

## 1. 总体指标

(1) 训练数据生产服务平台支持云化部署，能够与华为云云平台良好适配，兼容英伟达、昇腾系列 GPU(加速卡)。

## 2. 功能指标

### 2.1 训练数据生产服务平台

(1) 构建统一对外门户和权限管理，实现各子系统单点登录，权限统一管理。

(2) 提供业务运监态势，支持集中展示系统的运行状态、组件可用率、运行时长、访问量和在线用户数等数据状态信息。单击应用服务可查看该应用的详细信息。

(3) 提供基础资源监视功能，支持对云主机、云存储、数据库、中间件、业务组件等全要素纳管，同时对其运行状态的数据和日志进行实时采集、引接和集中监控。

(4) 支持运维监控，支持集中告警和故障分析。

(5) 支持配置标注任务模板，实现标注任务的灵活编排与动态调用。

(6) 能够根据任务执行监控情况，灵活分配标注任务，支持标注工具、标注 workflow 及团队人员的自动配置、自定义配置。

(7) 支持标注任务执行状态的可视化。

(8) 提供训练数据生产服务平台全部源码。

#### 2.1.1 训练数据引接汇聚子系统

##### 2.1.1.1 数据资源框架

（1）支持数据和服务资源注册及全生命周期管理功能，包括跨域数据资源空间构建、领域内数据接入管理、跨域数据接入。

（2）提供基于元数据注册信息构建资源模型并发布的能力，支持资源模型查询、权限管理、数据访问等功能。

（3）支持数据库、主题、数据服务接口等数据访问及接入能力，提供可视化数据源配置能力。在数据接入过程中，提供校验、融合等能力。

（4）支持通过上传导入增加新的数据标准，支持对已有标准更新内容、编辑应用权限、使用状态等的的能力，支持删除过期标准等。

（5）支持使用数据校验准则对数据资源池中的数据内容进行标准符合性审核，并给出标准符合性审核结论。

#### **2.1.1.2 数据资源引接**

（1）支持数据注册功能，能够对引接数据资源的特征、数据名称、存在位置、数据描述、数据属性（数据权属人、数据的生命周期、数据权限、数据状态、数据性质、数据合法性、数据质量等）等进行注册。

（2）具有数据引接控制台功能，能够适配关系型数据库、知识库、文件、大数据环境及数据服务接口等不少于5种的接入方式。

（3）支持对引接数据源连接链路状态的监控及控制。能够监控引接数据源连接链路的协议类型（TCP/UDP）、接口状态、

来源地址、引接数据量等信息。

（4）提供可视化引接控制台界面，支持控制当前各引接数据源链路的连接状态，可由用户决定当前链路是否启用/关闭。

（5）支持在通用的主题、数据库表、接口等来源之外，定制扩展其他的数据抽取来源，如 RPC 服务、WebService 接口、特定业务服务接口等数据资源。

（6）支持对数据抽取流程进行定制，对存在抽取顺序的引接数据，用户可以通过定制不同数据类型抽取顺序，完成使用场景数据的抽取，抽取过程支持数据转换、数据融合、二进制流处理等要求。

（7）提供引接数据资源展示功能，对引接的贴源数据、标签数据、主题数据、实体知识、关系知识、事理知识、场景知识、专用模型、通用模型等数据提供展示功能，用户能够查看每一路引接数据源中包含的详细数据内容。

