采购包1：

标的名称：语义化机器学习模型及工具;面向数字工品的标识数据同步与监管系统

|  |  |
| --- | --- |
| **序号** | **技术参数与性能指标** |
| 1 | 一、总体功能指标 1.1、语义化机器学习模型及工具支撑软件 多源异构标识数据处理：实现数据预处理、语义理解、特征提取、模式识别及语义融合。 模型训练与评估：支持模型训练、评估、推理及预测功能。 |
| 2 | 1.2 面向数字工品的标识数据同步与监管系统支撑软件 标识注册与管理：完成多源异构标识的注册、解析及动态更新。 数据同步与分发：构建分布式解析网络，实现低延迟数据同步与跨域分发。 安全与监管功能：包含安全与隐私保护、解析监控管理、日志管理、发码及终端用户服务。 |
| 3 | 1.3核心能力要求 数据整合：消除多源数据格式及语义差异，构建统一表征。 时空语义建模：解析数据时空关联性，支持动态场景语义推理。 智能分析预测：通过机器学习挖掘潜在模式，生成可解释性预测结果。 安全管控：集成加密传输、权限分级及审计追溯，保障数据隐私与系统稳定性。 |
| 4 | 二、总体技术指标 2.1 开发语言与接口 系统采用多语言协同开发架构，核心模块遵循“算法层灵活高效、服务层稳定可靠”原则： 算法开发：Python 3.10：集成 TensorFlow 2.13 与 PyTorch 2.1 框架，支持分布式训练（Horovod）与模型压缩（TensorRT），适配自然语言处理、计算机视觉等场景。C++ 20：基于 Eclipse CDT 开发，通过 pybind11 实现与 Python 高效交互，负责底层矩阵运算（如 OpenBLAS 加速）、数据加密（AES-256）及网络通信（asio 库）。 服务端架构：Java 17 + Spring Boot 3.1：构建 RESTful API 网关，基于 Spring Cloud Stream 实现消息驱动，接口响应时间 99%分位值≤200ms。OpenAPI 3.0 规范：提供交互式接口文档（Swagger UI），支持 JSON/Protobuf 双格式，数据传输压缩率≥70%。 |
| 5 | 2.2 内存管理与计算加速 2.2.1 内存优化技术 C++ 场景：采用 Google tcmalloc 内存分配器，通过线程本地缓存（Thread-Caching）减少锁竞争，内存分配延迟降低至 10ns 级，碎片率≤5%。基于 RAII 机制管理资源生命周期，结合智能指针（std::shared\_ptr）实现零内存泄漏。 Python 场景：通过objgraph可视化对象引用链，配合weakref.finalize实现延迟清理；采用\_\_slots\_\_减少类实例内存占用，平均节省 40% 空间。 |
| 6 | 2.2.2 异构计算架构 CUDA 加速栈：支持 NVIDIA A100/H100 GPU，集成 CUDA Graphs 实现计算流程固化，推理速度提升 40%；利用 Multi-Process Service（MPS）支持多进程共享 GPU 显存。8 卡 A100 集群下，PyTorch 分布式训练（NCCL 2.18）通信带宽达 1.5TB/s，模型训练时间较无CUDA 加速栈缩短不少于60%。 CANN 加速栈：适配昇腾 910B 芯片，通过 ACL 接口调用 CANN 算子库，语义解析任务推理延迟≤45ms，单位算力功耗降低 60%。 |
| 7 | 2.3 数据库与存储架构 2.3.1 湖仓一体架构设计 数据湖层： 基于 Apache Hudi 0.13 构建增量数据湖，支持 COW（Copy-On-Write）与 MOR（Merge-On-Read）存储模式，数据压缩比达 1:8（Snappy 算法）。 通过 Trino 405 实现跨源查询，支持 Hive/HBase/MySQL 联合查询，复杂 SQL（如多表 JOIN）响应时间≤3 秒。 数据仓库层： 分析型数据库采用 ClickHouse 23.3，基于 MergeTree 引擎实现千亿级数据秒级聚合，QPS 峰值达 10 万次。 事务型数据库使用 MySQL 8.1，按业务垂直分库（标识库、质量库、产业链库），单库水平分表（哈希分表 + Range 分表），单表数据量控制在 500 万条，写入 TPS 稳定在 2500。 |
