,其中单个数据上传接口,必须采用通过安全通道的 HTTPS 协议,HTTP 协议风格宜采用 `restful` 接口,便于理解使用。所有接口需要进行认证鉴权。

### 6.3. 矿山大模型与人工智能系统使能平台之间的接口

大模型下发训练任务给人工智能系统使能平台用于训练任务的启动,符合以下规范:

- a) 宜采用通过安全通道的 HTTPS 协议;
- b) 接口需要进行认证鉴权;
- c) HTTP 协议风格宜采用 `restful` 接口,便于理解使用。

## 7. 大模型对数据的技术要求

提供的图片数据宜不少于 10TB,中文文本数据宜不少于 1TB,语音数据宜至少满足 10 万小时要求。

### 7.1. L0 通用预训练大模型数据要求

通用预训练模型数据应遵循如下要求:

- a) 应由底层的通用平台提供，数据规模在 1 亿规模以上，且尽可能覆盖多种场景。
- b) 图片质量宜达到至少 720P 分辨率，不能出现过曝，低曝，运动模糊等不符合要求的图像。
- c) 对语音的数据要求：
  - 1) 大数据数据要求：中文语音音频数据，应满足16k 16bit要求。标注标签为中文简体汉字，如包含英文，统一大小写用词规范进行标准化转写，标注文件统一采用utf-8格式。音频数据符合信息安全规范要求，获取途径合规合法。大数据要求至少十万小时量级；
  - 2) 定制数据要求：语音数据要求同“大数据数据要求”，文本数据要求统一编码gbk格式，中文简体汉字，英文字母，阿拉伯数字表示成应用场景真实使用的汉字或者英文，无任何出简体汉字英文字母空格外的特殊符号。定制数据音频数据满足至少100小时，文本数据至少1M；
  - 3) 短板数据要求：语音数据要求同“大数据数据要求”，主要挖掘模型低置信度短板数据，弥补长尾数据覆盖度，100小时达到相对提升10%左右目标。

## 7.2. L1 行业预训练大模型数据要求

L1 行业预训练大模型数据应遵循如下要求：

- a) L1 行业预训练大模型的参数量至少保证在 1000 万量级以上。
- b) 对图片的数据要求：
  - 1) 图片质量应达到至少720P分辨率，不能出现过曝，低曝，运动模糊等不符合要求的图像；
  - 2) 矿山行业大部分场景环境特殊，光线效果差，成像图片需要清晰没有模糊（如煤泥遮挡，或者灯光弱，摄像头没有照到识别的主体）；
- c) 对语音的数据要求：同“L0 通用预训练大模型数据要求”。

## 7.3. L2 部署模型数据要求

L2 部署模型数据应遵循如下要求：

- a) 对图片的数据要求：
- 1) 应满足每类的场景大于50张图片；图片质量应达到至少720P分辨率，不能出现过曝，低曝，运动模糊等不符合要求的图像；
  - 2) 矿山行业大部分场景环境特殊，光线效果差，成像图片需要清晰没有模糊（如煤泥遮挡，或者灯光弱，摄像头没有照到识别的主体）；
- b) 对语音的数据要求：同“L0通用预训练大模型数据要求”。

## 8. 大模型对硬件平台的要求

### 8.1. 计算能力

计算能力应遵循如下要求：

- a) 基于大模型的训练和推理服务器，包含通用CPU训练/推理服务器和基于GPU/NPU的训练推理服务器。
- b) 训练服务器主要部署在中心侧，推理服务器主要部署在中心侧（集团中心侧）和边缘侧（矿山侧）。
- c) 边缘节点（矿山侧）宜包含视频管理服务器、推理服务器、一台边缘一体机组成，应符合下列规定：
  - 1) 边缘推理服务器配置应符合下列规定：
    - 支持通过管理软件对边缘节点进行统一纳管，并为边缘节点部署和更新业务应用（容器应用、AI模型文件），实现两端协同处理业务的功能；
  - 2) 视频接入平台配置应符合下列规定：
    - 接入、存储、管理摄像头视频应采用嵌入式Linux操作系统，支持7×24小时稳定运行；
    - 超融合，高性能；
    - 支持视频与图片混合直存，内置流媒体转发服务，相比传统视频监控架构，省去存储服务器以及转发服务器，具有绿色节能的特点；
    - 支持媒体块存储技术，消除磁盘碎片提升磁盘写入性能；
    - 支持录像备份，关键数据双重备份；