（8）具有在引接数据接入后，对引接数据的转换、清洗、标准化等预处理及数据缓存控制功能。

### **2.1.1.3 数据资源管理**

（1）提供数据资源存储管理和数据资源池化服务，保障资源池高质量运行和高效应用。

（2）提供适应不同模态数据的存储接口服务，包括但不限于结构化数据存储服务、半/非结构化数据存储服务、图模型数据存储服务。

(3) 提供对已授权数据资源按需查看规划能力，包括数据资源基本信息查询、数据资源内容查询和数据空间规划。

(4) 支持资源自动识别，能够基于规则 and 智能检测算法，自动扫描全域数据资源，识别重复数据资源、冲突数据资源和低效数据资源。

(5) 支持资源自动运维，能够对资源池中的资源状态定期检测，对于无法使用的资源及时告警提示，对服务启停、备份恢复等任务自动调度执行。

## **2.1.2 训练数据标注生产子系统**

### **2.1.2.1 数据解析处理**

(1) 支持从文本中提取出实体，实体要素包含但不限于武器装备、基础设施、时间、地点、国家、任务等。

(2) 支持从文本中提取出典型事件，典型事件要素抽取识别的内容包括事件信息、行动信息、以及关联数据。

(3) 支持针对文本智能组织管理需要，结合文本分类信息，构建大规模文本索引库，支持文本的高效检索。

(4) 能够对识别出的实体数据，形成实体库，并提供对实体库数据的查询、增加、删除、修改的能力。

(5) 支持识别和提取富文本文档中的文本、公式、表格、图片、公式等形式的载体要素信息，支持识别和还原复杂文档结构并生成对应文档的层次结构。

(6) 支持文档内容标注编辑，包括文档文本内容编辑、表

格编辑、公式编辑、内容解析问题反馈等。

(7) 支持通过对图片分辨率调整和数据增强操作(如旋转、缩放、裁剪、颜色变换等),对图像数据进行预处理。

(8) 支持利用多模态表征模型提取图片向量特征信息。

(9) 支持利用视觉模型进行物体检测与识别,提取图像特征。

(10) 支持利用多模态图片描述专家模型对图片内容进行识别,生成图片内容描述信息,用于图片多模态检索。

(11) 支持基于图片描述或标签信息,使用提示词和多模态专家模型生成针对图像的问答对。

(12) 支持采用数字信号处理技术,对音频语料进行降噪和音频标准化。

(13) 支持通过语音识别技术将语音内容转换为文本。

(14) 支持使用提问模型和大模型,基于音频和音频标签描述自动生成音频内容问答对。

(15) 支持通过多模态表征模型,对音频和文本问答对进行多模态向量特征提取,然后通过特征聚类、多阶段去重和难样本挖掘,获取高质量多模态音频数据。

(16) 支持通过视频预处理算法与工具能力解码视频,并分析提取关键帧集关联信息,提升视频语料数据质量。

(17) 支持视频 OCR 能力,将视频中的图片、字幕、弹幕等内容转换为文本。

（18）支持视频自动分段，提取视频关键帧，结合字幕与配音等信息，使用多模态大模型自动生成信息并进行内容总结，生成视频内容描述。

（19）支持基于提问模板和多模态大模型，基于视频和视频标签描述，自动生成视频内容问答对。

（20）支持智能擦除（去字幕、去标志、去水印）、旧视频修复（去噪、去划痕、去闪烁）、超高清增强（画质超分、插帧、HDR 等）。

（21）支持通过关键词对视频、视频描述内容进行通用检索，并对结果进行排序。

（22）提供面向结构化数据的知识抽取能力，通过对数据库的结构关系构建关联映射，实现结构化知识的抽取。

（23）提供某专业领域要素识别能力，实现某专业领域各类实体、活动识别等；提供关联关系识别，实现某专业领域实体关系、活动关系、实体与活动之间的关系识别。

（24）提供实体或活动的属性查询、关系查询、子图搜索、路径探测等服务能力。具备实体搜索能力、能够根据关键词定位实体或活动，查询其属性。

（25）具备关系搜索能力，能够根据关系类型、关系名称查询实体相关节点；具备邻居节点查询能力，能够快速查找某实体一跳或多跳关系的邻居节点。

（26）具备路径探测能力，能够查询两节点之间的关系路

径。

（27）提供时空索引创建、维护、更新和存储的功能，支撑了时空数据的快速检索的能力。

