

团 体 标 准

T/ISC XXX—XXXX

数智化转型 企业级多模态智能数据标注 平台能力要求

Digital-intelligent transformation — capability requirements for enterprise-level
multimodal intelligent data annotation platform

（征求意见稿）

2026 – – 发布

2026 – – 实施

中 国 互 联 网 协 会 发 布

目 次

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 符号和缩略语 2

5 企业级多模态智能数据标注平台能力要求总体架构 2

6 数据接入层 3

 6.1 多源数据接入 3

 6.2 数据预处理 4

 6.3 数据安全性与存储 5

7 标注引擎层 5

 7.1 多模态标注工具集 5

 7.2 智能预标注引擎 7

 7.3 多模态对齐引擎 7

 7.4 标注规则与模板引擎 8

8 平台服务层 8

 8.1 人员权限管理 8

 8.2 流程编排管理 9

 8.3 质量管控 10

 8.4 安全合规 11

 8.5 数据资产管理 11

 8.6 系统集成 12

9 业务应用层 12

 9.1 标注工作台 12

 9.2 项目管理 13

 9.3 数据集管理 13

 9.4 质量验收 14

 9.5 统计分析 15

 9.6 模型训练支撑 15

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。

本文件由中国互联网协会归口。

本文件起草单位：山西省智慧交通实验室有限公司、中国信息通信研究院、中国电力科学研究院有限公司、中电信数智科技有限公司、天翼云科技有限公司、中国铁塔股份有限公司（南京）科技创新中心、中移（苏州）软件技术有限公司、中移互联网有限公司、中电科大数据研究院有限公司、中电云计算技术有限公司、广电运通集团股份有限公司、南京市河西新城区国有资产经营控股（集团）有限责任公司、京东集团股份有限公司、北京铜牛奇安科技有限公司、申合信科技集团有限公司、政达研学（雄安）科技有限公司、本相空间（珠海）科技有限公司、福州市润桥数字科技有限公司、数融智联（徐州）信息有限公司、浪潮通用软件有限公司、成都协鑫数智科技有限责任公司、北京融信数联科技股份有限公司。

本文件主要起草人：徐恩庆、宋佳明、王翔、李琰、姜熠龙、张巍、杨立仁、闫超众、匡磊怀、王华、张嘉熙、简泳行、杨玉春、张林梁、郭佳、张泽华、张红康、明广鸿、苏蕾权、陈凯南、常博然、陈容与、徐琳、殷辉、李腾、税洋、饶旻、张广志。

引 言

当前，我国数字经济与人工智能产业高速演进，多模态大模型、计算机视觉、自然语言处理等技术在各行各业深度落地。随着企业智能化业务规模持续扩大，跨模态、大规模、行业垂类场景的数据标注需求快速增长，企业在建设或采购多模态智能数据标注平台过程中，面临平台架构不统一、多模态标注能力边界模糊、人机协同机制缺少规范、质量管控体系参差不齐、数据安全合规要求落实不到位等问题，造成标注数据集质量不稳定、平台互联互通难、选型评估缺少统一依据，制约人工智能数据集生产效率与产业规模化发展。

本文件立足国内企业数据标注产业实践现状，依据国家相关法律法规、网络安全及人工智能相关标准规范，定义企业级多模态智能数据标注平台总体技术架构，从数据接入层、标注引擎层、平台服务层、业务应用层四个层级明确平台各项能力要求，规范多模态数据接入处理、智能标注引擎、全流程业务管理、安全合规、模型训练支撑等关键技术指标，为企业级多模态智能数据标注平台的规划设计、开发建设、选型采购、测试评估、验收交付等全生命周期活动提供统一参考。本标准适用于自动驾驶、医疗、金融、工业、安防等多行业领域企业数据标注平台建设 with 数据集生产工作，引导多模态智能数据标注平台产业规范化、高质量发展。

对本文件中的具体事项，法律法规另有规定的，遵照其规定执行。

数智化转型 企业级多模态智能数据标注平台能力要求

1 范围

本文件规定了企业级多模态智能数据标注平台的能力框架与技术要求，包括数据接入层、标注引擎层、平台服务层和业务应用层四个层级的能力要求。

本文件适用于企业级多模态智能数据标注平台的规划、设计、建设、测试与评估，可为企业数据标注平台选型、供应商评价以及行业监管提供参考依据。

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。

3 术语和定义

GB/T 37950—2019、GB/T 42755-2023、GB/T 45674-2025界定的以及下列术语和定义适用于本文件。

3.1

企业级 Enterprise-level

指信息系统具备支撑一个独立法人或非法人组织主体整体业务运行所需的部署模式、统一管理、安全治理与合规保障能力的属性。

3.2

数据标注 data labeling

给数据样本指定目标变量和赋值的过程。

3.3

标注工具 labeling tool

数据标注方执行数据标注时使用的工具，标注管理方管理数据标注时使用的工具，数据需求方验收数据标注时使用的工具等所有流程相关的工具。

3.4

数据标注规则 annotation rule

生成式人工智能模型数据标注时所遵循的方法、要求的统称。

4 符号和缩略语

下列符号和缩略语适用于本文件。

AI — 人工智能 (Artificial Intelligence)
API — 应用程序编程接口 (Application Programming Interface)
BIM — 建筑信息模型 (Building Information Modeling)
CAD — 计算机辅助设计 (Computer-Aided Design)
CDN — 内容分发网络 (Content Delivery Network)
CoAP — 受限应用协议 (Constrained Application Protocol)
COCO — 微软常见物体图像数据集 (Common Objects in Context)
COS — 云对象存储 (Cloud Object Storage)
CSV — 逗号分隔值 (Comma-Separated Values)
DICOM — 医学数字成像和通信 (Digital Imaging and Communications in Medicine)
FTP — 文件传输协议 (File Transfer Protocol)
GraphQL — 图查询语言 (Graph Query Language)

