

中华人民共和国通信行业标准

YD/T [×××××]—[××××]
[代替 YD/T]

人工智能开发平台通用能力要求 第 1 部分：功能要求

The general capability requirements for artificial intelligence development
platform— Part 1: Function requirements

[点击此处添加与国际标准一致性程度的标识]

(报批稿)

[点击此处添加本稿完成日期]

[××××]-[××]-[××]发布

[××××]-[××]-[××]实施

中华人民共和国工业和信息化部 发布

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 缩略语 2

5 概述 2

6 平台通用功能指标 3

 6.1 数据处理 3

 6.2 模型构建 5

 6.3 模型部署 8

 6.4 支撑与服务 9

附 录 A（资料性） 功能分级定义 13

前 言

本文件按照GB/T 1.1—2020给出的规定起草

本文件是《人工智能开发平台通用能力要求》系列标准的一部分，该标准的结构和名称预计如下：

- 第1部分：功能要求；
- 第2部分：安全要求；
- 第3部分：性能要求。

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。

本文件由中国通信标准化协会提出并归口。

本文件起草单位：中国信息通信研究院、北京百度网讯科技有限公司、阿里云计算有限公司、华为技术有限公司、腾讯云计算（北京）有限责任公司、北京九章云极科技有限公司、北京金山云网络技术有限公司、厦门渊亭信息科技有限公司、中国工商银行股份有限公司、中国电信集团有限公司、中国联合网络通信集团有限公司、杭州海康威视数字技术股份有限公司、马上消费金融股份有限公司、杭州东信北邮信息技术有限公司、浪潮电子信息产业股份有限公司、新华三技术有限公司、联想（北京）有限公司、中国移动通信集团有限公司、北京研华兴业电子科技有限公司、中国信息通信科技集团有限公司、北京京东尚科信息技术有限公司、蚂蚁科技集团股份有限公司、中兴通讯股份有限公司、上海商汤智能科技有限公司、南京新一代人工智能研究院、北京小米移动软件有限公司、同盾科技有限公司、浙江中医药大学、南京甄视智能科技有限公司、上海湃道智能科技有限公司、京东方科技集团股份有限公司。

本文件主要起草人：董昊、曹峰、董晓飞、曹皓、靳伟、杨广贺、丁晓晗、刘海涛、方磊、于凯、晏晶晶、李笑如、黄博远、丁子建、刘翊、蒋宁、王帅、张宏伟、宋虎、李瑾瑜、刘倩、程帅、黄兵明、吴浩然、方家乐、缪小川、曹予飞、邸双朋、袁智、虞文明、史超、丁欣卉、王蕴韬、石霖、郑明智、黄琳、屈少杰、张昭智、姜幸群。

引 言

近年来，人工智能技术产业快速发展，各类人工智能开发平台也不断发展壮大，并被大量企业使用。人工智能开发平台既是连接产业上下游的重要枢纽，也是人工智能新基建的重要组成。此类平台主要提供传统机器学习和深度学习开发、部署、应用、运维的全流程服务，能够满足算法工程师、数据科学家、行业专家等不同类型开发者的使用需求，有效推进人工智能赋能各行各业。

目前，人工智能开发平台还处在发展初期，在主要功能层面已达成一定程度的产业共识，但是在功能全面性、性能稳定性、标准一致性等方面仍有待改善和规范，为了引导产业有序发展、规范行业自律，并为用户提供选型参考，特联合相关单位开展人工智能开发平台标准规范的编制工作。

该系列文件的结构、名称和目的如下：

——第1部分：功能要求。有助于对人工智能开发平台功能能力的构建、选型、评估、验收等工作。

——第2部分：安全要求。有助于对人工智能开发平台安全能力的构建、选型、评估、验收等工作。

——第3部分：性能要求。有助于对人工智能开发平台性能能力的构建、选型、评估、验收等工作。

本文件是“人工智能开发平台通用能力要求”系列文件第1部分：功能要求，是为人工智能开发平台产品的研发者和选型者提供评估参考规范。

人工智能开发平台通用能力要求 第 1 部分：功能要求

1 范围

本文件规定了人工智能开发平台通用功能的要求，主要包括数据处理、模型构建、模型部署、支撑与服务等核心环节的功能要求。

本文件适用于指导第三方测评机构对人工智能开发平台通用功能的评估、验收等工作。

2 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，仅所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 5271.31-2006 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