（28）提供对区域范围的划分维度制定与维护的功能，包括维度的建立、更新与删除功能，支撑了空间相关性的高效计算。

（29）提供对时间范围的划分维度制定与维护的功能，包括维度的建立、更新与删除功能，支撑了时间相关性的高效计算。

（30）提供基于区域范围划分与时间维度划分提供了时间相关性计算与空间相关性计算的功能，以及综合时间信息和空间信息关联的信息结果。

（31）通过多模态专家模型，基于时空数据进行预处理，解析识别提出时空数据的关键要素，支持对时空数据进行特征向量化。

（32）支持通过大语言模型，对时空数据内容及标签信息进行问答对的自动化生成，形成高质量的时空数据训练与评测用问答对。

#### **2.1.2.2 数据标注管理**

（1）支持自动标注，能够通过智能模型算法辅助，实现文本问答数据自动标注生成；能够学习人工标注模式、结果，基于特定任务场景，对图像、音视频等数据进行智能标注。

（2）提供编制长思维链数据功能。提供基于文档，自动批量生成高质量问答对数据能力，提供全自动、半自动、手工编制

问答对数据能力。

（3）支持多种数据格式，包括音、图、视、文的标注功能，标注模板支持在线自定义配置标签功能。

（4）支持组织标注任务功能，通过标注任务组织，可以实现人工对半自动标注结果检查和修正。

（5）支持数据导出功能，可以将标注后的数据导出到数据集模块进行集中管理。

（6）具备标注数据检验能力，包括机器检验人工标注准确性以及人工检验机器预标注准确性的能力，支持预标注结果修正功能。

（7）支持多人标注任务管理集多人协同标注，标注模块支持插件和非插件两种形态。

（8）具备标签管理能力，提供标签体系的能力建设、建模、生产评估、审核以及检索的能力，提供标签属性和标签层级的定义与维护功能，支持图像、语音、信号、气象水文等专业领域数据类型的标签，包括标签建模、标签生产任务管理、标签展示等功能。

（9）具备标签建模能力，提供标签规则库管理能力和标签模型库管理能力，支持按领域设置相应的标签规则和标签的模型，为标签的自动化生产提供规则与模型的支撑。

（10）提供了标签生产任务管理功能，支持标签任务的分配能力，包括数据的按比例切分等。



（11）提供自动打标签与手动打标签的能力，提供标签结果的评估与校正以及标签结果的审核功能。

（12）提供可视化展示标签数据的能力，提供标签的检索功能，为数据的搜索关联提供有力的支撑。

（13）支持多轮质检功能，系统支持配置多轮质检流程。质检可以将不合格的数据打回给前面节点进行返修。

（14）支持对标注任务的进度和质量进行查看和监控，支持导出统计报表。

### **2.1.2.3 知识构建融合**

（1）能够对文本、视频、图像、语音、数据库、格式报、矢量等数据进行知识抽取及关联构建，支持结构化数据的语义自动抽取，以及半结构化/非结构化数据的语义半自动抽取，支持对非结构化数据的实体抽取、属性抽取和关系抽取。

（2）具备知识语义管理能力，提供语义不同属性的定义。提供语义层级的维护以及语义属性维护的能力；提供语义树创建与编辑能力，用户能够通过图形化界面对图谱节点进行增、删、改等编辑；提供融合构建能力，对不同领域、不同业务之间的知识图谱，通过知识融合处理、语义治理的方法进行关联融合构建。

（3）提供基于文件名称、知识分类、文件类型、属性标签等的分类检索查询；提供知识属性查询能力，提供实体或活动的属性查询、关系查询、两条以上的子图搜索、节点间的路径探测等服务能力，提供模糊匹配能力。

（4）构建面向实体、词汇和意义的联合语义表示空间，支持实现实体指称的消歧，支持实体指称的链接。

（5）提供跨域多模态实体关联融合能力，基于上下文逻辑和因果关联分析的关系识别。

（6）支持知识组合的创建，通过知识组合将多个语义数据文件进行拼接，用于跨种属性标签、多种文档类型的融合的数据集创建。

（7）支持知识全生命周期管理，实现知识从生成到应用的全生命周期有效管理，形成从数据到知识、从管理到应用的完整闭环链路。

（8）支持通过知识图谱构建，经由数据抽取、清洗与关联等操作构建结构化语义网络，并依据业务逻辑与认知规律搭建知识框架，达成知识的有序管理；提供知识图谱可视化与统计分析功能，依托高效存储引擎保障知识安全存储与快速检索，打造动态更新、高效易用的知识生态系统。