5 企业级多模态智能数据标注平台能力要求总体架构

企业级多模态智能数据标注平台整体应划分为四个技术层级，依次为：数据接入层、标注引擎层、平台服务层、业务应用层，对应数据标注的常规流程：数据接入、数据标准、服务构建、业务应用。

数据接入层应具备多类型、多来源、多部署形态数据的统一接入与预处理能力，屏蔽底层存储与格式差异，实现多模态数据的安全、高效接入与分发，为上层标注引擎提供标准化数据输入。

标注引擎层是多模态智能数据标注平台的核心技术层，以多模态标注工具集为基础，通过智能预标注引擎、多模态对齐引擎与标注规则模板引擎，实现多模态数据的高效、精准、一致性标注，为上层平台服务层提供标准化的标注能力支撑。

平台服务层应具备标注全流程的管理与服务能力，为业务应用层提供统一的人员、流程、质量、安全、数据与系统集成服务支撑，保障标注业务的高效、规范、安全运行。

业务应用层作为面向用户的能力出口，是标注平台价值的直接呈现层，提供多元化的业务应用场景支撑，覆盖标注全流程的业务操作与管理需求。

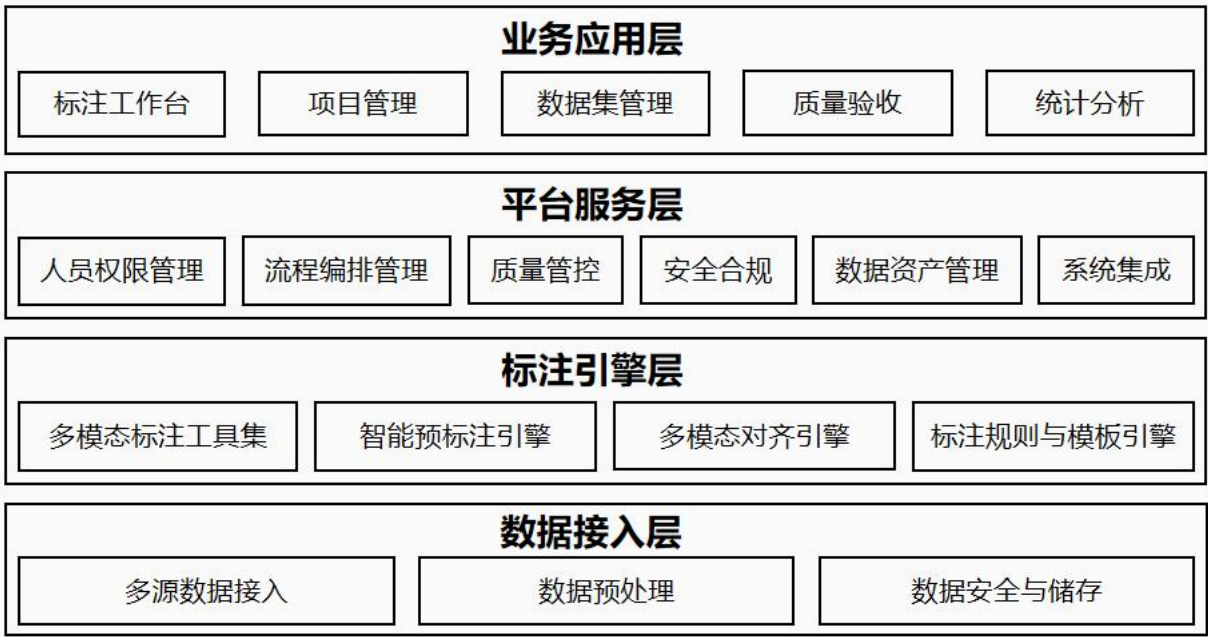


图1 企业级多模态智能数据标注平台总体架构

6 数据接入层

6.1 多源数据接入

6.1.1 多模态数据类型支持

- (1) 图像数据：应支持 JPG、PNG、BMP、TIFF、WebP 等主流图像格式，应支持超高分辨率图像、医学影像（DICOM）、遥感影像（TIFF、ENVI）等专业图像格式。
- (2) 视频数据：应支持 MP4、AVI、MOV、FLV 等主流视频格式，应支持帧提取、关键帧采样、视频片段标注等能力。
- (3) 文本数据：应支持 TXT、CSV、JSON、Word、PDF 等文本格式，应支持多语言文本、长文本、对话文本等多种文本形态。
- (4) 音频数据：应支持 WAV、MP3、FLAC、AAC 等主流音频格式，应支持长音频（时长≥2 小时）、多声道音频、语音对话等音频数据；支持音频采样率转换（8kHz-48kHz）、声道分离，满足不同模型训练的音频格式要求。
- (5) 点云数据：应支持 PCD、PLY、LAS 等点云格式，应支持 3D 点云、4D 毫米波雷达、LiDAR 点云等三维数据。
- (6) 多模态融合数据：应支持多模态数据的关联接入与打包，如“图像+文本”、“视频+音频”、“点云+图像”等组合数据。
- (7) 大模型训练专项数据：应支持指令 - 响应对、图文配对、多轮对话等生成式 AI 训练专属数据类型的打包接入。
- (8) 三维网格与模型数据：应支持 OBJ、STL、glTF、FBX 等主流三维模型格式，应支持数字孪生、工业仿真、VR/AR 等场景中的实体网格模型数据。
- (9) 高维栅格数据：应支持 NiftI、MHD/RAW、NRRD 等包含完整空间尺度信息（如层厚、像素间距、三维朝向）的多维密集数值数组。
- (10) 深度图数据：ROS bag、MCAP 等同步记录格式，应支持结构光、ToF 相机等采集的深度与彩色对齐序