3 术语和定义

GB/T 5271.31-2006和GB/T 5271.34-2006界定的以及下列术语和定义适用于本文件。

3.1

人工智能 artificial intelligence

共同构建、优化和应用模型的一组方法或自动化实体，使系统能针对一系列预定义任务实现预测、推荐或决策。

3.2

机器学习 machine learning

通过计算技术优化模型参数的过程，使模型的行为反映数据或经验。

3.3

深度学习 deep learning

通过训练具有许多隐层的神经网络来创建丰富层次表示的方法。

3.4

特征工程 feature engineering

从原始数据或者预处理后的数据中提取、变换为更易解决问题的特征数据的过程，旨在去除数据中杂质和冗余量。

3.5

模型评估 model evaluation

通过既定的各类人工智能任务评估指标，对训练生成的模型进行质量评判。

4 缩略语

下列缩略语适用于本文件。

AI	人工智能	Artificial Intelligence
API	应用程序接口	Application Programming Interface
CPU	中央处理器	Central Processing Unit
GPU	图形处理单元	Graphics Processing Unit
SDK	软件开发工具包	Software Development Kit

5 概述

人工智能开发平台是指集成端到端开发和支撑工具，提供涵盖数据处理、模型构建、模型部署、支撑与服务等人工智能开发应用全流程服务的技术平台。其功能框架如图1所示，四个核心环节在总体架构中的作用如下：

数据处理部分通过数据接入、数据预处理、数据标注等功能，将数据转换为模型训练可用的内容，并在此基础上进行数据管理和分析。

模型构建部分支持模型的开发、训练和评估，通过提供一系列的开发工具，帮助各类用户快速训练出与场景匹配的模型。

模型部署部分支持模型的管理、部署和推理，保证模型可以部署在计算硬件、操作系统、推理框架等各不相同的现实环境中，最大限度发挥资源效力并赋能业务系统。

支撑与服务部分为平台提供基础的运营、维护、管理等能力，一方面提供硬件资源的管理调度，以及工作流的支持；一方面提供认证、权限、日志和监控等运维保障。

本文件中的能力要求分为基本功能和高级功能。基本功能旨在覆盖当前行业内主流人工智能开发平台的能力，聚焦平台功能的系统性、全面性；高级功能关注人工智能开发平台未来的发展方向，聚焦平台的前瞻性、引领性。根据基本功能和高级支持的情况可以评估人工智能开发平台能力水平，详情可参见附录A中的功能分级定义。

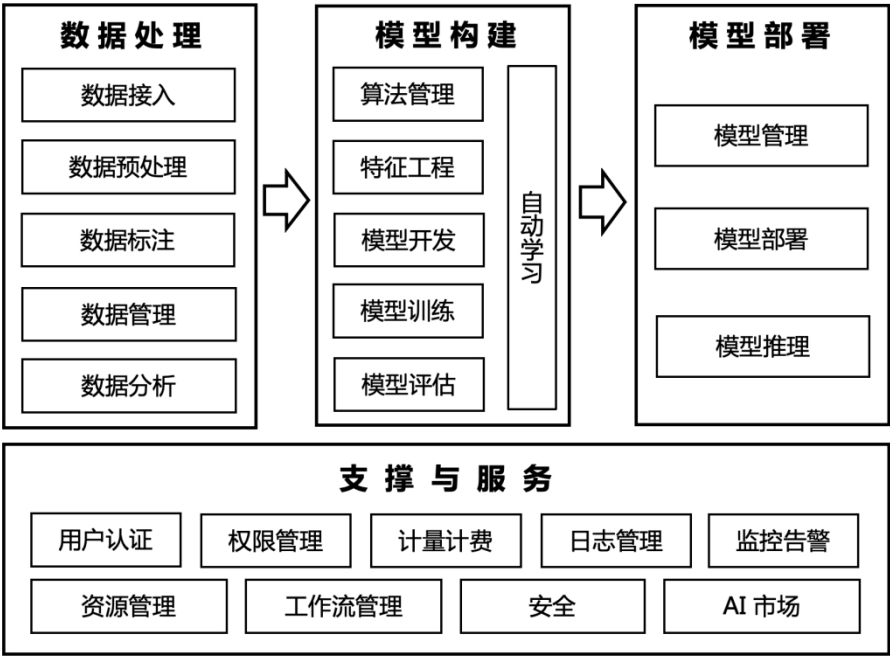


图 1 人工智能开发平台通用功能框架

6 平台通用功能指标

6.1 数据处理

6.1.1 数据接入

数据接入是人工智能开发平台的基础环节，根据项目需求，平台按照不同方式接入不同类型的数据，并在此基础上开展后续环节。该能力项可设置但不限于以下考察面：接入数据类型、接入数据方式、接入数据的参数配置等。