- 支持缓存补录技术，保证业务数据的完整性；
  - 3) 应配置高性能电脑，包含主机、显示器，需带独立显卡；
  - 4) 矿用隔爆或本安摄像机应高于200W像素，宜有变焦功能；
  - 5) 应配置补光灯，增加光照，保障成像质量。
- d) 应配置交换机、防火墙等网络设备，实现摄像头入网，边缘节点和矿上生产系统数据打通。

## 8.2. 软件能力

软件能力应具备如下要求：

- a) 通过该平台的应用，AI 开发人员可以通过低代码、可视化界面，以及配套的工作流程，快速高效地完成 AI 的端到端工作，包含数据导入、模型训练、模型部署等动作。
- b) AI 业务应用平台应实现矿山应用场景。平台采用集团侧和矿山侧业务应用两级架构，在集团侧实现数据集管理、模型调用、模型部署、模型下发、运行监控等模型管理功能，并实现统计分析和驾驶舱等功能，结合数字孪生场景，在矿山侧实现报警管理、业务分析、统计报表、联动处置等功能。
- c) 边缘节点场景应用应实现模型的正常部署、管理、升级等，同时能在边侧实现正常的推理功能，输出识别结果，实现 AI 应用在具体场景下的功能呈现。

## 9. 大模型的部署规范

### 9.1. 大模型部署架构

面向矿山生产领域，基于中心云和边缘节点，宜打造矿山智能化建设云边协同的二级架构，建设智能训练中心和人工智能系统应用平台，以支持矿山智能场景应用，云边协同架构如下图所示，中心云负责完成云端 L1 行业预训练大模型的训练工作，通过云端 L1 行业预训练大模型以及 AI 开发工作流输出 L2 部署模型，将 L2 部署模型分发到各个边缘节点上完成推理，达到训练部署的闭环。



图 1 云边协同架构图

部署在中心云侧的人工智能大模型平台，提供统一的开发、训练和运营运维能力。利用中心训练算力池进行模型训练、模型评估，完成训练和验证的模型，可按需部署到边缘推理服务器生产环境。

云端-边端模型关系及参数量：云端 L1 行业预训练大模型参数规模在十亿量级，边端部署模型由云端 L1 行业预训练大模型通过抽取以及蒸馏（可选）获得，其参数范围覆盖百万至十亿量级，支持边端多样化部署。

## 9.2. 中心云部署要求

集团中心应部署人工智能训练平台，构建算法模型开发与训练能力，实现模型统一部署管理，应符合以下要求：

- a) 应部署人工智能计算引擎，使用主流深度学习计算框架训练模型，完成业务仿真系统和开发者研究开发，使用先进的人工智能计算引擎对业务进行进一步的突破；
- b) 应部署人工智能算法模型开发与训练平台，为业务人员和开发人员建立人工智能算法模型的开发和训练环境，同时随着数据收集渠道的增加，可用的数据量快速膨胀，传统的单机训练已无法满足要求，开发训练环境应支持分布式训练能力，以持续提升训练效率和精度；

- c) 应实现人工智能模型管理与部署，随着算法模型不断丰富和应用范围的扩大，需要对人工智能算法模型进行有效管理并快速部署到生产环境，建立矿山等生产领域的人工智能生态。

### 9.3. 边缘节点部署要求

边缘节点应部署数字平台，集团侧和边缘侧两级管理，边缘侧应用系统与边缘算法模型配合使用，部署在各生产单位，满足生产单位业务需求，支撑矿山可监控应用运行。



## 附录 A

### (资料性附录)

#### 矿山大模型应用场景

人工智能对产业和社会变革的价值得到各行业广泛认同,煤矿行业拥有丰富的人工智能应用需求。煤矿行业对于人工智能技术的应用已经有了一些积累,大量涌现的智能 AI 应用为煤矿安全生产和智能化管理提供创新型的解决方案。开展人工智能在煤矿智能化领域的技术攻关和应用创新,打造智能矿山的产业生态圈,可以全面提升煤矿智能化水平。

依托人工智能平台、AI 算法方面的技术积累,引入 AI 大模型的理念,降低人工智能开发门槛,缩短模型训练周期,开启人工智能的“工业化开发模式”。以下是人工智能在各生产系统应用的典型应用场景。

图 1 矿山大模型应用场景表

大类	场景描述
1.综采面	智能放顶煤(视频、声音、煤层厚度等因子)
	综采面全景视频拼接
	滚筒、护帮板防碰撞检测
	煤壁片帮识别
	刮板输送机堆煤识别
	刮板输送机断链识别
	刮板输送机刮板损坏识别
	刮板输送机异物检测
	危险区域进入
	线槽电缆掉落检测
	转载机破碎滚筒挡煤链
2.掘进面	危险区域进入
	空顶作业检测
	敲帮问顶合规检测
	探放水作业合规检测
	瓦斯抽采作业合规检测
	作业序列视频分析
	卸压钻钻杆数统计
3.机电	智能化语音询问与调度
	设备运行状态监测
	设备故障诊断与智能运维

