AI+合规智能中枢 — 详细技术架构图

面向车企与工厂 | 全链路合规智能平台 | Detailed Architecture

车企研发/法务 — Web门户 + API调用
工厂EHS工程师 — 移动端 + Bot通知
采购/供应链 — PLM/ERP集成
管理层/审计 — Dashboard + 报表
外部供应商 — 合规声明上传

API Gateway / nginx / Traefik — TLS终止 | 限流熔断 | 路由分发 | 负载均衡 | JWT校验 | 请求日志

kbmp-service 知识库公开接口

- **知识库CRUD** — POST /workspace/create — 创建知识库空间
- **文件上传** — POST /files/upload — 文件登记 + 任务投递
- **检索编排** — POST /knowledge/retrieval — 意图识别→召回→重排→生成
- **Chunk召回** — POST /chunks/recall — 向量+关键词混合召回
- **任务投递** — 解析/索引任务→消息队列
- **Worker入口** — worker启动/心跳/状态上报

mcp-server 文档解析服务

- **阿里云解析** — POST /parse-document — 云端高精度解析
- **MinerU解析** — POST /mineru-parse — 本地多模态解析引擎
- **OCR引擎** — 版面感知 109语言支持
- **Markdown生成** — 结构化文本输出
- **表格/图片提取** — PDF/Word/Excel多格式
- **解析回退策略** — 阿里云→MinerU→本地Fallback

合规业务后端 法规 + 审查 + 推送

- **法规下载** — POST /compliance/regulations/download — 从互联网下载法规文档
- **法规更新/同步** — POST /compliance/regulations/update — 版本管理+增量索引同步
- **权限分级管理** — POST /compliance/access-control — 研发/生产/采购/法务四角色
- **智能合规审查** — POST /compliance/check — 条款级比对+风险评分
- **合规结果查询** — GET /compliance/query — 审查结果+风险项+整改建议
- **事件订阅推送** — POST /compliance/subscribe — Webhook+多渠道推送

法规感知引擎 Regulation Awareness Engine

- **法规源监控** — 定时爬取国标网/工信部/UN-ECE EUR-Lex/碳交易/行业通报
- **智能变更感知** — NLP比对新旧版本Diff 自动识别新增/修订/废止条款
- **自动解析入库** — 触发MinerU解析→条款分块 →BGE-M3嵌入→Milvus+PG写入
- **知识图谱同步** — Neo4j关系更新 条款义务映射+影响范围分析
- **差距分析** — AI对比企业制度与新法差距 自动生成变更影响评估
- **变更推送 & 整改触发** — 按角色/域精准推送摘要 自动创建整改任务→闭环跟踪

AI推理引擎 RAG + LLM + 图谱

- **混合检索** — BM25关键词 + BGE-M3向量 本地+网络双路召回
- **BGE-M3嵌入** — 中英双语 8192 tokens Dense+Sparse+Multi-vec
- **Reranker精排** — Cross-Encoder语义精排 Top-K结果重排序
- **LLM生成** — DeepSeek/Qwen2.5 引文锚定+置信度评分
- **知识图谱** — Neo4j法规实体关系图 多跳推理+条款关联
- **NLP分析** — 实体识别/文档分类 隐患实体抽取

Worker集群 异步任务执行

- **解析Worker** — 消费解析任务→调用mcp-server
- **向量化Worker** — 文本清洗→切分→嵌入→入库
- **合规Worker** — 比对法规→风险评分→报告
- **感知Worker** — 法规变更检测→增量重索引
- **推送Worker** — 消息分发→Email/Bot/站内
- **调度框架** — Celery + Cron定时 失败重试+死信队列

服务 ↔ 中间件 双向通信

- **Milvus** — 向量数据库 Dense+Sparse+Hybrid
- **PostgreSQL** — 关系数据库 元数据/权限/任务
- **S3 / MinIO** — 对象存储 原始文件/解析产物
- **Neo4j** — 图数据库 法规实体关系图谱
- **RabbitMQ** — 消息队列 异步任务分发
- **Redis 7.x** — 缓存/会话 热数据/分布式锁

中间件 ↔ AI模型 调用链路

嵌入模型

- **BGE-M3 (主模型)** — 中英双语 100+语言
- **bge-large-zh-v1.5** — 中文专项嵌入
- **多语言E5** — 跨语言检索备选
- **8192 token上下文** — 长文档向量化支持