应包含以下基本功能：

- 支持接入结构化数据，如库表等格式；
- 支持接入多种半结构化数据，包括但不限于：csv、xls、xlsx等格式；
- 支持接入多种文本类非结构化数据，包括但不限于：txt、doc、docx等格式；
- 支持接入多种图片类非结构化数据，包括但不限于：png、jpg、jpeg、bmp等格式；
- 支持接入多种音频类非结构化数据，包括但不限于：mp3、wav等格式；
- 支持接入视频类非结构化数据，如mp4、avi、mov等格式；
- 支持接入多种标签数据，包括但不限于：JSON、XML等格式；
- 支持接入压缩包文件，如zip、rar、tar等格式；
- 支持本地文件接入数据，如本地单个上传、批量上传等；
- 支持网络协议接入数据，如FTP（File Transfer Protocol，文件传输协议）、URL（Uniform Resource Locator，统一资源定位符）、RTSP（Real Time Streaming Protocol，实时流传输协议）、共享存储等；

——支持通过用户界面配置数据接入的参数，如数据源、数据类型、接入方式等。

可包含以下高级功能：

- 支持接入实时数据流，如结构化数据流、非结构化数据流等；
- 支持周期性地数据接入，如按照设置的时间间隔定期接入数据。

6.1.2 数据预处理

经过清洗、转换等操作，数据预处理部分可以解决数据可能存在的质量问题（如不一致、无效、缺失、重复等），将数据加工为模型开发能够直接使用形式。该能力项可设置但不限于以下考察面：数据清洗、数据转换、数据增强等。

应包含以下基本功能：

- 支持结构化/半结构化数据的清洗，如数据去重、异常值检测、缺失值填充等；
- 支持非结构化数据的清洗，如根据特定规则剔除不符合要求的非结构化数据；
- 支持不同格式数据标签之间的转换，如XML标签与JSON标签之间的转换；
- 支持集成脚本语言进行数据加工，如Python、PySpark、SQL等；
- 支持自定义数据预处理功能，如用户自定义预处理算法逻辑。

可包含以下高级功能：

- 支持结构化/半结构化数据的自动预处理；
- 支持非结构化数据的自动预处理，如图像数据的自动预处理；
- 支持有监督数据增强技术，也即基于现有样本的单样本数据增强、多样本数据增强，如图像的翻转、缩放、裁剪等；
- 支持无监督数据增强技术，也即基于深度学习等算法的数据增强，如基于GAN（Generative Adversarial Networks，生成式对抗网络）的数据增强、神经风格转换等。

6.1.3 数据标注

数据标注是认知数据特征的重要过程，标注质量与模型效果息息相关，平台应提供面向不同数据的人工标注及自动标注工具，并提供可灵活扩展的团队标注模式。该能力项可设置但不限于以下考察面：标注类型、标注工具、团队标注、智能标注等。

应包含以下基本功能：

- 支持文本类数据的多种（两种或以上）标注工具或模板，如文本分类、命名实体等；
- 支持图片类数据的多种（两种或以上）标注工具或模板，如图像分类、目标检测、OCR（Optical Character Recognition，光学字符识别）等；
- 支持音频类数据的多种（两种或以上）标注工具或模板，如声音分类、语音分割等；
- 支持视频类数据的多种（两种或以上）标注工具或模板，如视频分类、目标跟踪等；
- 支持标注信息的管理，如标注标签、标注属性等的编辑、删除和查询；
- 支持可视化标注功能，标注信息可在原始数据上直观呈现；
- 支持团队标注的基本功能，如任务管理、人员管理等功能；
- 支持标注评估或质检，如评估标注的准确性、有效性等；
- 支持标注结果导出，如已标注的数据及标签导出。

可包含以下高级功能：

- 支持自定义标注功能，如基于平台提供的标注组件，配置组合为不同的标注工具，用于不同类型的标注任务；
- 支持团队标注的高级功能，如人效分析、报表导出等；
- 支持智能标注功能，如直接调用预定义算法或外部服务自动标注数据，或者在少量人工标注基础上通过训练算法进行自动标注等。

6.1.4 数据管理

数据管理是人工智能开发平台的支撑环节，平台应支持用户对其权限内的数据进行统一管理，并以数据集的形式服务于后续环节。该能力项可设置但不限于以下考察面：数据集的操作、数据集的管理等。