（9）支持通过知识标签生成、分类汇总、社区发现以及社区摘要等步骤，实现知识从标识、聚合到要点提炼的有序管理流程。

（10）知识通过实体对齐技术精准识别并合并不同来源的同一知识实体，同时融合知识属性。

（11）支持知识加工生产，知识生成式生产借助大语言模型和深度学习算法自动产出创新、专业内容；文档知识处理通过文

档对比、勘误、审核、标记、内容查重等功能确保知识质量。

（12）支持高效知识检索，全文检索对知识全文深度解析并建立索引，实现精准内容定位；混合检索融合关键词检索、语义检索与智能推荐，提升检索的智能化与全面性水平。

（13）支持多模态知识图谱构建，涵盖知识建模、知识抽取、知识对齐、知识融合、知识推理等 5 项关键能力。

（14）支持文件知识存储、图谱知识存储、向量知识存储 3 种存储方式。

（15）具备检索增强生成（RAG）、常规知识问答、图检索增强生成（graphRAG）、多跳查询、多模态内容生成等 5 种能力。

（16）支持知识更新，具备知识版本管理、知识生命周期管控、基于用户反馈的优化等 3 项能力。

#### **2.1.2.4 数据增强扩展**

（1）能够通过基于知识图谱的语义信息拓展、基于大模型的提示生成、面向多模态数据的数据转换等方式对数据集进行增强。

（2）支持仿真样本生成，利用对抗生成网络、大模型生成等技术手段，生成多维度、多风格、多模态仿真样本数据，实现样本数据的增广，有效丰富大模型训练数据集。

（3）支持多模态数据增强，是构建跨域多模态实体关系向量化表示模型，综合不同模态实体间关联关系和属性信息，对实体语义表示向量进行维度补全，以此丰富样本数据。

(4) 具备数据描述生成能力，指结合大语言模型技术，构建与之适配的提示工程和语料生成框架，实现按需批量生成训练数据。

### **2.1.3 训练数据应用服务子系统**

#### **2.1.3.1 数据服务管理**

(1) 支持用户授权管理，提供用户对数据资源的申请访问的功能，支持管理员对数据资源的权限进行分配。

(2) 支持设置有权限使用的人员单位范围、设置数据资源的类别以及数据资源有效期设置的功能。

(3) 支持数据资源审核发布，提供数据资源的可用性校验、数据资源的符合性校验以及数据资源发布的能力。

#### **2.1.3.2 数据查询推荐**

(1) 提供对已授权数据资源按需查看规划能力，包括数据资源基本信息查询、数据资源内容查询和数据空间规划。

(2) 提供数据资源基本信息查询，包括资源名称、资源描述、资源数据格式、资源来源、资源发布人、发布机构、发布时间等，方便用户快速了解该资源数据概况。

(3) 提供数据资源内容查询能力，在资源描述信息中提供出可检索属性，用户可在基础数据访问接口中添加检索属性参数对数据资源内容按需查询。

(4) 支持针对不同用户进行授权管理，规划形成不同数据空间，根据不同用户需求，按照空间属性授予用户访问权限。

（5）支持根据用户需求，提供统一数据访问和高价值数据资源推荐能力。

（6）支持基于用户的数据需求，结合数据资源来源信息，生成获取数据资源服务接口，提供数据访问服务。

（7）提供多模态数据集检索能力，支持对各类异构数据集按领域、时间、密级、内容等进行多维检索，支持模糊匹配与查询结果的可视化展示。