列数据。

(11) 运动传感器数据：BVH、C3D 等行业运动格式，应支持 IMU、可穿戴传感器、运动捕捉等多轴运动与姿态数据。

(12) 图与知识图谱数据：应支持 RDF (Turtle/N-Triples)、GraphML、GML 等图数据格式，应支持社交网络图、药物分子图、企业知识图谱等关联数据。

(13) 工程数据：应支持 SOLIDWORKS (.sldprt/.sldasm)、Autodesk Inventor (.iam)、Revit (RVT)、IFC、DWG 等机械 CAD 与建筑 BIM 格式，应支持产品设计、数字孪生建筑等工程数据。

(14) 网络捕获数据：应支持 PCAP、PCAPNG 等网络流量捕获格式，应支持网络安全分析、协议解析等场景中的原始数据包数据。

6.1.2 多数据源接入

(1) 本地存储接入：应支持本地文件系统、内网 FTP/SFTP 服务器等本地数据源的批量接入。

(2) 云存储接入：应支持主流公有云对象存储（如 OSS、S3、COS）、私有云存储等云端数据源接入。

(3) 数据库接入：应支持关系型数据库、NoSQL 数据库、数据仓库等结构化与半结构化数据源接入。

(4) 实时数据流接入：应支持消息队列、流式数据接口等实时数据流的持续接入与处理。

(5) API 接口接入：应支持通过 RESTful API、GraphQL 等接口方式接入第三方系统数据。

(6) 数据中台/数据湖接入：应支持与企业数据中台、数据湖平台的元数据与数据链路对接。

(7) 离线介质接入：应支持移动硬盘、光盘等离线存储介质批量导入数据，适配敏感、离线内网等无网络传输场景。

(8) 物联网协议直连接入：应支持 MQTT、CoAP、Modbus TCP/RTU、OPC UA 等工业与物联网协议，应支持传感器、PLC、边缘网关等设备的直接消息与流式数据接入。

(9) 大数据存储与数据湖格式接入：应支持 Apache Hudi、Iceberg、Delta Lake 等数据湖表格式，应支持直接读取 Hive Metastore、Glue Catalog 等元数据服务中的海量结构化数据。

6.2 数据预处理

6.2.1 数据清洗

(1) 应支持数据去重、损坏数据检测与过滤、无效数据自动剔除等基础清洗能力。

(2) 应支持图像去模糊、视频抽帧去重、文本去噪（去除乱码、冗余空格）、音频降噪（去除背景杂音、电流声）等模态特定清洗能力

(3) 应支持基于规则与 AI 模型的异常数据自动识别与过滤，如识别图像中的模糊帧、文本中的敏感违规内容、音频中的无效片段等。

6.2.2 数据格式转换

(1) 应支持同模态内不同格式数据的自动转换与标准化。

(2) 应支持数据压缩、分辨率调整、采样率转换等参数标准化处理。

(3) 应支持多模态数据的格式统一与打包输出，打包格式支持 ZIP、RAR 等。

(4) 应支持标注结果在主流数据集格式间的双向转换，满足不同模型训练框架的格式要求。

6.2.3 数据增强

(1) 应支持图像翻转、裁剪、旋转、亮度调整等基础数据增强能力。

(2) 应支持文本回译、同义词替换、随机插入等文本数据增强能力。

(3) 应支持音频变速、变调、加噪等音频数据增强能力。

6.3 数据的安全与存储

6.3.1 数据安全

(1) 应支持数据传输加密，确保数据在传输过程中的安全性。建议采用 TLS 1.3 或国密 SSL 协议，实现端到端加密。

(2) 应支持数据存储加密，应支持数据存储加密，实现敏感数据的加密存储。宜支持透明数据加密（TDE）及字段级加密，兼容国密算法（SM2/SM3/SM4）。

(3) 应支持数据脱敏，对人脸、车牌、身份证号等敏感信息进行自动脱敏处理。宜同时支持静态脱敏与动态脱敏两种模式。

(4) 应支持数据水印，实现数据版权保护与溯源追踪。宜支持可见/不可见水印嵌入，并能通过水印提取实现泄露溯源。

(5) 应支持数据安全审计，记录数据访问、操作及分发日志，满足合规追溯与安全事件调查需求。

(6) 应支持数据分类分级管理，依据数据敏感程度和重要性划分安全等级，并据此实施差异化的加密、脱敏及访问控制策略

(7) 宜支持异常访问行为的自动识别与实时预警，对高风险行为自动触发告警并启动防护处置。

6.3.2 存储适配

(1) 应支持分布式存储架构，具备 PB 级数据存储能力。宜采用多副本与纠删码机制，支持在线水平扩展。

(2) 应支持冷热数据分层存储，优化存储成本与访问性能。将高频访问的标注中数据、待标注数据作为热数据，存储于高速存储介质（SSD）；将已完成标注、长期未访问的数据作为冷数据，存储于低成本存储介质（HDD）。