应包含以下基本功能：

——支持数据集的基本操作，包括但不限于：创建、修改、删除、导入、导出、发布等；
——支持数据集信息的展示和查询，包括但不限于：原始数据、数据标注信息、标签信息等内
容；

——支持数据集的管理操作，如权限管理、版本管理、标签管理等；
——支持基于数据集的二次处理，如拆分、合并、结构自定义等操作；
——支持数据集和训练结果的关联关系管理；
——支持提供 API 接口，对外提供数据，如接入线上业务系统、第三方平台等。

可包含以下高级功能：

——支持数据集的共享，如在团队内部的共享等。

6.1.5 数据分析

数据分析支持使用统计方法分析数据并提取有效信息，及时发现数据特征或分布上的问题，从而有针对性地优化处理。该能力项可设置但不限于以下考察面：数据预览、数据集分析等。

应包含以下基本功能：

——支持结构化数据的预览功能；
——支持非结构化数据的预览功能，如文本、图片、视频、音频等；
——支持数据集的分析，如统计特征、质量特征分析等；
——支持数据分析的可视化，如数据分布、标签分布等的可视化。

可包含以下高级功能：

——支持数据集探索分析形成新的数据集，如数据清洗、集合、填充、过滤等；
——支持数据集关联关系的分析和查看，如数据集和算法、模型、工作流、任务等的关联；
——支持提供数据质检功能，如对数据样本的样本数量、完整度、分布等进行检测。

6.2 模型构建

6.2.1 算法管理

算法管理提供了业内较为成熟的各类基础算法，供开发者使用，可以降低使用门槛、节省开发时间。平台可根据功能、场景等维度对算法进行分类管理，并支持开放自定义算法接口。该能力项可设置但不限于以下考察面：预置算法类型、算法操作等。

应包含以下基本功能：

——支持多种（两种或以上）传统机器学习算法，如分类、回归、聚类等；
——支持多种（两种或以上）深度学习算法，如卷积神经网络、循环神经网络等；
——支持多种（两种或以上）计算机视觉类算法，如目标检测、图像分类、文字识别等；
——支持多种（两种或以上）语音类算法，如语音识别、语音合成等；
——支持多种（两种或以上）自然语言处理类算法，如词法分析、序列标注、语义匹配等；
——自定义算法开发，如自定义名称、唯一标识、算法组件等；
——支持算法的基本操作，包括但不限于：信息查询、增加、删除、分类、检索等操作；

——支持算法的版本管理，如版本号管理、版本发布说明等。

可包含以下高级功能：

- 支持迁移学习算法；
- 支持强化学习算法；
- 支持自定义算法与预置算法的混合使用，算法间支持数据接口、模型格式接口的兼容性。

6.2.2 特征工程

特征工程提供各类基础工具，帮助用户对数据进行加工，如提取特征、分析特征、变换特征，以便更好地为模型训练做准备。该能力项可设置但不限于以下考察面：特征选择、特征分析、特征转换、数据降维、自动特征工程等。

应包含以下基本功能：

- 支持特征选择功能，如支持基于卡方检验的特征选择、基于建模的特征选择等；
- 支持特征组合功能，如将多个特征组织或衍生为新的特征等；
- 支持特征分析功能，如分析特征的重要性、权重等；
- 支持特征转换功能，如数据归一化、标准化、分箱、类型转换等方法；
- 支持数据降维功能，如主成分分析、线性判别分析等方法。

可包含以下高级功能：

- 支持自定义特征工程，如自定义新的特征工程算法；
- 支持特征分析可视化，如特征重要性指标的图表可视化等；
- 支持特征异常评估，如基于孤立森林的异常点检测、Z-score（Standard Score，标准分数）异常值检测等；
- 支持特征库的管理操作，如特征的存储、管理、分享、特征库的接入等；
- 支持多种（两种或以上）自动特征工程功能，如自动特征选择、自动特征衍生、自动特征数据增强等策略。

6.2.3 模型开发

模型开发旨在提供方便、系统、专业的开发工具和环境，通过更加人性化的操作接口展示，帮助不同层次的开发者开发算法，该部分尤其关注对主流机器学习框架的集成和支持。该能力项可设置但不限于以下考察面：开发库（AI框架等）、建模方式等。

应包含以下基本功能：

- 支持多种（两种或以上）传统机器学习框架，如Spark、Scikit-learn、XGBoost等；
- 支持多种（两种或以上）深度学习框架，如TensorFlow、PyTorch、PaddlePaddle等；
- 支持提供预训练模型，如自然语言处理、计算机视觉等领域的预训练模型；
- 支持Python等开发语言；
- 支持交互式编码建模，如Notebook等；
- 支持可视化建模，如拖拽组件等；
- 支持本地IDE开发，如通过集成SDK访问平台服务；
- 支持自定义开发环境，如以镜像方式提供可自定义的开发环境。