	钢丝绳表面状态监测
4.主运输	皮带异物识别
	转载点堆煤识别
	皮带跑偏识别
	皮带撕裂识别
	煤量识别（煤量视频图像分析）及煤流调速
	危险区域进入
	皮带卡子损坏检测
	光纤传感托辊检测
5.地质	矿压危险等级评估
	矿井水害预测及分析
	隐蔽致灾地质构造预测
	冲刷带发育区预测
	瓦斯富集区预测
	构造煤发育区预测
	煤层厚度分布预测
	煤岩异构区域分布检测
6.洗煤厂	煤流超粒度识别
	溜槽、胶带机机头卡堵识别
	浅槽分选机异常状态识别
	磁选机翻花断流识别
	烟雾、火焰识别
	煤泥水泄漏识别
	筛机煤层厚度监测
	胶带机煤量精准预测
	振动筛异常分析
	智能洗选参数优化
7.人员行为	出入井人脸识别
	出入井人数统计
	人员违章概率排名及人员跟踪
	人员未按安全防护规范穿戴检测
	人员睡岗识别
	人员离岗识别
	人员摔倒识别
	抽烟识别
人员违规跨越电子围栏	
8.矿山环境保护与修复	矿山生态环境空-天-地协同监测
	矿区开采沉陷监测
	矿区土地利用变化遥感监测
	矿区植被退化监测
	矿区水体质量监测
	矿区生态修复效果评价
	矿区生态退化预警

9.其它	安全隐患分析
	危险区域周界防护
	智慧溜井管理

## 附录 B

### (资料性附录)

#### 人工智能系统功能参考架构

##### B.1 人工智能系统在智能化矿山中的技术参考架构图

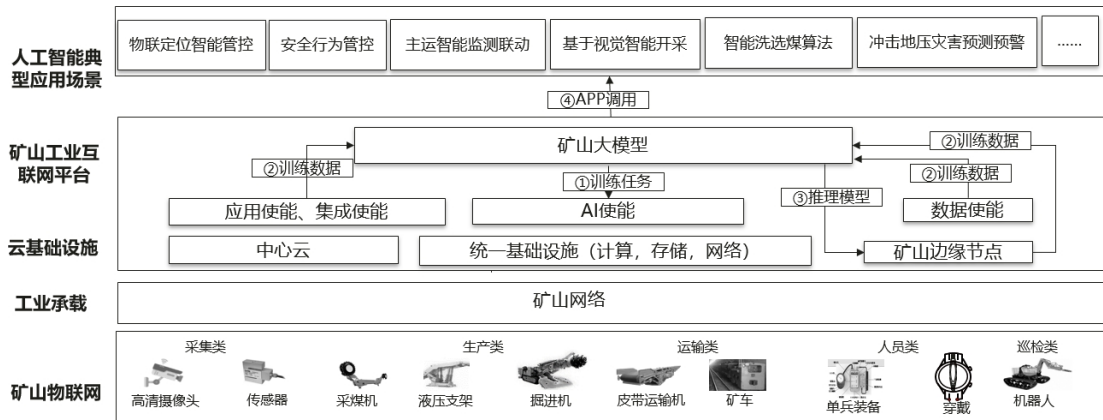


图 2 智能化矿山技术参考架构图

参考国家能源局《煤矿智能化建设指南（2021年版）》，人工智能系统在智能化矿山技术参考架构图中的位置如图 2 所示。各层的功能描述如下：

- a) 矿山物联网：通过各种感知、信息传输与处理技术，构成矿山中人与人、人与物、物与物相联，实现智能化识别、定位、跟踪监控和管理。
- b) 工业承载：实现矿山物联网和云平台之间传送各种数据业务。
- c) 云基础设施：提供云服务的硬件层，组件通常分为计算、网络和存储等。
- d) 矿山工业互联网平台：面向矿山行业信息化、数字化、智能化需求，构建基于海量数据采集、汇聚、分析的服务体系，支撑矿山行业泛在连接、弹性供给、高效配置的工业云平台。
- e) 人工智能典型应用：基于矿山工业互联网平台构建 AI 应用场景。