| 8 | 2.3.2 数据治理能力 元数据管理：基于 Apache Atlas 构建元数据中心，自动采集 Hive/MySQL 等数据源的表结构、字段描述，支持数据血缘可视化（DAG 图展示）。 数据质量：通过 Apache Griffin 定义规则（如唯一性校验、值域校验），缺失率≤0.1%、错误率≤0.05%，异常数据触发 Kafka 告警。 |
| 9 | 2.4 网络通信与安全 2.4.1 通信架构 内部服务：采用 gRPC 1.54 通信，基于 HTTP/2 协议实现多路复用，单连接延迟≤10ms，吞吐量达 10 万 RPC / 秒。 外部接口：RESTful API 基于 Spring Cloud Gateway 3.1 实现，支持限流（令牌桶算法，峰值 QPS≤5000）、熔断（Hystrix 兼容模式）。 |
| 10 | 2.4.2 安全防护体系 认证授权：OAuth 2.0+JWT 实现用户认证，令牌有效期≤15 分钟；RBAC 角色权限细化至接口级别，权限变更实时生效。 数据加密： 传输层：TLS 1.3 + 国密 SM4 混合加密，密钥更新周期≤7 天。 存储层：数据库透明加密（TDE）+ 字段级加密（AES-GCM），敏感数据（如企业地址）模糊化处理覆盖率 100%。 安全运营：部署 WAF 防火墙（抵御 OWASP Top 10 攻击）、IDS 入侵检测（Suricata 规则库），漏洞扫描与代码审计频率≥每月 1 次。 |
| 11 | 2.5 扩展技术支持 2.5.1 边缘计算协同 基于 KubeEdge 1.13 构建云边一体化架构，边缘节点部署轻量级容器（Runc），运行 TinyYOLOv3 模型（模型体积≤10MB），设备异常检测延迟≤500ms。 采用 Delta Sync 机制实现边缘与云端数据同步，仅传输变更数据块，带宽占用降低 80%。 |
| 12 | 2.5.2 容器化与 DevOps 容器编排：Kubernetes 1.27 集群支持自动扩缩容（HPA+VPA），Pod 启动延迟≤3 秒，服务恢复时间≤5 分钟。 镜像构建：使用 Buildah 1.29 替代 Docker，镜像体积缩小 40%，构建速度提升 30%；集成 Tekton 实现 CI/CD 流水线自动化。 |
| 13 | 三、数据中台技术要求 3.1 数据治理体系 3.1.1 元数据管理 支持自动采集 Hive、HBase、MySQL等10+数据源元数据，覆盖率 100%；通过 UI 界面可视化展示数据血缘关系（如字段来源表、加工逻辑）。 元数据更新延迟≤1小时，支持版本管理（可回溯至前30天历史版本）。 |
| 14 | 3.1.2 主数据管理 建立产品、企业、供应商三大主数据实体，定义唯一标识符（如产品ID采用UUID v5），数据一致性≥99.9%。 主数据审核流程支持多级审批（三级审批权限，如提交→初审→终审），变更记录可追溯。 |
| 15 | 3.2 数据交换规范 3.2.1 多协议支持 实时接口：Kafka 3.4 消息队列，单主题吞吐量≥10MB/s，消息延迟≤50ms；gRPC 接口支持流式传输（如实时设备数据上报）。 批量接口：FTP/SFTP 文件传输，支持断点续传，日均处理文件量≥10 万份；ETL 工具（Apache NiFi）支持复杂数据转换流程编排。 |
| 16 | 3.2.2 安全交换机制 数据沙箱基于 Apache Ranger 实现细粒度权限控制，支持按字段设置访问权限（如隐藏供应商联系方式）。 同步链路采用 “源端加密→传输加密→目标端解密” 模式，敏感数据字段加密率 100%。 |
| 17 | 四、数字工品语义化机器学习模块 4.1 多模态数据预处理 4.1.1 文本数据处理 命名实体识别（NER）：基于 spaCy 3.7 模型，准确率≥92%，支持自定义实体类型（如 “产品型号”“生产标准”）。 语义向量化：TF-IDF+Word2Vec 生成 200 维词向量，余弦相似度计算误差≤0.05，支持行业术语表（如机械制造领域词库）导入。 |
| 18 | 4.1.2 图像数据处理 目标检测：YOLOv5s 模型（640×640 输入）检测准确率≥95%，推理速度≥30FPS；支持动态阈值调整（如低光照环境提高检测阈值）。 特征提取：ResNet50 输出 2048 维特征向量，通过对比学习（Contrastive Loss）实现跨模态对齐，图像 - 文本语义匹配准确率≥90%。 |
