团体标准

发 布

中国电机工程学会

20XX—XX—XX实施

20XX—XX—XX发布

电力人工智能平台架构及技术要求

Architecture and technical requirements of power artificial intelligence platform

（征求意见稿）

T/CSEE XXXX—YYYY

ICS 35.020

CCS K85

目 次

[前 言 3](#_Toc126154203)

[1 范围 4](#_Toc126154204)

[2 规范性引用文件 4](#_Toc126154205)

[3 术语和定义 4](#_Toc126154206)

[4 缩略语 6](#_Toc126154207)

[5 架构要求 6](#_Toc126154208)

[5.1 概述 6](#_Toc126154209)

[5.2 总体架构 6](#_Toc126154210)

[5.3 功能架构 8](#_Toc126154211)

[5.4 技术架构 9](#_Toc126154212)

[5.5 数据架构 10](#_Toc126154213)

[6 技术要求 11](#_Toc126154214)

[6.1 功能要求 11](#_Toc126154215)

[6.2 性能要求 12](#_Toc126154216)

[6.3 安全要求 12](#_Toc126154217)

[6.4 硬件要求 12](#_Toc126154218)

[7 算法模型共享应用要求 13](#_Toc126154219)

[7.1 算法模型共享方式 13](#_Toc126154220)

[7.2 算法模型文件 13](#_Toc126154221)

[7.3 算法模型基本应用方式 14](#_Toc126154222)

前 言

本文件按照《中国电机工程学会标准化管理办法》、《中国电机工程学会标准化管理办法实施细则》的要求，依据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。

本文件由中国电机工程学会提出。

本文件由中国电机工程学会人工智能标准专业委员会技术归口和解释。

本文件起草单位：中国电力科学研究院有限公司、国网四川省电力公司、国网山东省电力公司、国家电网有限公司大数据中心、国网信息通信产业集团有限公司、国网宁夏电力公司、国家电网有限公司信息通信分公司。

本文件主要起草人：\*。

本文件首次发布。

本文件在执行过程中的意见或建议反馈至中国电机工程学会标准执行办公室（地址：北京市西城区白广路二条1 号，100761，网址：http：//www.csee.org.cn，邮箱：[cseebz@csee.org.cn](mailto:cseebz@csee.org.cn)）。

电力人工智能平台架构及技术要求

1. 范围

本文件规定了电力人工智能平台建设的架构要求、技术要求、算法模型共享应用要求。

本文件适用于电力人工智能平台的规划、设计、建设和运维。

1. 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件，仅注日期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 5271.1—2000 信息技术 词汇 第1部分：基本术语

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

GB/T 5271.29—2006 信息技术 词汇 第29部分：人工智能语音识别与合成

GB/T 5271.31—2006 信息技术 词汇 第31部分：人工智能机器学习

GB/T 5271.34—2006 信息技术 词汇 第34部分：人工智能神经网络

1. 术语和定义

下列术语和定义适用于本文件。



人工智能 artificial intelligence

一门交叉学科，通常视为计算机科学的分支，研究表现出与人类智能（如推理和学习）相关的各种功能的模型和系统。

[GB/T 5271.28—2001,定义28.01.01]

机器学习 machine learning

功能单元通过获取新知识或技能，或通过整理已有的知识或技能来改进其性能的过程。

[GB/T 5271.31—2006,定义31.01.02]

神经网络 neural network

由加权链路且权值可调整连接的基本处理元素的网络，通过把非线性函数作用到其输入值上使每个单元产生一个值，并把它传送给其他单元或把它表示成输出值。

[GB/T 5271.34—2006,定义34.01.06]

训练（在神经网络中） training（in neural network）

教会神经网络在输入值的样本和正确输出值之间作出结合的步骤。

[GB/T 5271.34—2006,定义34.03.18]



样本数据 sample data

其具备的特征能够反映总体数据情况的一部分个体数据。



推理 inference

从已知前提导出结论的推理方法。

在人工智能领域中，前提是事实或规则。

术语“推理”既指过程也指结果。

[GB/T 5271.28—2001,定义28.03.01]

深度学习 deep learning

深度学习是机器学习的一种，深度学习的概念源于人工神经网络的研究，含多个隐藏层的多层感知器就是一种深度学习结构。



深度学习框架 deep learning framework

一种支持深度学习模型设计、训练和推理的工具。



资源 resource

执行所要求的操作而必需的数据处理系统的任何组成部分。

[GB/T 5271.1—2000,定义01.01.23]