可包含以下高级功能：

- 支持对建模任务进行封装，如封装为SDK等形式。

6.2.4 模型训练

模型训练为用户提供多维度的训练支持，包括训练资源的调度、训练操作的支持、训练优化的支持等，平台为用户屏蔽了底层算力设施的复杂组网和配置，通过简易设置即可实现不同的训练模式。该能力项可设置但不限于以下考察面：训练类型、训练操作、训练优化等。

应包含以下基本功能：

- 支持多种 CPU 训练方式，包括但不限于：单机训练、分布式训练；
- 支持多种 GPU 训练方式，包括但不限于：单卡训练、单机多卡、多机多卡训练；
- 支持设置训练资源规格，如 CPU 核数、GPU 个数、内存等；
- 支持自定义训练参数，包括但不限于：算法参数、运行参数、训练数据、验证数据等；
- 支持训练任务的多种操作，包括但不限于：创建、查询、开启、终止、删除、修改等；
- 支持对正在执行训练任务的多种操作，包括但不限于：中止、断点恢复、断点重做等；
- 支持训练任务的信息查看，包括但不限于：训练状态、训练进度、训练结果、训练失败原因等信息；

- 支持模型的微调（fine-tune），如基于预训练模型、自动训练模型的二次训练；
- 支持模型训练过程的可视化，如训练参数和指标、模型图等可视化。

可包含以下高级功能：

- 支持异构计算资源训练，如不同架构芯片、加速卡用于异构加速训练任务；
- 提供自动调参工具，可根据模型及数据量设定合适的参数；
- 支持训练优化技术，如混合精度训练、编译优化等；
- 支持分布式计算框架的优化，如显存优化、线性加速比的提升、通信优化等。

6.2.5 模型评估

模型评估支持按照各类任务评估指标对训练完成的模型进行质量评价，帮助用户选择合适的模型，平台提供模型指标可视化工具，以图表形式呈现不同模型版本的指标对比。该能力项可设置但不限于以下考察面：评估类型、评估报告、评估可视化等。

应包含以下基本功能：

- 支持对分类模型的评估，并支持相关的评估指标，如精确率、准确率、召回率、F 函数、ROC Curve（Receiver Operating Characteristic Curve，受试者工作特征曲线）、AUC（Area Under Curve，曲线下面积）、混淆矩阵等；
- 支持对回归模型的评估，并支持相关的评估指标，如平均绝对误差、均方误差、均方根误差等；
- 支持对聚类模型的评估，并支持相关的评估指标，如轮廓系数、兰德系数等；
- 支持对序列预测模型的评估，并支持相关的评估指标；
- 支持生成评估报告，如评估结果可按照固定格式输出为报告；
- 支持模型评估报告的对比，如同一模型不同版本的报告对比；
- 支持评估指标的可视化展示。

可包含以下高级功能：

- 支持用户自定义模型评估指标，用户可以自定义评估指标及计算公式；
- 支持基于评估指标的异常样本检测，通过修正数据标签、挖掘潜在噪声样本等手段供用户优化模型；
- 支持模型的性能评估，如响应时间、业务并发等；
- 支持提供模型优化建议，如自动生成模型训练等方面的优化建议；
- 支持模型可解释性评估，如模型结构的可视化、特征重要性等。

6.2.6 自动学习

自动学习可以将特征工程、超参数调优、模型选择等环节进行自动化处理，同时通过模板化、可视化、向导式的建模工具，降低用户所需的技术门槛。该能力项可设置但不限于以下考察面：超参数搜索、模型结构设计、自动学习模板等。

应包含以下基本功能：

- 支持自动数据增强，如遗传进化、可微分等技术；
- 支持多种（两种或以上）超参数搜索技术，如网络搜索、随机搜索、贝叶斯优化等；
- 支持模型结构的自动设计，如神经网络架构搜索；
- 支持通过向导式的操作帮助用户完成模型训练，如模式定义、模型选择、参数/超参数设置等操作；
- 支持自动化建模任务的信息查看，如模型指标、资源消耗、训练时长等。