| 19 | 4.2 模型训练与推理优化 4.2.1 训练效率提升 分布式训练：PyTorch DDP 模式支持 8 卡同步训练，通信效率≥90%；混合精度训练（FP16+BF16）减少显存占用 50%，训练速度提升 2 倍。 模型压缩：采用 TensorRT 8.6 对 PyTorch 模型进行优化，参数压缩比达 4:1，推理延迟降低 60%。 |
| 20 | 4.2.2 推理服务能力 服务部署：基于 TorchServe 实现模型服务化，支持动态负载均衡，单实例吞吐量≥1000req/s。 边缘推理：TinyBERT 模型（参数规模≤100MB）部署于边缘节点，设备日志语义解析延迟≤200ms。 |
| 21 | 五、标识数据同步与监管系统 5.1 标识全生命周期管理 5.1.1 编码规则设计 采用 “urn: [行业代码]:[体系代码]:[原标识]”三段式编码，兼容 Handle、Ecode、GS1 等标准，编码冲突率≤0.001%。 支持自定义扩展字段（如 “生产批次”“地域代码”），扩展字段长度≤32 位。 |
| 22 | 5.1.2 分布式解析架构 基于 DHT（Chord 协议）构建解析网络，节点自组织能力≥99%；Redis 7.0 缓存热点标识（TTL≤24 小时），命中率≥90%，解析延迟≤50ms。 支持批量解析（单次请求≤1000 条标识），响应时间≤200ms。 |
| 23 | 5.2 监管与可信机制 5.2.1 实时监控体系 监控指标：涵盖标识注册量（日均≥10 万条）、解析成功率（≥99%）、节点 CPU 利用率（≤80%）等 20 + 指标，通过 Prometheus 采集，Grafana 可视化延迟≤10 秒。 告警策略：支持多级告警（预警→严重→故障），通过邮件 / SMS 通知，响应时间≤30 秒。 |
| 24 | 5.2.2 区块链存证 基于 Hyperledger Fabric 2.5 构建联盟链，标识注册 / 变更记录上链延迟≤2 秒，数据不可篡改率 100%。 国密算法应用：SM2 签名算法（密钥长度 256 位）、SM3 哈希算法（输出长度 256 位），符合 GM/T 0034-2014 标准。 |

采购包2：

标的名称：数字工品质量评价数据挖掘算法与模型；基于数字标识的产业链环节评估算法与工具

|  |  |
| --- | --- |
| **序号** | **技术参数与性能指标** |
| 1 | 一、总体功能指标 1.1 采购内容 1.1.1. 数字工品质量评价数据挖掘算法与模型支撑软件 支持质量信息数据库数据字段建设。 支持质量信息数据库行业分类管理。 支持质量信息数据库数据深度挖掘。 支持质量评价成熟度因子管理。 支持质量评价报告生成。 |
| 2 | 1.1.2. 基于数字标识的产业链环节评估算法与工具支撑软件 实现产业链韧性指数高置信算法。 实现产业链韧性指数监测模块植入贯通支撑平台。 支持多条产业标识参数接入，实现多条产业链韧性指数实时显示。 |
| 3 | 1.2 主要功能或目标 1.2.1. 数字工品质量评价方面： 数据标准化：构建统一的质量信息数据库字段与行业分类体系，支持多源异构数据的整合。 深度挖掘能力：通过算法挖掘质量数据中的关键因子，动态管理质量评价成熟度模型。 智能化输出：自动生成结构化质量评价报告，提供可落地的改进建议。 |
| 4 | 1.2.2. 产业链环节评估方面： 算法精准化：构建高置信度韧性指数模型，支持多维度参数融合分析与动态优化。 平台集成化：将监测模块无缝植入现有产业链支撑平台，实现数据贯通与可视化展示。 实时响应能力：支持≥5条产业链标识参数并行接入，实现秒级数据更新与风险预警。 |
| 5 | 二、总体技术指标 2.1 开发语言与接口 系统采用多语言协同开发架构： 算法开发：Python 3.10：集成 TensorFlow 2.13 与 PyTorch 2.1 框架，支持分布式训练（Horovod）与模型压缩（TensorRT），适配自然语言处理、计算机视觉等场景。C++ 20：基于 Eclipse CDT 开发，通过 pybind11 实现与 Python 高效交互，负责底层矩阵运算（如 OpenBLAS 加速）、数据加密（AES-256）及网络通信（asio 库）。 服务端架构：Java 17 + Spring Boot 3.1：构建 RESTful API 网关，基于 Spring Cloud Stream 实现消息驱动，接口响应时间 99% 分位值≤200ms。OpenAPI 3.0 规范：提供交互式接口文档（Swagger UI），支持 JSON/Protobuf 双格式，数据传输压缩率≥70%。 |