(3) 应支持数据缓存与 CDN 加速，提升大规模数据访问效率。宜支持边缘节点缓存及中心-边缘协同加速。

(4) 应支持数据生命周期管理，根据预设策略自动执行数据归档、迁移或过期清理，降低长期存储成本。

(5) 应支持存储容量的弹性扩缩容，可根据业务负载自动调整存储资源，避免资源浪费或瓶颈。

6.3.3 数据分发

(1) 应支持按任务、按人员、按批次的数据智能分发策略。

(2) 应支持数据的增量同步与版本管理，同步过程支持断点续传，版本更新可追溯，支持历史版本回溯，避免数据丢失或错乱。

(3) 应支持数据访问的流量控制与并发控制。

7 标注引擎层

7.1 多模态标注工具集

7.1.1 图像标注工具

(1) 2D 框标注：应支持矩形框、旋转框、多边形框等多种框选标注方式，应支持目标检测、物体识别等场景。

(2) 语义分割：应支持像素级语义分割、实例分割、全景分割等精细标注能力。

(3) 关键点标注：应支持人体关键点、面部关键点、物体关键点等关键点标注。

(4) 图像分类：应支持单标签分类、多标签分类、层级分类等图像分类标注。

- (5) 属性标注：应支持颜色、材质、状态、姿态等物体属性的标注。
- (6) OCR 标注：应支持文本检测、文本识别、表格结构识别等 OCR 相关标注。

7.1.2 视频标注工具

- (1) 目标跟踪：应支持视频序列中的目标跟踪标注，应支持多目标跟踪。
- (2) 视频分类：应支持视频级分类、片段级分类、动作分类等；支持分类标签的自定义配置，支持批量视频分类标注，支持视频片段的快速截取与标注。
- (3) 时序标注：应支持事件检测、动作定位、时序关系标注等。
- (4) 视频分割：应支持视频对象分割、视频语义分割等。
- (5) 关键帧标注：应支持关键帧提取与标注，应支持帧间插值自动标注。

7.1.3 文本标注工具

- (1) 文本分类：应支持主题分类、情感分类、意图分类等文本分类标注。
- (2) 序列标注：应支持命名实体识别、分词、词性标注等序列标注。
- (3) 关系抽取：应支持实体关系、事件关系等关系标注。
- (4) 文本摘要：应支持抽取式摘要、生成式摘要等摘要标注。
- (5) 对话标注：应支持对话状态、对话意图、对话质量等对话相关标注；支持多轮对话的关联标注，支持对话标签的自定义配置，适配智能客服、对话机器人模型训练需求。
- (6) 大模型标注：应支持 RLHF 人类反馈标注、指令微调数据标注、偏好对比标注等大模型训练数据标注。
- (7) 业务实体与关系标注：宜支持业务实体的识别与关系抽取，并支持基于标注结果触发业务流程引擎进行自动化调度。

7.1.4 音频标注工具

- (1) 语音转写：应支持语音内容转写标注，应支持多语言、多方言。
- (2) 音频分类：应支持语音分类、环境音分类、音乐分类等。
- (3) 声纹标注：应支持说话人识别、说话人分离等声纹相关标注。
- (4) 情感标注：应支持语音情感、语气、语速等声学特征标注。
- (5) 事件标注：应支持音频事件检测、时序标注等。
- (6) 业务场景音频标注：应支持业务场景音频的意图分类与关键承诺提取。

7.1.5 3D 点云标注工具

- (1) 3D 检测框：应支持 3D bounding box 标注，应支持立方体、圆柱体等多种形状。
- (2) 点云分割：应支持点云语义分割、实例分割等。
- (3) 点云分类：应支持点云级分类、物体级分类等。
- (4) 车道线标注：应支持 3D 车道线、路面标识等自动驾驶场景标注；支持车道线类型（直道、弯道、坡道等）的标注，支持路面标识（斑马线、停止线等）的关联标注。

7.1.6 多模态融合标注工具

- (1) 图文融合标注：应支持图像与文本的关联标注，如图像描述、视觉问答等；支持图像区域与文本描述的关联绑定，支持视觉问答的问题与答案标注，适配多模态大模型训练需求。
- (2) 音视频融合标注：应支持音频与视频的同步标注，如字幕生成、唇语识别等。

- (3) 点云图像融合标注：应支持 3D 点云与 2D 图像的联合标注，如多传感器融合标注。
- (4) 多模态对齐标注：应支持多模态数据在时间、空间、语义维度的对齐标注。

7.2 智能预标注引擎

7.2.1 自动预标注

- (1) 应支持基于 AI 模型的自动预标注能力，覆盖图像、视频、文本、音频、点云等各模态。
- (2) 预标注准确率应满足：通用场景目标检测 $\geq 90\%$ ，通用场景图像分类 $\geq 92\%$ ，通用场景文本实体识别 $\geq 88\%$ ，行业垂类复杂场景预标注准确率不低于 75%。
- (3) 应支持预标注结果的置信度评估，自动区分高置信度与低置信度标注结果。
- (4) 应支持批量预标注，具备大规模数据的快速预标注处理能力。

7.2.2 模型管理

- (1) 应支持内置主流预标注模型，覆盖常见标注场景。
- (2) 应支持自定义模型接入，提供标准化模型接入接口（如 TensorFlow、PyTorch 模型接口）；支持模型的上传、部署与调用，无需复杂配置，适配企业自研模型的接入需求。
- (3) 应支持模型版本管理，实现模型的迭代升级与回滚；版本更新记录可追溯，支持历史版本模型的调用与对比，避免模型升级导致预标注效果下降。
- (4) 应支持模型性能监控，包括准确率、推理速度、资源占用等指标。

7.2.3 主动学习

- (1) 应支持主动学习策略，自动选择最有价值的样本进行人工标注。
- (2) 应支持不确定性采样、多样性采样等多种主动学习策略。
- (3) 宜支持标注-训练-预标注的闭环迭代，实现模型持续优化；人工标注完成的数据可自动用于模型微调，微调后的模型重新用于预标注，形成闭环，预标注准确率随迭代次数逐步提升。
- (4) 应支持主动学习效果评估，量化标注效率提升比例。

7.2.4 人机协同

- (1) 应支持智能辅助标注，提供智能补全、智能推荐、智能纠错等辅助能力。
- (2) 应支持交互式标注，通过少量人工交互实现快速精准标注。
- (3) 应支持标注效率提升不低于 50%，复杂场景标注效率提升不低于 30%。
- (4) 支持标注经验积累与复用，将人工标注的优质案例、修正方案存入知识库，后续遇到同类标注场景可自动调用，提升标注效率与一致性；支持标注人员与智能引擎的协同反馈，人工修正的结果可用于优化智能辅助标注能力。