可包含以下高级功能：

- 支持小样本学习，在小样本的情况下训练出较高精度模型；
- 支持多个模型融合，如 AveragingEnsemble（Averaging Ensemble，平均法融合）、GreedyEnsemble（Greedy Ensemble，贪婪法融合）、StackingEnsemble（Averaging Ensemble，堆叠法融合）等集成方法；
- 支持多种（两种或以上）图像类自动建模场景模板，如图像分类、物体检测、图像分割、文字识别等；
- 支持多种（两种或以上）文本类自动建模场景模板，如文本分类、文本匹配、序列标注等；
- 支持多种（两种或以上）语音类自动建模场景模板，如语音识别、声音分类等；
- 支持多种（两种或以上）视频类自动建模场景模板，如目标跟踪、视频分类等；
- 在明确标注数据和业务场景的情况下，支持自动学习生成模型，自动完成特征工程、模型选择、超参数调优、模型训练等环节。

6.3 模型部署

6.3.1 模型管理

模型管理负责平台上的人工智能模型的管理，同时，为适配不同的部署和推理环境，还需提供模型转换、模型压缩等优化和适配功能，该能力项可设置但不限于以下考察面：模型文件管理，模型操作管理、模型压缩、模型转换等。

应包含以下基本功能：

- 支持模型仓库的管理和配置，包括但不限于：模型存储（如自定义模型、订阅模型）、模型版本控制等；
- 支持多种模型文件格式，如 ONNX（.onnx 等）、TensorFlow（.pb 等）、Pytorch（.pt 等）等相关的格式；
- 支持模型文件的常见操作，包括但不限于：查看、导入/导出、校验、修改、删除等；
- 支持模型文件的管理操作，包括但不限于：查询、排序、分类、展示等；
- 支持模型文件的格式转化，如以标准模型格式为中介支持各主流模型格式间的互相转换；
- 支持模型可视化能力，如模型结构、网络层级、网络权重等的可视化；
- 支持模型溯源信息，如查看模型与数据集、算法间的关系。

可包含以下高级功能：

- 支持基于任务维度的模型评估，如基于同一任务对模型进行对比；

——支持模型压缩，可根据用户定义参数完成模型压缩，如模型量化、剪枝等；
 ——支持提供模型适配云端、边缘端、终端等多种异构硬件、多种操作系统的能力，如通过 TensorRT、OpenVINO、TVM 等软件对模型进行适配优化。

6.3.2 模型部署

模型部署是指将平台上管理的模型，按照与推理环境相匹配的方式部署到指定环境中，并以指定的接口形式与其他业务应用集成。根据具体业务需求，模型可以部署在云端、边缘端、终端等不同的位置。该能力项可设置但不限于以下考察面：部署测试、部署模式、部署资源、部署管理等。

应包含以下基本功能：

- 支持用户自定义推理服务使用的资源规格，如计算资源类型、资源数量等；
- 支持以容器镜像的方式部署模型；
- 支持将模型部署为在线服务；
- 支持将模型部署为批量服务；
- 支持以 SDK 的方式部署模型；
- 支持多种模型部署及测试策略，包括但不限于：滚动更新、灰度测试、A/B 测试等；
- 支持面向业务场景的多模型编排，将多模型编排后以统一接口提供模型推理服务；
- 支持查看模型部署的信息，如部署状态、失败信息、离线日志等。

可包含以下高级功能：

- 支持推理加速框架模型的部署，如 TensorRT、openVINO、TVM 等；
- 支持边缘端、终端设备的模型部署；
- 支持面向业务场景的可视化模型编排；
- 支持模型热更新，如根据预设的条件（如模型评估结果）更新模型版本；
- 支持端云协同的服务部署管理，如云端支持对边缘端、终端设备的模型下发和更新。；
- 支持提供模型适配云端等多种异构硬件、多种操作系统的能力；
- 支持提供模型适配边缘端、终端等多种异构硬件、多种操作系统的能力。

6.3.3 模型推理

模型推理是人工智能开发平台对外提供服务的窗口，根据用户实际需求，平台分配相应的计算资源，运行指定的模型并输出预测结果。该能力项可设置但不限于以下考察面：推理框架、推理信息查询、推理操作等。

应包含以下基本功能：

- 支持深度学习推理框架，如 TensorFlow Serving 等；
- 支持模型推理服务的弹性资源调度，如弹性扩容、缩容；
- 支持模型推理服务的基本操作，如任务启动、停止等；
- 支持模型推理服务的接口信息查询及展示，如版本、实例数、接口格式等；
- 支持模型推理服务状态信息的查询及展示，如运行状态、调用量、调用成功率等；
- 支持模型推理服务的管理操作，如流量分配、服务限流、负载均衡等。