配置 configuration

信息处理系统中的硬件和软件组织和互连起来的方式。

[GB/T 5271.1—2000,定义01.01.26]



接口 interface

两个功能单元共享的边界，它由各种特征（如功能、物理互连、信号交换等）来定义。

[GB/T 5271.1—2000,定义01.01.38]

计算机视觉 computer vision

功能单元获取、处理和理解可视数据的能力。

[GB/T 5271.28—2001,定义28.01.19]

语音识别 speech recognition

利用功能单元进行的，从语音信号到语音内容的某一表示的转换。

[GB/T 5271.29—2006,定义29.01.30]

1. 缩略语

下列缩略语适用于本文件。

SVM：支持向量机（Support Vector Machines）

SVD：奇异值分解（Singular Value Decomposition）

PCA：主成分分析（Principal Component Analysis）

OLS：最小二乘法（Ordinary Least Squares）

CNN：卷积神经网络（Convolutional Neural Networks）

DNN：深度神经网络（Deep Neural Networks）

YOLO：YOLO模型（You Only Look Once）

TFLOPS：万亿次浮点运算每秒（Tera Floating-point Operations Per Second）

FMI：FMI指数（Fowlkes and Mallows Index）

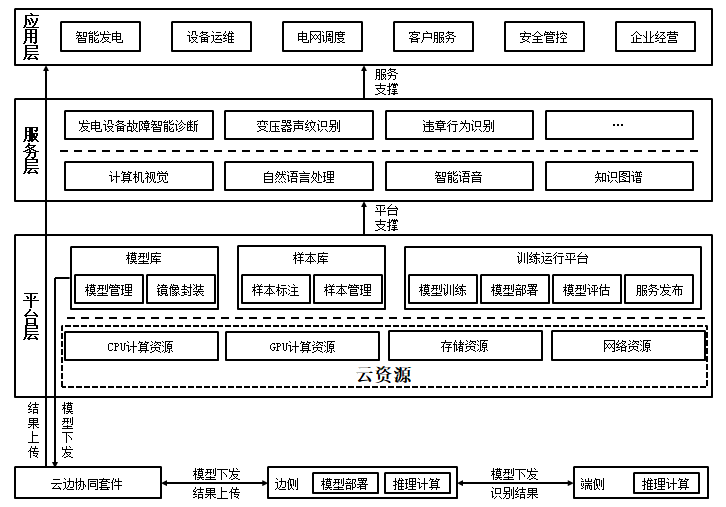
2. 架构要求
   1. 概述

电力人工智能平台架构要求包括：

1. 总体架构：规定电力人工智能平台总体结构以及和其它平台及组件的关系；
2. 功能架构：规定电力人工智能平台总体功能结构；
3. 技术架构：规定电力人工智能平台主体组件的技术选型和技术范围；
4. 数据架构：规定电力人工智能平台数据的架构。
   1. 总体架构
      1. 概述

电力人工智能平台架构应包括：模型库、样本库和训练运行平台3部分，这3部分可基于已有云计算提供的基础资源共同构成平台层。平台层和服务层构成人工智能中台，支撑应用层的各类业务应用。电力人工智能平台总体架构如下图1，各部分要求为：

1. 电力人工智能平台的建设应基于云的CPU计算资源、AI芯片计算资源、存储资源和网络资源；
2. 实验训练应使用样本库的样本集，通过训练得到可用的模型，模型可输出至模型库；
3. 模型库存放训练出的算法模型或采购来的第三方算法模型；
4. 人工智能平台应提供模型推送功能，可通过云边协同套件推送至边侧设备；
5. 服务层应包括：计算机视觉、自然语言处理、智能语音和知识图谱等基础服务，以及变压器渗漏油识别等专用服务；
6. 人工智能中台可支撑设备运维、电网调度、客户服务、安全管控和企业经营等电力业务；
7. 边缘侧设备的样本数据可通过云边协同套件上传至样本库；
8. 平台层支撑各类人工智能服务，并通过服务支撑应用层的业务应用。