7.3 多模态对齐引擎

7.3.1 时空对齐

- (1) 应支持时间维度对齐，实现音频、视频、文本等时序数据的时间同步。
- (2) 应支持空间维度对齐，实现 2D 图像与 3D 点云的空间配准。
- (3) 时间对齐精度应达到毫秒级，空间对齐精度应达到像素级。
- (4) 应支持多传感器数据的时空同步标定。

7.3.2 语义对齐

- (1) 应支持跨模态语义关联，实现不同模态数据的语义映射。
- (2) 应支持多模态标注结果的语义一致性校验。
- (3) 应支持跨模态实体对齐，确保同一实体在不同模态中的标注一致性。

7.3.3 融合标注

- (1) 应支持多模态数据的联动标注，一个模态的标注自动关联到其他模态。
- (2) 应支持多视图联动，实现不同视角数据的同步浏览与标注。
- (3) 应支持融合标注结果的统一存储与管理。

7.4 标注规则与模板引擎

7.4.1 标注规则配置

- (1) 应支持可视化标注规则配置，无需编码即可定义标注规范。
- (2) 应支持标注类别、属性、关系、约束等规则元素的灵活配置；支持类别层级管理（父类、子类），支持属性的必选、可选配置，支持关系的类型与方向配置，满足不同行业场景的标注规则需求。
- (3) 应支持层级化规则体系，应支持父类子类继承、属性组等复杂规则。
- (4) 应支持规则版本管理，实现标注规则的迭代与追溯。
- (5) 支持规则的批量导入与导出，实现跨项目、跨平台规则复用。

7.4.2 标注模板

- (1) 应提供丰富的行业标注模板，覆盖自动驾驶、医疗、金融、安防等主流场景。
- (2) 应支持自定义标注模板，应支持模板的保存、复用与分享。
- (3) 应支持模板导入导出，实现跨平台模板迁移。

7.4.3 标注约束

- (1) 应支持标注约束校验，实时检查标注结果是否符合规则。
- (2) 应支持必填项校验、范围校验、逻辑校验等多种约束类型。
- (3) 应支持标注错误的实时提示与引导修正。
- (4) 支持约束校验结果的统计与分析，识别高频错误类型，优化标注规则与培训内容。
- (5) 应支持自定义约束规则，可针对特定业务场景灵活配置专属校验逻辑，满足差异化标注管理需求。
- (6) 应支持约束规则的批量启用与禁用，可根据不同标注任务的阶段要求快速调整校验策略。
- (7) 应支持约束校验日志的全链路留存，可追溯每一次标注不合规的触发场景、修改记录，为标注规则优化提供数据支撑。
- (8) 应支持多级约束优先级配置，当多个约束规则冲突时，可按照预设优先级自动处理校验结果，避免校验逻辑混乱。

8 平台服务层

8.1 人员权限管理

8.1.1 组织架构管理

- (1) 应支持多级组织架构管理，适配企业复杂组织形态。

- (2) 应支持团队、项目组、标注小组等多种团队形态；支持团队的创建与管理，可将人员分配到不同团队，支持团队间的协作与数据隔离。
- (3) 应支持人员信息管理，包括基本信息、技能标签、绩效数据、培训记录等；支持人员信息的批量导入与导出，支持人员信息的修改与归档，人员信息更新实时同步到各业务模块。
- (4) 应支持组织成员的快速调入调出、组织架构调整操作与流程审批，适配企业人员流动与组织架构动态迭代的需求。
- (5) 应支持跨组织协作场景下的成员权限隔离，可按需配置跨组织访问权限，满足多方联合标注项目的组织管理要求。
- (6) 应支持组织架构与人员信息的批量导入导出，可对接企业现有人力资源管理系统，实现存量组织信息的快速迁移接入。

8.1.2 角色权限管理

- (1) 应支持基于角色的访问控制（RBAC），提供管理员、项目经理、质检员、标注员等默认角色。
- (2) 应支持自定义角色与权限配置，实现细粒度权限控制。
- (3) 应支持数据权限隔离，确保不同人员只能访问授权范围内的数据。
- (4) 应支持操作权限控制，对标注、审核、修改、删除等操作进行权限管控。
- (5) 应支持权限变更操作的全流程留痕归档，所有权限调整记录可查询可追溯。
- (6) 应支持角色权限的批量复制与批量分配，提升多项目、大批量人员场景下的权限配置效率。

8.1.3 人员能力管理

- (1) 应支持标注人员技能标签管理，应支持按技能匹配任务。
- (2) 应支持人员绩效考核，自动统计标注量、准确率、效率、返工率等绩效指标；绩效数据实时更新，支持按日、周、月统计，支持绩效报表自动生成与导出，为人员考核与激励提供依据。
- (3) 应支持人员培训与认证管理，建立标注人员能力认证体系。
- (4) 应支持创建不同等级、不同方向的认证项目，支持上传培训课程材料、设置线上考核环节，自动记录人员的培训进度与考核结果；
- (5) 支持对认证结果进行分层分级，不同等级认证对应可承接不同精度要求的标注任务，
- (6) 支持认证有效期管理，到期自动触发重认证提醒，保障标注团队整体能力水平持续适配各类项目的标注质量要求。

8.2 流程编排管理

8.2.1 workflow 定义

- (1) 应支持可视化 workflow 编排，灵活定义标注、审核、质检、验收等环节。
- (2) 应支持串行、并行、分支、条件跳转等多种流程模式。
- (3) 应支持自定义流程节点，应支持节点属性与规则配置。
- (4) 应支持 workflow 模板管理，提供主流行业的标准化 workflow 模板，可直接调用使用。
- (5) 应支持 workflow 版本管理，支持不同版本的对比查看与回滚操作，方便流程迭代过程的追溯管理。