可包含以下高级功能：

- 支持异构计算资源推理，如不同架构芯片、加速卡用于异构加速推理任务；
- 支持模型推理服务的过程采样，如新样本采集、难预测样本筛选，以支持模型升级迭代。

6.4 支撑与服务

6.4.1 资源管理

资源管理支持对基础的计算、存储、运行环境等软硬件资源的管理调度，为不同的人工智能业务场景提供基础支撑。该能力项可设置但不限于以下考察面：资源类型、资源虚拟化、资源配置等。

应包含以下基本功能：

- 支持多种物理计算资源，包括但不限于：CPU、GPU 等；
- 支持 CPU 资源虚拟化；
- 支持镜像仓库，提供内置镜像、自定义镜像的存储和管理；
- 支持镜像的基本操作，包括但不限于：镜像的详情查看、导入、删除、修改、排序、查找等；
- 支持多种存储资源，如对象存储、块存储、文件存储等；
- 支持异构计算资源的管理和调度，如不同架构的芯片、加速卡用于加速计算；
- 支持资源的逻辑隔离，如计算及存储资源基于项目、用户等维度的逻辑隔离；
- 支持根据任务弹性调度分配计算、存储等资源；
- 支持设置资源调度的颗粒度，如 CPU 核数、GPU 卡数、内存数量等。

可包含如下高级功能：

- 支持 GPU 资源的虚拟化、池化等；
- 支持高速 RDMA 网络资源，如 Infiniband 网络、RoCE 网络等。

6.4.2 workflow 管理

workflow 管理支持将 AI 业务中的各类功能单元封装为一系列可独立执行的 workflow，帮助用户更好地管理开发应用流程，如更快速的测试验证、更快捷的流程复用等。该能力项可设置但不限于以下考察面：workflow 操作、workflow 编排、workflow 管理、workflow 模板等。

应包含以下基本功能：

- 支持 workflow 的基本操作，包括但不限于：创建、配置、拷贝、删除、信息查看等；
- 支持多种 workflow 编排方式，包括但不限于：可视化（图形拖拽）、代码编辑等；
- 支持 workflow 的定制化执行，如一键运行、定时执行、重复执行等；
- 支持自定义 workflow 算子，如使用 Python 等；
- 支持 workflow 执行实例的对比，如对比同一任务不同 workflow 实例中的模型性能等；
- 支持提供 workflow 模板，如典型 AI 业务流程的模板；
- 支持端到端 AI 业务流程管理，包含数据处理、模型构建、模型部署等；
- workflow 支持基于 API 的交互。

可包含以下高级功能：

- 支持多用户协同开发 workflow。

6.4.3 用户认证

用户认证部分提供统一的用户身份鉴别和用户认证管理能力，并在此基础上开展后续环节。该能力项可设置但不限于以下考察面：用户身份验证、组合认证方式等。

应包含以下基本功能：

- 支持对系统用户进行身份标识和鉴别，或支持对接第三方用户体系的接口；
- 支持服务访问请求中的认证校验，如签名信息、应用授权等；
- 支持两种及以上组合鉴别技术进行用户认证。

6.4.4 权限管理

权限管理部分的核心是管理用户与权限，避免因权限控制缺失或操作不当引发的风险问题，平台可以管理用户访问其功能、资源、数据等核心要素的权限，并应支持用户资源的隔离、权限划分等基本功能。该能力项可设置但不限于以下考察面：角色管理、权限管理、多租户管理等。

应包含以下基本功能：

- 支持角色管理，包含但不限于：角色的增加、删除、修改、查询等功能；
- 支持权限管理，包含但不限于：权限类别和操作策略的增加、删除、修改、查询等；
- 支持租户管理，如多租户的认证、权限管理；
- 支持租户间的隔离策略，如数据隔离、资源隔离、运行隔离等；
- 支持租户内部的资源共享，如数据、任务、工作流等；
- 支持对接第三方平台的权限审批系统。

6.4.5 计量计费

计量计费支持对平台资源的使用统计，并提供明确的计费标准。该能力项可设置但不限于以下考察面：资源使用统计、计费标准等。

应包含以下基本功能：

- 支持资源使用量的统计，如 CPU、GPU 等的使用数量和时长；
- 支持从不同维度进行资源使用统计，如租户、用户、任务等；
- 支持 API 调用的统计，如 API 调用次数、频率等；
- 提供明确的计费标准说明（适用于公有云企业）。