1. 电力人工智能平台总体架构图
   * 1. 训练运行平台

训练运行平台应包括：模型训练、模型部署、模型评估和服务发布等4部分。各部分应满足的要求为：

1. 不同框架的环境创建采用统一的操作入口；
2. 模型开发提供命令行式全功能操作方式；
3. 模型训练包括命令行式训练、可视化建模和自动化建模；
4. 模型评估包括模型选择、验证集选择、验证环境选择和结果展示。
   * 1. 模型库

模型库应包括：模型管理和镜像封装2部分。模型库支撑的服务层的具体要求为：

1. 基础模型服务可包括：
   1. 智能语音：语音识别和语音合成等；

2）自然语言处理：信息检索和问答对话等；

3）计算机视觉：图像识别和视频分析等；

4）知识图谱：图谱构建和图谱服务等。

1. 业务模型服务可包括但不限于以下几方面：

1）变压器渗漏油识别；

2）发电设备故障智能诊断；

3）客服语音质检；

4）电网设备故障事件抽取；

5）线路金具缺销钉识别；

6）电力文本查重；

7）变压器声纹识别；

8）电力事件演进关系抽取；

9）未戴安全帽违章识别；

10）舆情分析预警；

11）输变电设备故障诊断；

12）缺陷设备属性抽取。

1. 模型管理服务应包括：

1）对模型进行统一生命周期管理；

2）对模型进行标签、收藏等特定服务；

3）支持多厂商多框架多功能模型统一纳管。

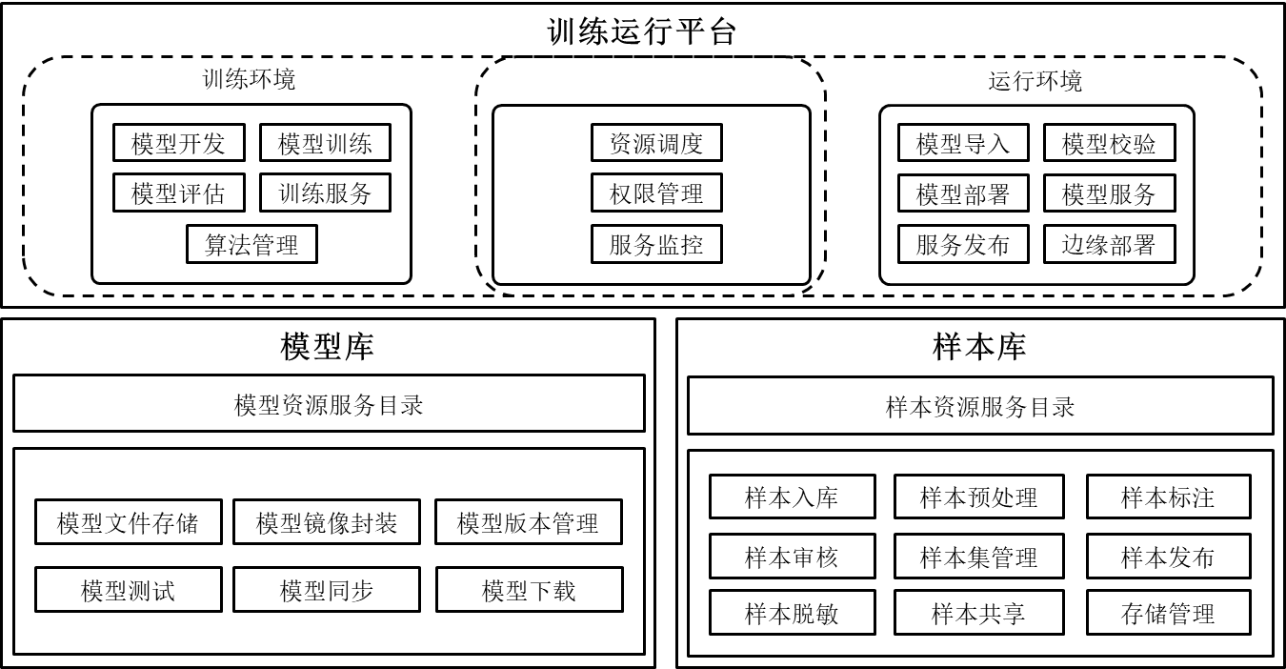
* + 1. 样本库

样本库应包括：样本管理和样本标注2部分，应满足如下要求：

1. 样本管理包括：样本数据集管理、标签管理和样本信息管理等；
2. 样本标注包括：文本标注、视频标注、音频标注和图像标注。
   1. 功能架构

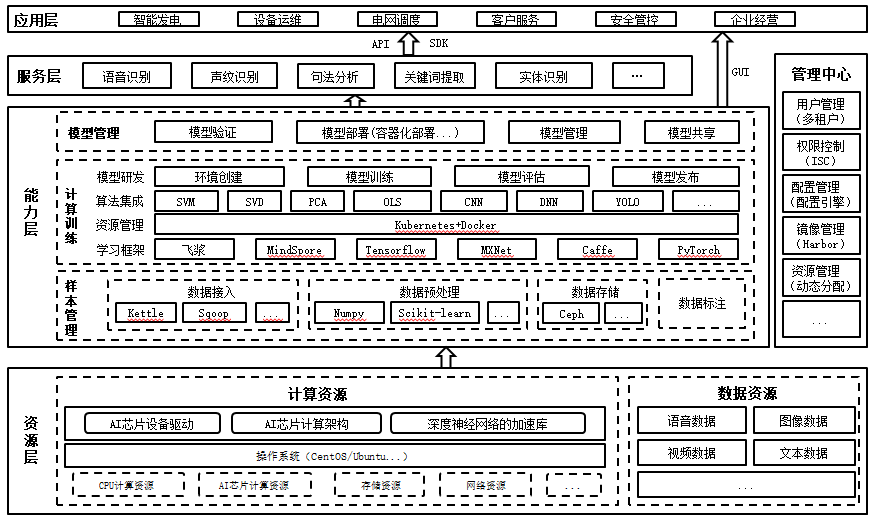
电力人工智能平台功能应包括：样本库、模型库和训练运行平台。电力人工智能平台功能架构如下图2，功能架构应满足如下要求：