8.2.2 任务管理

- (1) 应支持任务创建、分配、调度、跟踪、关闭全生命周期管理。
- (2) 应支持任务批量创建与智能分配，根据人员技能与负载自动分配任务。
- (3) 应支持任务优先级管理，应支持紧急任务插队与资源调度。

- (4) 应支持任务进度实时监控，提供任务完成率、剩余量、预计完成时间等指标。
- (5) 支持任务拆分与合并，可将大型任务拆分为多个子任务，分配给不同人员并行处理；可将多个小型任务合并为一个任务，提升管理效率。
- (6) 应支持任务超时自动预警，可针对不同优先级任务自定义设置超时阈值，超时后自动通过系统消息、邮件等多渠道推送提醒给任务处理人员与项目管理者，避免项目整体交付逾期。
- (7) 应支持任务多维度筛选检索，支持按优先级、负责人、完成状态、创建时间等条件快速筛选定位目标任务，方便管理者统筹查看项目进展。
- (8) 应支持任务完结后归档管理，支持按项目、标注人员、交付时间对已完成任务分类存储，方便后续数据集溯源查询与复用。

8.2.3 流程监控

- (1) 应支持流程执行状态实时监控，可视化展示各环节流转情况。
- (2) 应支持流程异常告警，及时发现并预警流程阻塞、超时等异常。
- (3) 应支持流程效率分析，识别流程瓶颈与优化点。

8.3 质量管控

8.3.1 质检流程

- (1) 应支持多种质检模式，包括全检、抽检、交叉质检、专家质检等。
- (2) 应支持多级质检体系，实现自检、互检、专检等多层质量把控。
- (3) 应支持质检任务自动分配与质检结果自动流转。
- (4) 支持质检流程的自定义配置，可根据行业场景与质量要求调整质检模式、质检比例与质检环节；支持专家质检团队的管理，邀请行业专家参与复杂数据的质检，提升质检准确性。

8.3.2 质量指标

- (1) 应支持标注准确率、召回率、F1 值、一致性等核心质量指标自动计算；计算结果实时更新，支持按人员、任务、项目、时间等维度统计，为质量评估提供依据。
- (2) 应支持标注一致性评估，提供 Kappa 系数、一致性百分比等指标。
- (3) 应支持质量趋势分析，按时间、人员、项目等维度统计质量变化。
- (4) 采用专家基准抽检方式评估，通用场景标注准确率应 $\geq 95\%$ ，复杂场景标注准确率应 $\geq 90\%$ ，标注一致性 Kappa 系数应 ≥ 0.85 ，应支持按人员、项目、数据类型等多维度计算质量指标，支持质量趋势追踪。

8.3.3 质量问题管理

- (1) 应支持质量问题分类与标签管理，将质量问题分为标注错误、规则理解偏差、数据异常等类型，为每个问题添加对应标签，便于问题统计与分析；问题类型与标签可自定义配置，适配不同行业场景。
- (2) 应支持质量问题跟踪与闭环处理，从发现到整改形成完整闭环。
- (3) 应支持质量问题统计分析，识别高频问题与改进方向。
- (4) 支持质量问题知识库管理，将高频问题、整改方案、优化建议存入知识库，供标注人员与质检员学习参考，避免同类问题重复出现；支持质量问题反馈机制，标注人员可反馈质量问题与改进建议，提升质量管控的针对性。

8.3.4 智能质检

- (1) 宜支持基于规则的自动质检，对常见标注错误进行自动识别。
- (2) 宜支持基于 AI 模型的智能质检，自动发现标注异常与低质量标注。

(3) 宜支持质检模型的持续优化与迭代，基于人工质检结果与质量问题反馈，对质检模型进行微调，提升质检准确性与覆盖范围；支持质检模型的版本管理，实现模型的迭代升级与回滚。

8.4 安全合规

8.4.1 数据安全

- (1) 应支持数据分级分类管理，根据数据敏感级别实施差异化保护。
- (2) 应支持数据脱敏，自动识别并脱敏人脸、车牌、身份证等敏感信息。
- (3) 应支持数据防泄漏，包括屏幕水印、操作审计、下载管控等。
- (4) 应提供数据安全销毁功能，使已删除数据无法通过平台常规手段恢复。

8.4.2 系统安全

- (1) 应支持身份认证，应支持账号密码、SSO、LDAP 等多种认证方式。
- (2) 应支持访问控制，实现细粒度的功能与数据权限管控。
- (3) 应支持操作审计，记录所有用户操作日志，应支持审计追溯。
- (4) 应符合网络安全等级保护二级以上要求，涉及重要数据的应符合三级要求。

8.4.3 合规管理

- (1) 应符合数据安全、个人信息保护相关法律法规的要求。
- (2) 应支持数据合规性检查，自动识别数据合规风险。
- (3) 应支持合规报告生成，满足监管审计要求。
- (4) 应支持合规培训与管理，定期开展数据合规培训，提升用户合规意识；支持合规文档管理，存储相关法律法规、合规制度、审计报告等文档，便于查询与管理；支持合规风险评估，定期评估平台合规风险，制定优化措施，降低合规风险。
- (5) 应完整留存合规检查、风险处置、报告生成等全流程操作记录，满足监管溯源查验要求。

8.5 数据资产管理

8.5.1 数据集管理

- (1) 应支持数据集的创建、编辑、删除、版本管理全生命周期管理。
- (2) 应支持数据集分类与标签管理，应支持多维度检索。
- (3) 应支持数据集统计分析，包括数据量、标注类型、质量分布等。
- (4) 应支持数据集共享与授权管理。
- (5) 可提供共享数据集前置脱敏、溯源水印添加等安全能力，有效防范共享过程中的数据泄露风险，保障数据集共享全流程安全合规。