## B.2 基于矿山大模型的人工智能参考架构图

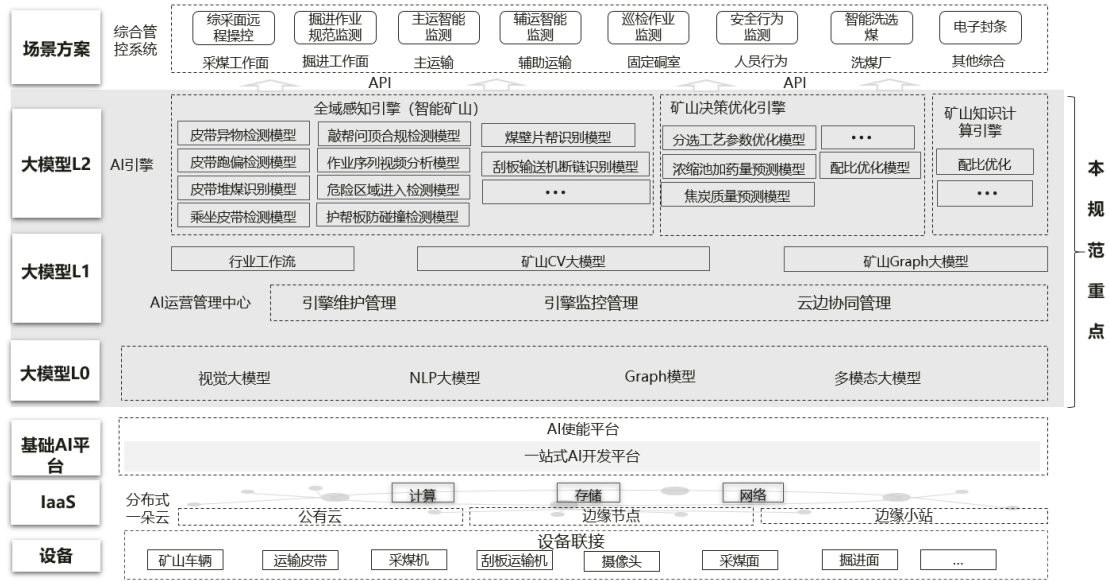


图 3 基于矿山大模型的人工智能参考架构图

基于矿山大模型的人工智能功能架构如图 3 所示，各层的功能描述如下：

- 设备层为工业互联网平台提供底层数据输入，实现海量工业数据的精准、实时采集和集成。
- IaaS 基础设施层包括硬件服务器、数据存储、5G 网络及虚拟化技术相关的基础设施，可以为工业互联网平台的安全、稳定运行提供硬件支撑。
- 基础 AI 平台提供全流程的 AI 开发与推理服务，训练作业管理、多开发框架支持、模型统一管理、服务按需部署能力，支持 GPU、CPU、VPU、FPGA 的异构资源调度与统一管理，帮助用户快速创建和部署模型、管理全周期 AI workflow，助力应用开发者快速完成模型开发与上线，使能行业创新 AI 业务。
- L0 通用预训练大模型应从大量训练数据（有标签或无标签）中学习高效特征表示方法，训练出有巨大参数量的神经网络模型。包括视觉预训练大模型、自然语言处理预训练大模型、多模态预训练大模型等，图 3 中所示模块按需部署。
- L1 行业预训练大模型基于 L0 通用预训练大模型，通过海量行业数据训练得到的行业预训练大模型，同时配套行业场景模型开发工具链，支持基于 L1 行业预训练大模型生产 L2 部署模型，图 3 中所示模块按需部署。

- f) L2 大模型为部署模型，基于 L1 行业预训练大模型和少量场景化数据微调，产出的适用于场景化 AI 推理模型，满足实际落地场景的推理性能，图 3 中所示模块按需部署。

### 参考文献

- [1] GB/T 15259—2008 煤矿安全术语
  - [2] GB/T 34679-2017 智慧矿山信息系统通用技术规范
  - [3] 煤矿智能化建设指南（2021年版） 国家能源局、国家矿山安全监察局
  - [4] 煤矿安全规程（2022版） 应急管理部
  - [5] GB/T 5271.28-2001 信息技术 词汇 第28部分:人工智能 基本概念与专家系统
  - [6] KSSJ/BM12-2023 井工煤矿数据分类及编码规范 第1部分：基础类
  - [7] KSSJ/BM12-2023 井工煤矿数据分类及编码规范 第2部分：生产类
  - [8] KSSJ/BM12-2023 井工煤矿数据分类及编码规范 第4部分：管理类
  - KSSJ/YY11-2023 数据共享基本要求
  - [9] A. Vaswani et al., Attention is all you need, in NeurIPS, 2017.
  - [10] T. B. Brown et al., Language models are few-shot learners, in NeurIPS, 2020.
-