1. 训练运行平台可分为训练环境和运行环境，训练环境包括：模型开发、模型训练、模型评估、模型服务、算法管理、资源调度、权限管理和服务监控，运行环境包括：模型导入、模型校验、模型部署、模型服务、服务发布、边缘部署、资源调度、权限管理和服务监控；
2. 样本库包括：样本资源服务目录、样本入库、样本预处理、样本标注、样本审核、样本集管理、样本发布、样本脱敏、样本共享和存储管理；
3. 模型库包括：模型资源服务目录、模型文件存储、模型镜像封装、模型版本管理、模型调试、模型同步和模型下载。



1. 电力人工智能平台功能架构图
   1. 技术架构

电力人工智能平台技术架构要求主要包括：应用层、服务层、能力层和资源层的要求。管理中心处于服务层和能力层。电力人工智能平台技术架构如下图3：



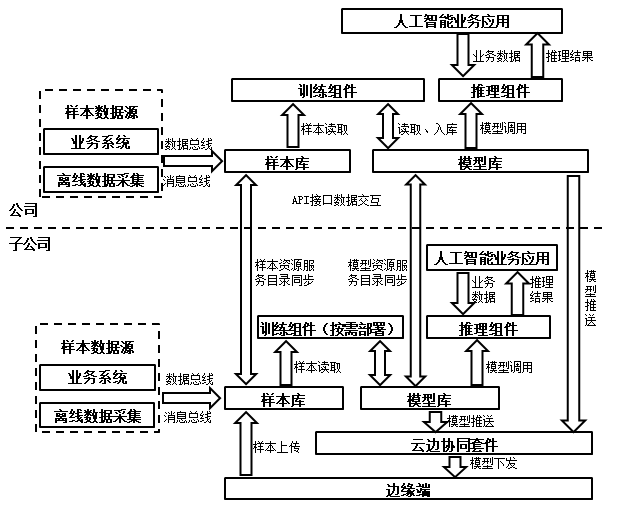
1. 电力人工智能平台技术架构图

电力人工智能平台技术架构要求为：

1. 应用层通过API和SDK两种方式使用平台的服务，通过GUI的方式访问能力层提供的能力；
2. 模型管理中的模型部署应采用容器化部署方式，应使用Kubernetes和Docker组件；
3. 算法集成应包括算法模块：SVM、SVD、PCA、OLS、CNN、DNN、YOLO等；
4. 学习框架应包括主流的开源深度学习框架；
5. 数据接入宜采用Kettle和Sqoop等组件；
6. 数据预处理宜支持Numpy、Scikit-learn等组件；
7. 数据存储宜支持Ceph等组件；
8. 配置管理模块宜采用配置引擎的方式开发；
9. 镜像管理应支持Harbor；
10. 资源应支持动态分配；
11. AI芯片计算资源包括但不限于：GPU和NPU等；
12. 操作系统宜采用宜采用Ubuntu 18.04及以上或openEuler 20.09及以上版本。
    1. 数据架构