8.5.2 标注结果管理

- (1) 应支持标注结果的统一存储与结构化管理。
- (2) 应支持标注结果多格式导出，如 COCO、PASCAL VOC、YOLO 等主流格式。
- (3) 应支持标注结果版本管理，应支持历史版本回溯与对比。
- (4) 应支持标注结果的质量校验，可识别错标、漏标、标注不一致等异常问题，支持人工复核修正。
- (5) 应支持多维度筛选查询标注结果，可根据标注人员、标注时间、标注类型等条件快速定位目标内容。

(6) 应支持多人协同标注的结果合并与冲突处理，可对同一数据的不同标注结果进行比对，辅助解决标注分歧。

8.5.3 元数据管理

- (1) 应支持数据全链路元数据采集与管理。
- (2) 应支持数据血缘追踪，记录数据从接入到标注完成的全流程链路。
- (3) 应支持元数据搜索与发现，快速定位所需数据。
- (4) 宜支持元数据自定义配置，允许用户根据不同业务场景、不同数据类型新增、调整元数据字段与属性，满足个性化管理需求。
- (5) 宜支持元数据与原始数据、标注结果的关联映射，可通过元数据直接跳转访问对应目标数据，提升数据使用效率。
- (6) 宜支持元数据权限管控，可根据用户角色配置不同的元数据访问、修改权限，避免敏感元数据泄露。
- (7) 宜支持元数据变更记录管理，记录元数据的修改主体、修改时间与修改内容，支持变更历史回溯。

8.6 系统集成

8.6.1 API 服务

- (1) 应提供标准化 RESTful API 接口，覆盖平台核心功能。
- (2) 应支持 API 权限控制与调用审计，实现 API 调用可追溯、可管控。
- (3) 应提供 SDK 开发工具包，应支持多语言开发集成。

8.6.2 第三方系统集成

- (1) 应支持与企业 OA、HR、项目管理等系统的用户与组织同步。
- (2) 应支持与数据平台、AI 平台、模型训练平台的对接。
- (3) 应支持与存储系统、计算平台的基础设施对接。
- (4) 宜支持与数据要素流通平台对接，实现标注数据集的上架、交易与流通。

8.6.3 扩展性

- (1) 应采用微服务架构，应支持功能模块的灵活扩展。
- (2) 应支持插件化扩展机制，应支持自定义标注工具、质检规则等。
- (3) 应支持水平扩展，满足业务规模增长需求。

9 业务应用层

9.1 标注工作台

9.1.1 标注界面

- (1) 应提供直观、易用的图形化标注界面，应支持多模态数据的统一标注体验。
- (2) 应支持快捷键操作，提升标注效率。
- (3) 应支持标注界面自定义布局，适配不同用户的操作习惯。
- (4) 应支持撤销/重做、历史记录回溯等操作。
- (5) 应支持标注结果自动保存、手动保存和异常中断恢复，避免因网络或系统异常造成标注结果丢失。
- (6) 应支持已标注、未标注、待复核等不同状态数据的快速切换与定位。

9.1.2 标注辅助

- (1) 应提供标注规则实时提示与引导。
- (2) 应提供智能辅助工具，如自动吸附、智能补全、批量操作等。
- (3) 应提供标注参考示例与标准样例展示。

9.1.3 协作标注

- (1) 应支持多人协同标注，实现任务拆分与并行处理。
- (2) 应支持标注实时同步，确保多人标注结果的一致性。
- (3) 应支持标注评论与讨论，便于沟通与问题反馈。
- (4) 应支持数据编辑锁定、占用状态提示和操作冲突检测，避免多人同时编辑造成结果覆盖。
- (5) 应保留协作过程中的任务流转、标注修改、评论及问题处理记录，实现协作过程可追溯。

9.2 项目管理

9.2.1 项目创建与配置

- (1) 应支持项目基本信息配置，可编辑修改，确保信息准确可追溯。
- (2) 应支持项目标注规则与模板配置。
- (3) 应支持项目人员配置与角色分配。
- (4) 应支持项目 workflow 配置。

9.2.2 项目监控

- (1) 应提供项目进度看板，实时展示项目整体进度。
- (2) 应提供项目质量监控，实时展示项目质量指标与趋势。
- (3) 应提供项目人员绩效统计，按人员维度展示工作量与质量。
- (4) 应支持项目风险预警，及时发现进度延迟、质量下降等风险。
- (4) 应提供项目产能监控，展示日均标注量、人均产能、任务处理时长和团队负载情况。
- (5) 应提供项目人员工作情况统计，包括任务领取量、完成量、完成效率、质量结果和返工情况。

9.2.3 项目报表

- (1) 应支持项目报表自动生成，包括进度报表、质量报表、成本报表等。
- (2) 应支持多维度统计分析，按时间、人员、任务类型等维度进行分析。
- (3) 应支持报表导出与分享，可设置分享权限，保障信息安全。
- (4) 应支持报表模板配置，根据不同项目和管理角色自定义报表内容。
- (5) 应支持项目报表归档，并保留报表版本和生成记录，实现历史查询和审计追溯。
- (6) 宜支持项目人力成本、算力成本、数据存储成本自动核算报表，支撑项目预算与成本管控。

9.3 数据集管理

9.3.1 数据集浏览

- (1) 应提供数据集可视化浏览界面，应支持数据预览与标注结果查看。
- (2) 应支持多维度数据筛选与检索。
- (3) 应支持数据详情查看，包括数据属性、标注信息、质量状态等。
- (4) 应根据图像、视频、文本、音频、点云等不同数据模态提供相适配的预览方式。

- (5) 应支持图文、音视频、点云与图像等关联数据的联动浏览，展示不同模态数据之间的关联关系。
- (6) 应支持查看数据样本详情，包括数据属性、来源信息、标注内容、标注人员、审核结果、质量状态及历史处理记录等。
- (7) 应支持数据集统计信息展示，包括样本数量、数据规模、类别分布、标注进度、质量分布等。