| 6 | 2.2 内存管理与计算加速 2.2.1 内存优化技术 C++ 场景：采用 Google tcmalloc 内存分配器，通过线程本地缓存（Thread-Caching）减少锁竞争，内存分配延迟降低至 10ns 级，碎片率≤5%。基于 RAII 机制管理资源生命周期，结合智能指针（std::shared\_ptr）实现零内存泄漏。 Python 场景：通过objgraph可视化对象引用链，配合weakref.finalize实现延迟清理；采用\_\_slots\_\_减少类实例内存占用，使用内存优化技术后较优化前空间节省率≥40%。 |
| 7 | 2.2.2 异构计算架构 CUDA 加速栈：支持 NVIDIA A100/H100 GPU，集成 CUDA Graphs 实现计算流程固化，推理速度提升 40%；利用 Multi-Process Service（MPS）支持多进程共享 GPU 显存。8 卡 A100 集群下，PyTorch 分布式训练（NCCL 2.18）通信带宽达 1.5TB/s，模型训练时间缩短 60%。 CANN 加速栈：适配昇腾 910B 芯片，通过 ACL 接口调用 CANN 算子库，语义解析任务推理延迟≤45ms，单位算力功耗较未使用CANN 加速栈前降低不少于60%。 |
| 8 | 2.3 数据库与存储架构 2.3.1 湖仓一体架构设计 数据湖层： 基于 Apache Hudi 0.13 构建增量数据湖，支持 COW（Copy-On-Write）与 MOR（Merge-On-Read）存储模式，数据压缩比达 1:8（Snappy 算法）。 通过 Trino 405 实现跨源查询，支持 Hive/HBase/MySQL 联合查询，复杂 SQL（如多表 JOIN）响应时间≤3 秒。 数据仓库层： 分析型数据库采用 ClickHouse 23.3，基于 MergeTree 引擎实现千亿级数据秒级聚合，QPS 峰值达 10 万次。 事务型数据库使用 MySQL 8.1，按业务垂直分库（标识库、质量库、产业链库），单库水平分表（哈希分表 + Range 分表），单表数据量控制在 500 万条，写入 TPS 稳定在 2500。 |
| 9 | 2.3.2 数据治理能力 元数据管理：基于 Apache Atlas 构建元数据中心，自动采集 Hive/MySQL 等数据源的表结构、字段描述，支持数据血缘可视化（DAG 图展示）。 数据质量：通过 Apache Griffin 定义规则（如唯一性校验、值域校验），缺失率≤0.1%、错误率≤0.05%，异常数据触发 Kafka 告警。 |
| 10 | 2.4 网络通信与安全 2.4.1 通信架构 内部服务：采用 gRPC 1.54 通信，基于 HTTP/2 协议实现多路复用，单连接延迟≤10ms，吞吐量达 10 万 RPC / 秒。 外部接口：RESTful API 基于 Spring Cloud Gateway 3.1 实现，支持限流（令牌桶算法，峰值 QPS≤5000）、熔断（Hystrix 兼容模式）。 |
| 11 | 2.4.2 安全防护体系 认证授权：OAuth 2.0+JWT 实现用户认证，令牌有效期≤15 分钟；RBAC 角色权限细化至接口级别，权限变更实时生效。 数据加密： 传输层：TLS 1.3 + 国密 SM4 混合加密，密钥更新周期≤7 天。 存储层：数据库透明加密（TDE）+ 字段级加密（AES-GCM），敏感数据（如企业地址）模糊化处理覆盖率 100%。 安全运营：部署 WAF 防火墙（抵御 OWASP Top 10 攻击）、IDS 入侵检测（Suricata 规则库），漏洞扫描与代码审计频率≥每月 1 次。 |