宜构建样本资源目录多级管理体系，实现资源目录上送、下发的双向同步及样本资源目录、标注和原始数据的推送分发功能。应将数据按照图像、视频、语音、文本等类型存储，经过数据标注后形成样本库。电力人工智能平台数据架构如下图4，数据架构要求为：

1. 样本数据应来源于业务系统或通过离线数据采集，经由数据总线和消息总线推送至样本库，也可来源于边缘端；
2. 训练组件读取样本库的数据和模型库的模型，并将训练好的模型入库存储到模型库；
3. 推理组件调用模型库的模型，使用业务应用推送的业务数据，推理计算，将推理结果反馈给业务应用；
4. 公司级平台和子公司级平台的样本库可实现样本目录同步；
5. 公司级平台和子公司级平台的模型库可实现模型目录同步；
6. 模型库的模型可通过云边协同套件下发到边缘端。



1. 电力人工智能平台数据架构
2. 技术要求
   1. 功能要求
      1. 训练运行平台

训练运行平台功能要求为：

1. 项目管理：项目创建、项目切换、项目删除和项目监控；
2. 模型训练：应至少兼容飞桨、MindSpore、TensorFlow、MXNet、PyTorch和Caffe等开源深度学习框架，训练任务宜支持单卡内存级分配；
3. 模型评估：评估指标和评估报告管理，按算法类型分为：
4. 1）分类算法：准确率指标评估、召回率指标评估、F1指标评估；
5. 2）回归算法：误差平方和、决定系数和校正决定系数；
6. 3）聚类算法：紧密型、间隔性、纯度、标准化互信息、兰德指数、FMI、Jaccard指数等。
7. 算法管理：算法上架、算法分类和算法下架等
8. 容器管理：容器镜像文件导入和导出。
   * 1. 模型库

模型库功能要求为：

1. 模型管理：模型导入、模型删除、版本管理、模型标签、模型收藏和模型共享；
2. 模型测试：模型部署、在线测试和服务管理，模型测试服务发布应支持向导模式，宜支持一键自动发布测试服务，模型测试服务宜支持单卡内存级分配。
   * 1. 样本库

样本库功能要求为：

1. 数据接入：本地文件导入、HDFS数据导入、FTP文件导入、NFS数据导入和数据库数据导入；
2. 数据预处理：特征提取、灰度校正和平滑去噪等；
3. 数据标注：宜支持智能标注，以模型来给未标注数据进行标注。
   1. 性能要求
      1. 响应时限

人工智能平台响应时限应符合DL/T 1731-2017 电力信息系统非功能性需求规范的规定。

* + 1. 可靠性

人工智能平台应提供主要网络设备、通信线路和集群系统的硬件冗余，保证高可用性，在无不可抗力环境下应满足7×24小时服务不中断，具体要求为：

1. 数据完整性：存储节点发生故障时，应确保数据完整；
2. 算力完整性：计算节点发生故障时，应不影响训练和推理的执行结果；
3. 消息完整性：消息队列节点发生故障时，应确保消息不丢失，且不影响消息正常提交和消费；
4. 任务调度完整性：任务调度节点发生故障时，应不影响任务调度和执行；
5. 网络完整性：网络发生故障并恢复后，系统和任务、服务均应自动继续运行。
   * 1. 可扩展性

人工智能平台应支持资源横向扩展和系统平滑升级，资源扩展和系统升级过程应不影响现有训练任务和推理服务。

* 1. 安全要求

人工智能平台的安全要求包括：

1. 应符合GB/T 18336—2015 《信息技术 安全技术 信息技术安全评估准则》的规定；
2. 应符合GB/T 22239-2019《信息安全技术网络安全等级保护基本要求》的规定；
3. 平台中人工智能算法在支撑业务应用时，应充分考虑算法计算精度突然降低、计算结果出错、计算结果超时等状况下对业务系统造成的不利影响。
   1. 硬件要求