#### **2.1.3.3 数据权责管理**

（1）提供数据访问权限管理功能，支持根据用户的身份特征、使用需求、审批情况和数据的业务范围、数据来源、密级等要求，对数据访问进行权限审核。

（2）提供数据维护权限功能，根据不同用户的身份级别、数据的业务范围、数据来源、密级等要求，对数据维护的权限进行审核和授权。

#### **2.1.3.4 数据质量管理**

（1）支持按照数据质量维度（包括但不限于完整性、规范性、一致性、准确性、唯一性、关联性）对数据质量进行评估。

（2）支持制定数据错误预防方案，并针对性优化当前数据问题。

（3）支持通过对数据实时监控，维护已改善的效果。

#### **2.1.3.5 数据统计分析**

（1）支持结合知识图谱、用户画像、标签等，自动化生成

训练数据相关统计数据 and 报表。

(2) 支持数据资产地图，实现数据资产统一管理，支持数据血缘追踪。

(3) 支持在资源调用管理过程中，对资源的每次调用过程生成调用链信息，覆盖数据来源、数据组装交换、数据调用方、数据响应时间、异常情况等信息。

(4) 支持对收集的调用链信息进行分析，分析其中的高频调用来源、信息依赖关系、链路环路、性能瓶颈等情况，基于分析结果，采用资源缓存、资源组合、链路重构等方式对资源调用链路进行优化。

(5) 提供反馈数据训练机制，支持用户行为采集、意见反馈等信息采集途径和反馈数据集自动化构建能力，包括数据搜索、快/慢浏览、查看、标注行为、批注、修改、重要设置等。

#### **2.1.3.6 数据服务**

(1) 支持数据目录，提供领域、对象、类型等不同维度不同层级的数据集概览能力，支持对各类数据资产信息进行统计，以柱状图、饼状图、雷达图等形式展现数据态势；可以依据数据集的分类特征通过数据目录路径快速定位所需数据集。

(2) 提供用户数据集订阅申请及权限审核的能力。提供单个数据集，或者依据领域、日期、类型等特征的批量数据集订阅服务。

(3) 能够根据订阅需求，将已经发布的数据集批量发送给

订阅用户；提供数据集的实时更新能力，能够将新更新的数据集实时同步分发。

（4）能够根据用户的查询、浏览、订阅行为自动推荐相关的配套数据集，包括基于训练阶段的推荐、基于数据配比的推荐、基于场景的推荐、基于模型规模的推荐等。

（5）能够根据用户的应用场景需求自动进行数据配比形成综合训练数据，包括领域数据与通用数据的无监督配比，面向不同能力的指令微调数据配比，及分批训练的数据配比。

（6）提供用户反馈渠道，能够进行训练效果，应用效果、数据错误等方面的服务反馈，以支持训练数据集的持续优化，提高训练数据质量。

#### **2.1.4 训练数据安全保障子系统**

##### **2.1.4.1 数据汇聚安全**

（1）提供内容安全审查能力，包括意识形态错误、价值观错误、高密级数据、攻击性提示等类型，不安全数据的检出率不低于 85%。

（2）提供数据脱敏能力，针对数据开发测试场景，对数据中含有的敏感信息进行保护，对数据进行批量化处理，按照脱敏策略、规则和算法对策略进行脱敏处理，将脱敏后数据提供至数据使用方，保证开发测试过程中敏感数据不外泄。