9.3.2 数据集处理

- (1) 应支持数据集拆分，按比例拆分为训练集、验证集、测试集。
- (2) 应支持数据集合并与采样，满足测试、质检等场景需求。
- (3) 应支持数据集格式转换，导出为多种主流格式。
- (4) 应支持数据去重、异常样本剔除、低质量样本过滤和标注结果筛选等处理操作。
- (5) 应支持数据类别分布分析和样本均衡性检查，通过可视化方式体现各类别统计结果。

9.3.3 数据集版本

- (1) 应支持数据集版本管理，记录每次版本变更。
- (2) 应支持版本对比，查看不同版本之间的差异。
- (3) 应支持版本回滚。
- (4) 应记录各版本的数据来源、样本范围、标注版本、处理规则、创建人员、创建时间和变更说明等信息。
- (5) 应支持基于已有版本创建分支版本，满足不同模型、场景或实验的数据需求。

9.4 质量验收

9.4.1 验收流程

- (1) 应支持标准化验收流程，从提交验收到验收完成的全流程管理。
- (2) 应支持验收标准自定义配置，明确验收通过与驳回条件。
- (3) 应支持验收问题反馈与整改跟踪。
- (4) 应支持全量验收、随机抽样、分层抽样和风险抽样等多种验收方式。
- (5) 应支持自动生成验收样本，并记录抽样范围、抽样方法、抽样比例和样本明细。
- (6) 应支持验收任务分配，将验收任务分配给指定验收人员或专家。
- (7) 应支持验收人员查看原始数据、标注结果、审核记录、质量状态及相关标注规则。
- (8) 应支持验收结果记录，包括通过、不通过、存在争议及需整改等状态。
- (9) 应支持根据验收指标和判定规则自动计算验收结果，并允许具备权限的验收人员进行人工确认。
- (10) 应支持验收问题分类、问题描述、问题定位、责任人指派、整改时限和整改要求等信息管理。
- (11) 应支持验收问题与具体数据样本、标注对象、标注人员和标注规则关联。
- (12) 应支持验收问题整改跟踪，记录问题提交、整改、审核和关闭全过程。
- (13) 应支持整改完成后重新发起复验，并保留历次验收和复验结果。
- (14) 应支持验收异议、问题申诉和复核处理，并记录处理意见和最终结论。

9.4.2 验收报告

- (1) 应自动生成验收报告，包含质量指标、问题清单、验收结论等。
- (2) 应支持验收报告的审核与签发。
- (3) 应支持验收报告归档与查询。
- (4) 应支持验收报告关联相应的数据集版本、标注规则版本、验收标准版本和抽样样本。
- (5) 应支持验收报告导出，导出格式宜包括 PDF、Word 等通用文档格式。
- (6) 应支持验收报告访问权限控制和操作审计，保障验收信息安全。

9.5 统计分析

9.5.1 运营统计

- (1) 应提供平台整体运营数据统计，涵盖标注量、用户数、项目数等核心指标。
- (2) 应提供产能统计分析，按时间、团队、人员等维度统计产能。
- (3) 应提供效率统计分析，统计人均效率、平均标注时长等指标。
- (4) 应提供标注任务统计分析，包括任务创建量、完成量、完成率、延期情况等指标。
- (5) 应提供数据资产统计分析，包括数据规模、数据类型、数据来源、数据集版本数量等信息。
- (6) 应支持自动生成运营分析报告。

9.5.2 质量分析

- (1) 应提供整体质量趋势分析，展示质量变化趋势。
- (2) 应提供质量问题分布分析，识别主要质量问题类型。
- (3) 应提供人员质量排名与对比分析。
- (4) 应支持低质量数据样例快照图，用于直观展示标注数据质量问题。
- (5) 应支持质量问题闭环分析，跟踪问题发现、整改、验证全过程。

9.5.3 可视化看板

- (1) 应提供可视化数据看板，应支持图表化展示统计数据。
- (2) 应支持看板自定义配置，根据需求调整展示内容与布局。
- (3) 应支持数据下钻，从宏观指标下钻到明细数据。
- (4) 根据不同角色配置展示内容，包括管理员、项目负责人、标注人员等。

9.6 模型训练支撑

9.6.1 数据准备

- (1) 应支持模型训练数据的快速准备与导出。
- (2) 应支持数据增强与数据平衡处理。
- (3) 应支持特征工程辅助工具。
- (4) 应支持训练数据格式转换，满足不同模型训练框架的数据格式要求。

9.6.2 平台对接

- (1) 宜支持与模型训练平台的对接，实现标注-训练的对接。
- (2) 宜支持训练效果反馈，根据模型效果指导标注优化。
- (3) 宜支持模型迭代的数据持续供给。
- (4) 宜支持数据集权限控制与访问管理。
- (5) 宜支持模型训练平台的数据回流到标注平台，针对模型推理显示标注异常的数据集进行重新标注。

9.6.3 大模型训练支撑

- (1) 宜支持大模型预训练数据标注与清洗。
- (2) 宜支持指令微调数据标注。
- (3) 宜支持 RLHF 人类反馈数据标注，包括偏好对比、奖励建模等。
- (4) 宜支持大模型评测数据标注与构建。

- (5) 宜支持多模态大模型预训练数据全流程治理，涵盖清洗、去重、合规校验、质量分级；
 - (6) 宜支持检索增强生成（RAG）场景训练数据标注，包括文档片段相关性、问答对质量、引用准确性等维度。
 - (7) 宜支持利用大模型辅助生成训练样本，提高数据构建效率。
 - (8) 宜支持人工生成数据审核，包括真实性、一致性、有效性检查。
 - (9) 宜支持生成数据与真实数据混合管理。
-