人工智能平台中运行训练组件、推理组件和具有预标注功能的标注组件的服务器应具备独立的AI芯片计算资源。具体要求如下：

1. 对训练平台，AI芯片计算资源单卡内存不应低于8GB，宜达到32GB或以上，单卡内存带宽不应低于700GB/秒，单卡算力应达到以下条件之一：
2. 1)单精度浮点算力至少达到9TFLOPS；
3. 2)半精度浮点算力至少达到18TFLOPS。
4. 在运行环境，AI芯片计算资源单卡内存不应低于4GB，单卡内存带宽不应低于100GB/秒，单卡算力应达到以下条件之一：
5. 1)单精度浮点算力至少达到9TFLOPS；
6. 2)半精度浮点算力至少达到18TFLOPS；
7. 3)INT8精度算力至少达到36TOPS。
8. 训练平台宜使用固态硬盘作为训练样本集的大容量缓存。
9. 算法模型共享应用要求
   1. 算法模型共享方式
      1. 算法模型文件方式共享

算法模型宜支持以模型文件方式共享，可利用算法模型文件实现二次训练和部署应用，算法模型文件应包括但不限于以下文件：

1. 算法模型源文件；
2. 算法模型配置文件；
3. 算法模型运行脚本文件。
   * 1. 算法模型容器方式共享

算法模型应支持以容器方式共享，容器应包括算法模型和算法模型运行所需的基础环境，可利用容器直接部署应用，容器中算法模型文件应包括但不限于以下文件：

1. 算法模型二进制文件；
2. 算法模型配置文件；
3. 算法模型运行脚本文件。
   1. 算法模型文件
      1. 算法模型源文件

算法模型源文件应包括由训练框架和数据集经过模型训练后得到的所有参数存储文件。所涉及的训练框架包括但不限于Caffe、PyTorch、TensorFlow、MXNet、飞桨、MindSpore等深度学习框架。根据不同描述语言和开发框架，算法模型源文件对应要求如下：

1. Caffe框架。算法模型源文件宜包括caffemodel文件和prototxt文件，caffemodel存储模型参数，prototxt存储模型网络结构；
2. PyTorch框架。算法模型源文件宜包括pth文件，用于存储模型的网络结构和参数；
3. TensorFlow框架。模型源文件宜包括meta文件、data文件和index文件，meta文件存储模型的网络结构，data文件存储模型的网络参数，index文件为张量描述列表或网络结构和参数整合后的h5/pb文件；
4. MXNet框架。算法模型源文件宜包括params文件和json文件，params文件存储模型参数，json文件存储模型网络结构；
5. 其他框架。算法模型源文件可采用通用PMML预言模型标记语言描述，文件格式为xml，可用于描述和存储算法模型。
   * 1. 算法模型配置文件

模型配置文件描述所有可配置的变量，针对不同框架的配置文件格式要求如下：

1. Caffe框架。算法模型配置文件应为prototxt格式；
2. PyTorch框架。算法模型配置文件应为py格式；
3. TensorFlow框架。算法模型配置文件应为config格式；
4. MXNet框架。算法模型配置文件应为params格式；
5. 其他框架。算法模型配置文件可采用xml格式。

各类配置文件中可调参数应包含但不限于以下字段：

1. 可调参数说明

|  |  |  |
| --- | --- | --- |
| **序号** | **字段名称** | **含义说明** |
| **1** | BatchSize | 描述批处理参数 |
| **2** | Width | 描述输入数据的宽度 |
| **3** | Height | 描述输入数据的高度 |
| **4** | Channel | 描述输入数据通道数 |

* + 1. 算法模型运行脚本文件

算法模型运行脚本文件应对模型加载、训练和推理提供代码支持，并以模型配置文件所指定的参数运行模型文件。

* 1. 算法模型基本应用方式
     1. 部署应用

算法模型应支持以部署方式进行应用，部署方式应支持容器部署，宜支持模型文件部署。

* + 1. 二次训练应用

算法模型宜支持以二次训练方式应用。根据算法模型的开发语言、深度学习训练框架、数据集和描述文档等内容，宜支持对算法模型进行相应参数调优。

* + 1. 应用接口

算法模型应支持以接口方式对外提供服务，接口方式应至少支持API和SDK两种接口方式中的一种。

**━━━━━━━━━━━**