LLM大模型

- **DeepSeek-V3 / R1** — 推理能力强, 国产开源
- **Qwen2.5-72B** — 中英双语, 合规场景优化
- **本地私有化部署** — vLLM/TGI推理加速
- **引文锚定生成** — 输出含原文出处+页码

文档解析模型

- **MinerU (多模态)** — 版面感知PDF解析
- **阿里云文档解析** — 云端高精度解析
- **版面感知OCR** — 109语言扫描件识别
- **表格/图片识别** — 复杂版面结构提取

专项模型

- **Cross-Encoder** — Reranker语义精排
- **NLP实体抽取** — 法规条款/隐患实体
- **SIF风险评分** — 高严重性事件潜力预测
- **合规分类器** — 法规域/文档类型分类

AI模型 ← 法规数据源 学习与检索

- **车辆安全法规** — GB 7258 · GB 18384 · UN-ECE R155/156
- **数据安全法规** — PIPL · DSL · GDPR · GB/T 35273
- **工厂EHS法规** — GB 6441 · AQ/T系列 · ISO 45001 · IATF 16949
- **碳排放法规** — NEV积分 · CCER · CBAM碳边境税
- **企业内部文档** — Confluence · SharePoint · 历史报告 · 审计记录
- **行业案例库** — 处罚案例 · 事故通报 · 整改最佳实践

核心数据流 (Data Flows)

1. **法规感知闭环:** 定时爬取法规源 → NLP变更感知(Diff) → 自动解析入库(MinerU+嵌入) → Milvus+PostgreSQL+Neo4j同步 → 差距分析 → 按角色推送 → 触发整改 ↺ 持续监控
2. **上传→解析→入库:** 用户上传 → API Gateway → kbmp-service → 队列 → Worker → mcp-server解析 → 文本切分 → BGE-M3嵌入 → Milvus+PostgreSQL写入
3. **检索→问答:** 用户提问 → 意图识别 → BM25+向量双路召回 → Cross-Encoder精排 → LLM生成(引文锚定) → 返回结果
4. **合规审查:** 文件上传 → OCR解析 → 条款级分块 → 法规域匹配 → 语义比对 → 风险评分 → 整改建议 → 报告生成
5. **EHS隐患:** 巡检文本NLP → 隐患实体抽取 → SIF风险评分 → 四维根因分析 → 整改工单 → 验收关闭 → 模型优化

安全与治理

- Token鉴权 (JWT/OAuth2)
- RBAC角色权限矩阵
- 知识库分区隔离
- 敏感数据脱敏
- 全链路审计日志
- PIPL/DSL数据主权
- 私有化部署 (数据不出厂)
- WAF & DDoS防护
- TLS端到端加密

运维观测

- Prometheus指标采集
- Grafana可视化面板
- Loki/ELK日志聚合
- 分布式链路追踪
- 告警规则引擎
- SLA可用性监控
- 性能基线管理
- 容量规划报表

基础设施

- Kubernetes容器编排
- Docker容器运行时
- GPU集群 (A100/H100)
- vLLM/TGI推理加速
- CI/CD流水线
- Nginx/Traefik网关
- VPN & 网络隔离
- 数据备份 & 灾备

核心开源技术栈

- **LangChain** RAG编排框架
- **LlamaIndex** 数据索引引擎
- **RAGFlow** 文档理解
- **BGE-M3** 多语言嵌入
- **MinerU** 文档解析OCR
- **Milvus** 向量数据库
- **Neo4j** 知识图谱
- **FastAPI** API框架
- **Celery** 异步任务队列
- **DeepSeek** 推理LLM
- **Qwen2.5** 双语LLM
- **PyMuPDF** PDF处理

法规感知自动更新闭环

- ① **法规源监控**
- 定时爬取国标网 · 工信部 · UN-ECE
- ② **智能变更感知**
- NLP Diff · 新增/修订/废止识别
- ③ **自动解析入库**
- 解析→分块→嵌入→Milvus+PG
- ④ **知识图谱同步**
- Neo4j关系更新 · 影响范围分析
- ⑤ **差距分析&推送**
- AI比对制度差距 · 按角色推送
- ⑥ **触发整改闭环** ↺
- 自动创建整改任务 · 闭环跟踪

三类合规闭环

- **法规变更闭环**
- 监控→感知→更新→推送→整改→归档
- **文档审查闭环**
- 上传→解析→比对→标注→整改→复审
- **EHS安全闭环**
- 发现→评级→派发→跟踪→验收→优化