| 12 | 2.5 扩展技术支持 2.5.1 边缘计算协同 基于 KubeEdge 1.13 构建云边一体化架构，边缘节点部署轻量级容器（Runc），运行 TinyYOLOv3 模型（模型体积≤10MB），设备异常检测延迟≤500ms。 采用 Delta Sync 机制实现边缘与云端数据同步，仅传输变更数据块，带宽占用降低 80%。 |
| 13 | 2.5.2 容器化与 DevOps 容器编排：Kubernetes 1.27 集群支持自动扩缩容（HPA+VPA），Pod 启动延迟≤3 秒，服务恢复时间≤5 分钟。 镜像构建：使用 Buildah 1.29 替代 Docker，镜像体积缩小 40%，构建速度提升 30%；集成 Tekton 实现 CI/CD 流水线自动化。 |
| 14 | 三、数据中台技术要求 3.1 数据治理体系 3.1.1 元数据管理 支持自动采集 Hive、HBase、MySQL 等 10 + 数据源元数据，覆盖率 100%；通过 UI 界面可视化展示数据血缘关系（如字段来源表、加工逻辑）。 元数据更新延迟≤1 小时，支持版本管理（可回溯至前 30 天历史版本）。 |
| 15 | 3.1.2主数据管理 建立产品、企业、供应商三大主数据实体，定义唯一标识符（如产品 ID 采用 UUID v5），数据一致性≥99.9%。 主数据审核流程支持多级审批（提交→初审→终审），变更记录可追溯。 |
| 16 | 3.2 数据交换规范 3.2.1 多协议支持 实时接口：Kafka 3.4 消息队列，单主题吞吐量≥10MB/s，消息延迟≤50ms；gRPC 接口支持流式传输（如实时设备数据上报）。 批量接口：FTP/SFTP 文件传输，支持断点续传，日均处理文件量≥10 万份；ETL 工具（Apache NiFi）支持复杂数据转换流程编排。 |
| 17 | 3.2.2 安全交换机制 数据沙箱基于 Apache Ranger 实现细粒度权限控制，支持按字段设置访问权限（如隐藏供应商联系方式）。 同步链路采用 “源端加密→传输加密→目标端解密” 模式，敏感数据字段加密率 100%。 |
| 18 | 四、数字工品质量评价模型 系统应支持多源数据融合挖掘，能够实时接入企业多源数据，并进行高效的清洗加载。应支持特征工程，能够自动筛选重要特征，并进行特征重要性排序。系统应构建完善的评价体系，包括定量指标和定性指标，并能通过模型进行性能评估，支持预测结果的解释。 |
| 19 | 4.1 多源数据融合挖掘 数据整合：通过 Kafka Connect 实时接入企业 ERP（如 SAP）、LIMS（实验室信息系统）数据，完整率≥95%，T+1 小时内完成清洗加载。 特征工程：LightGBM 自动筛选重要特征（如 “设备故障率”“客户投诉率”），特征重要性排序准确率≥90%。 |
| 20 | 4.2 评价体系构建 指标体系： 定量指标（功能占比70%）≥20项，主要为国家工业和信息化部14条重点产业链相关通用指标，如尺寸精度（公差 ±0.01mm）、化学成分（偏差≤0.5%）。 定性指标（功能占比 30%）≥16项：通过专家打分法评估工艺先进性、环保合规性等。 模型性能：LightGBM 模型在医疗装备数据集上 AUC≥0.95，F1-score≥0.92，支持 SHAP 值解释预测结果。 |
| 21 | 五、产业链环节评估算法 系统应支持韧性指数建模，能够从多个维度进行评估，并采用优化算法优化指数权重，支持重点产业链的实时监测。系统应支持风险传导与应对，能够构建产业链图谱，进行多级传导路径分析，并能通过算法生成风险应对策略，提供策略推荐和快速响应能力。 |
| 22 | 5.1 韧性指数建模 评估维度： 供应稳定性：指标≥6项，需包含原材料交付周期波动率（目标≤5%）、供应商集中度（CR3≤60%）。 生产协同性：指标大于等于8项，需包含订单响应速度（≤48 小时）、设备联网率（≥80%）。 算法优化：采用 PPO 强化学习算法优化指数权重，预测误差率≤5%，支持国家工业和信息化部14条重点产业链质量体系、韧性体系实时监测。 |
| 23 | 5.2 风险传导与应对 知识图谱：基于 Neo4j 5.1 构建产业链图谱，节点数≥10 万，边数≥50 万，支持 “供应商→制造商→分销商” 三级传导路径分析，预测准确率≥85%。 策略生成：通过 GNN（图神经网络）生成风险应对策略，如 “备选供应商切换”“安全库存提升”，策略推荐准确率≥80%，响应时间≤1 小时。 |