##### **2.1.4.2 数据存储安全**

（1）支持对数据进行存储加密和分级分类管理，对数据资

产进行分级分类识别与打标，根据数据敏感程度进行分级，按照敏感级别进行分级管理。

（2）提供数据防泄漏功能，从网络侧与终端侧对数据输出采取防泄漏措施，识别出各类敏感数据的外发行为，及时切断数据传输通道，避免敏感数据泄漏。

（3）制定完善的数据备份策略，对重要数据、关键数据进行定期备份、实时备份，重要数据库管理系统实施热备份。

#### **2.1.4.2 数据标注安全**

（1）支持留存系统操作日志，能够记录用户操作和系统活动，支持日志回溯、审计。

（2）支持数据访问权限控制，在数据标注过程中，用户只能使用经授权访问的数据集。

（3）支持设置安全性标注规则，能够指导标注人员围绕数据及生成内容的安全防线进行标注，能够给出规避安全风险的正向引导性响应信息。

（4）支持安全审核，能够对标注生产的数据进行内容安全检测。

#### **2.1.4.4 数据流转安全**

（1）在数据共享过程中，对数据产生、传输、处理等行为进行基于标签或用户身份的全链路认证与监控，记录存证，保障数据知悉范围的可控可查。

（2）提供降密脱敏措施，保障模型开发测试过程中敏感数



据不外泄。

(3) 提供数据水印功能，实现数据跨网共享的全链路可追溯。

#### **2.1.4.5 数据使用安全**

(1) 支持对应用数据资源访问进行持续验证，根据用户身份属性信息改变、访问行为异常、终端访问场景变化等安全情况进行动态授权。

(2) 数据访问控制系统与身份认证基础设施集成，实现统一用户管理与身份认证。

### **2.2 训练数据集**

(1) 引接汇聚和标注生产支持的数据类型不少于 6 类，包括但不限于文本、图像、视频、音频、信号、时空数据。

(2) 支持预训练数据集、微调训练数据集、精调训练数据集、应用数据集等类型的数据集标注生产。

(3) 训练数据集规模满足百亿级参数大模型训练需要，其中文本数据集不少于 6000 亿个词元 (Token)，评测数据集不小于 1000 个指令。

(4) 开源军事对话音频数据不少于 4000 小时，内容包括但不限于军事记录片、军事新闻报道等。

(5) 开源军事视频数据不少于 1000 小时，内容涵盖纪录片、军事报道、开源军事情报等。

(6) 提供并定期更新维护开源军事资料数据集，内容包括

但不限于智库研究报告、国际动态、国际军事发展动态、防务数据、社交媒体信息、官方媒体新闻等。

（7）针对某专业领域装备，可通过三维模型渲染方式生成近似现实场景下的装备可见光图片、视频数据，每型装备提供不少于 5 万张生成式训练样本，渲染场景至少包括森岭、平原、沙漠、城市等类型，模型可在装备机械结构展开或折叠状态下生成样本，可水平方向到垂直方向多角度模拟航空拍摄视角，可按实际需求更改装备涂装。

（8）持续提供全球实时飞机航迹数据，每日数据增量不少于 6 亿条，数据更新频率不低于 20 秒/次，包含 ADS-B、ADS-C、UAT/ADS-R、MLAT、TIS-B、Mode-S 等多种信号源，数据字段包括但不限于飞机动态航迹角、高度、速度、经纬度、机场状态、机场 ICAO 等。

（9）持续提供全球实时船舶航迹数据，每日数据增量不少于 0.6 亿条，数据更新频率不低于 5 分钟/次，包含岸基 AIS、卫星 AIS 等多种信号源，数据字段包括但不限于船舶经纬度、转向率、航向、航行状态、服务状态等。

（10）积累多模态训练数据集不少于 100TB，内容包含文本、图像、视频、音频、信号、时空数据 6 类。

（11）提供高质量 SFT 微调数据集，数量不低于 30 万对，涵盖基础知识问答、行动策略生成、指令执行、任务规划等多种类型。

(12) 提供数据标注生产服务，辅助模型训练数据集构建，开展数据生产加工培训。

(13) 建立完善的数据标注规范体系，形成标注指南或标准草案。

### 2.3 训练数据集评估验证

(1) 建立合理可行的训练数据集评估体系。设计多维度评估指标体系，包含质量、冗余度、语义一致性、多样性、时效性等核心维度，支持权重动态调整，确保评估结果客观准确，满足不同场景需求。

(2) 支持大模型评测集的构建，内容至少包括业务分类、专题分类、题目类型、题目内容（文本、图片、音频、视频等）、题目提示信息、题目答案、题目解释说明、题目难度分级、题目关联知识点。