6.4.6 日志管理

日志管理对平台各个环节设置日志记录，有助于及时发现、预防故障隐患，避免系统运行事故的发生。该能力项可设置但不限于以下考察面：日志类型、日志操作等。

应包含以下基本功能：

- 支持平台多个环节的日志采集及存储，包括但不限于：数据处理部分、模型构建部分、模型部署部分、资源使用部分、用户操作部分等；
- 支持日志的基本操作，包括但不限于：日志的查询、筛选、备份等。

可包含以下高级功能：

- 支持对平台任务日志的分析。

6.4.7 监控告警

监控告警可以感知平台资源的使用情况，并对潜在的风险及时发出告警。该能力项可设置但不限于以下考察面：平台资源监控、告警策略设置、监控告警自定义等。

应包含以下基本功能：

- 支持对平台资源的监控，包括但不限于：全局基础资源、训练资源、推理资源等；
- 支持不同维度的监控，如任务、用户、租户等；
- 支持告警自定义，包括但不限于：告警规则、告警方式（邮件、短信、电话等）等的自定义。

可包含以下高级功能：

- 告警时支持提供参考意见和解决方法，方便用户快速定位和解决问题。

6.4.8 AI 市场

AI 市场主要管理人工智能开发平台中的 AI 资产，诸如模型、算法等资产内容，并为平台的使用者和开发者提供 AI 资产共享、交易的场所。该能力项可设置但不限于以下考察面：预置 AI 资产、AI 资产信息设置、AI 管理操作等。

应包含以下基本功能：

- 支持预置多种 AI 资产，如数据集、特征、算法、模型、服务、工具等；
- 支持 AI 资产的信息设置，如名称、分类、标签、详情、版本信息等；
- 支持 AI 资产的发布管理，包括但不限于：发布、编辑、删除、权限设置等；
- 支持 AI 资产发布的审核，允许平台管理员审核 AI 资产的发布与否；
- 支持 AI 资产的基本用户操作，包括但不限于：按分类或标签浏览、评论、订阅等。

可包含以下高级功能：

- 支持对 AI 资产的评价，如用户打分、评论等。

6.4.9 安全

安全部分关注跟 AI 开发平台核心业务强相关的安全内容，比如模型等相关的安全要求。该能力项可设置但不限于以下考察面：模型安全、平台安全等。

应包含以下基本功能：

- 支持数据的加密保护，如数据的加密存储、加密传输等；
- 支持模型文件的加密保护，如模型文件的加密存储、加密传输等；
- 支持模型预测服务加密能力，如模型预测服务需要获取密钥后才能正常访问；
- 支持模型的灾备迁移，如当资源异常时可将该资源上的模型迁移到其他可用资源上；
- 支持代码加密能力，可用于模型训练等环节的 Python 应用层代码加固处理；
- 平台核心组件支持完善的漏洞检测和配置合规性检查，包括但不限于：AI 框架等。

可包含以下高级功能：

- 支持模型的安全性评估，如对抗样本攻击评估等。

附 录 A
(资料性)

功能分级定义

人工智能开发平台系统功能分级定义如表A.1所示，共分为基础级、增强级、专业级、领先级、旗舰级五类，能力定义要求依次增高。功能分级的依据主要是人工智能开发平台功能的全面性、智能化、可视化，可参考功能分级的定义描述进行对标评估。

表A.1 人工智能开发平台通用功能要求分级定义

级别	名称	能力定义
1 级	基础级 Basic level	支持基础的数据标注功能，具备小规模集群环境下的模型训练和部署能力，可帮助小型团队完成开发工作。
2 级	增强级 Enhanced level	支持常见数据类型的标注和管理，具备大规模集群环境下的模型训练和部署能力，提供较为完备的资源管理、运营维护等能力，具备一定的可视化支持。
3 级	专业级 Professional level	支持常见数据类型的标注和管理，支持智能标注功能；提供多样化的模型构建方式、高性能分布式训练加速能力、一定程度的自动建模；具备较强的模型适配和部署能力，支持主流的云边端推理环境；支持弹性动态异构的资源调度能力，以及完善的平台运营维护。平台具备较高的可视化支持。
4 级	领先级 Excellent level	支持较全面的数据智能化标注和管理；提供丰富的满足多类型开发者的模型构建方式，提供高水平的自动建模，在常见场景下无需人工干预；支持常见云边端设备的协同部署和推理管理；提供精细化自动化的资源管理能力，以及智能化的运营维护。具备全面的可视化支持。
5 级	旗舰级 Ultimate level	支持高度智能化、自动化的数据标注、模型训练、模型部署、运维管理等能力，可满足用户在大部分场景下的无代码或低代码开发，具备全方位的可视化交互。