(3) 提供基于评测集的大模型评估功能，评估指标包括但不限于精确率、召回率、F1 得分、BLEU 值、CIDER、ROUGE-L 等。

(4) 支持可视化统计训练数据集领域、来源分布，具备直观展示训练数据集多样性的能力。

(5) 支持训练数据集语种覆盖度评估。检测数据集中各语种占比及分布，支持小语种专项分析，确保多语种场景下数据均衡性，输出语种覆盖雷达图。

(6) 支持训练数据集的格式多样性评估。自动识别文本、

图片、视频、音频、遥感影像、电磁信号等数据格式，统计格式分布及合规性，检测格式错误并生成修复建议，确保多模态数据兼容性。

（7）支持训练数据集的时效性评估。分析数据时间戳分布，计算时效性评分，支持对过期数据自动标记并提醒更新。

（8）支持训练数据集标注填充率评估。检测必填标注字段的缺失率，支持输出缺失字段统计表。

（9）支持基于采样统计等方式的训练数据标注准确率评估，其中人工抽检数据条目数不少于 200 条。

（10）支持按领域选择不同的评估模板，根据领域需求，配置不同的评估项，实现领域评估的自适配性，并依据选定的评估模板，对训练数据集进行评估。

### **3. 性能指标**

#### **3.1 训练数据生产服务平台**

（1）系统能够支持 500 个并发用户同时处理标注任务，系统响应时间不超过 1 秒。

（2）系统页面的加载时间平均不超过 2 秒，如果页面中存在大量图片，千兆网络环境下每加载 50 张图片（单张图片不超过 2MB）耗时不超过 3 秒。

（3）系统支持图像、文本、音频、视频等标注任务切换，切换耗时不超过 2 秒。

（4）系统执行数据导入时，导入 1000 张图片（单张图片不

超过 2MB) 或 5 万条文本数据, 千兆网络环境下耗时不超过 3 分钟。

(5) 系统执行数据导入时对数据进行自动去重、格式转换。

(6) 系统执行数据导出时, 每 10 万条文本数据导出时间 <90 秒; 每 1000 张图片 (单张图片不超过 2MB) 导出时间 <120 秒。

(7) 提供亿级别标签数据的存储和访问的能力。

### **3.2 训练数据集**

(1) 数据智能标注准确率: 目标识别类标签准确率不低于 90%, 知识类标注准确率不低于 90%。

## **4. 配套服务要求**

### **4.1 服务团队要求**

(1) 投标供应商应针对本项目成立专门的项目组, 项目组必须由具有相关软件开发、测试、集成和实施经验的人员组成, 投标供应商需提交项目组成员名单与个人简历及相应资质证明材料, 并明确该项目组成员的职责任务情况。

(2) 中标供应商应保证关键岗位人员不能随意变更, 如因正当理由需要调整项目主要人员的, 应当提前 1 个月通知采购方, 获得书面同意后方可更换。

(3) 采购方有权对中标供应商提供的人员情况进行检查, 在发现有偏差时, 向中标供应商提出质疑, 并要求给予答复和更正。

## 4.2 售后服务要求

(1) 投标供应商须承诺提供系统平台的技术培训、技术支持和维修巡检服务，服务内容包括培训、维护、巡检、优化等，并在投标文件中提供相应书面方案。

(2) 维修巡检服务包括但不限于重保巡检、定期巡检、漏洞维护等。供应商应重视巡检过程中发现的异常情况，及时通知采购需求单位维护人员，查找原因，排除故障隐患。

(3) 供应商自终验合格之日起，即进入售后服务期。售后服务期包括质量保证期和质量维护期，售后服务期内供应商须对所提供的系统平台和为本项目所研制开发的软件、模型提供永久使用许可和免费服务，遇有严重以上安全漏洞时，须在 6 小时内修复，支持地点以最终指定的地点为准。

(4) 如系统平台发生故障，供应商维护人员应在接到采购需求单位报修后 30 分钟内响应。如需现场处理，供应商维护人员应在 3 小时内赶到现场处置。故障解决后，供应商维护人员应在 72 小时内提供完整的故障处理报告，报告内容要有针对性的预防措施。

## 4.3 培训要求

(1) 投标供应商应在应答书中提出详细的培训内容和培训计划。

(2) 投标供应商提供业务操作、运行维护等全面的技术培训。使采购方人员达到能独立进行平台管理使用、维护测试和故

障处理等工作的技术水平。

（3）培训内容应至少包括：投标供应商所提供平台软件的相关技术原理和操作使用方法，维护管理的技术，实际的操作练习。

（4）采用集中培训和现场培训结合的方式，提供人次不限的初级培训。

（5）投标供应商应根据采购方人员在系统使用过程中反馈的突出问题，及时与采购方进行充分沟通，制定针对性强、合理可行的培训计划，组织临时培训。