

个人简介

10年+全栈研发与架构经验，前7年深耕**Java企业级架构**（Spring Cloud 微服务、高并发交易、金融安全），如主导青岛市公交的公交调度，gis算法，乘车码支付码等核心算法与安全和部分运维。近3年转型**AI Agent 系统与 Neuro-Symbolic 混合架构**，交付了多个大型复杂系统，研发产品上过央视新闻，世界互联网大会等。

核心优势：具备从产品需求分析→架构设计→多语言开发（Java/Python/Rust/Go/TypeScript）→AI大模型与Agent集成→前端可视化→CI/CD部署运维的**全流程独立交付能力**，单人可完成传统5-8人团队的工程产出。

核心能力标签：AI大模型 & Multi-Agent系统 · Neuro-Symbolic混合架构 · Java企业级架构（10年）
· Rust高性能引擎 · 计算机视觉边缘部署 · 运筹优化算法 · 全栈全流程交付

核心技术能力

AI大模型与Agent系统 精通LangGraph、CrewAI、MetaGPT、Google ADK等主流Agent编排框架，具备Neuro-Symbolic（神经符号）混合架构设计能力——将LLM语义理解与符号推理（HTN/知识图谱/规则引擎）深度融合，解决大模型幻觉与不可控问题。深入掌握RAG/Agentic RAG/GraphRAG企业级落地方案，熟练运用Semantic Router实现意图分发。具备LORA微调Llama/GPT模型的实战经验。熟练使用Claude Code、Codex CLI、Cursor等AI开发工具构建生产级系统。

Java企业级架构（10年） 精通Spring Cloud Alibaba全家桶（Nacos/Sentinel/Gateway/Seata），具备大型微服务架构设计与治理经验（百万级用户、7,000+车辆调度）。精通基于Netty的高性能网络编程，深入理解DDD领域驱动设计。具备金融级安全系统设计能力（国密SM2/SM3/SM4、双离线支付、API动态防御网关），通过国家保密局安全审计。熟练掌握Kafka/Flink实时数据流水线、TiDB分布式数据库、Redis Cluster。

Python AI工程化 精通FastAPI高性能框架，具备20+Agent协同集群、7阶段情报分析管线、12步战术决策状态机等复杂AI系统的架构与实现经验。熟练运用Instructor结构化输出、Pydantic数据建模、asyncio异步编程。

Rust高性能系统 具备Rust生产级开发经验：自研混合A*全地形路径规划引擎（7,200+行，PyO3/Maturin桥接Python）、符号推理Sidecar微服务（Axum/Tokio，JSON Logic规则引擎+多维风险评分）。熟练掌握Rust异步编程与FFI跨语言调用。

Go微服务开发 具备Go生产级开发经验：自研设备控制平台（DDD分层架构，设备接入/遥测监控/指令下发），熟练掌握Go并发编程与微服务开发。

TypeScript全栈开发 具备TypeScript前后端开发经验：应急指挥系统前端（React+Cesium三维数字孪生+WebSocket实时通信，4.3万+行）、情报分析前端（React+态势可视化）。

计算机视觉与边缘AI 自研edge-vision边缘视觉引擎，支持YOLOv11/SAM3/InsightFace/PaddleOCR多模型协同推理。五种推理后端自适配（ONNX Runtime/TensorRT/OpenVINO/PaddlePaddle/PyTorch），具备Jetson Orin NX边缘部署与模型量化（FP16/INT8）实战经验。

运筹优化算法 自主研发Offroad A*（Rust，全地形越野路径规划）、MCTS（蒙特卡洛树搜索）、NSGA-II/III多目标优化、VRP/VRPTW（OR-Tools）、CSP能力-资源匹配（CP-SAT）、离散事件仿真等算法引擎。

前端与可视化 熟练掌握React/Vue/Cesium/Mars3D三维数字孪生/ECharts，具备Android原生（Kotlin）与Uni-app跨端开发能力。

DevOps 与全流程交付 熟练运用 Docker / Kubernetes 进行微服务编排与模型服务部署，具备完整 CI/CD 流水线搭建经验。从需求分析、产品设计、架构设计、多语言开发、测试到部署运维，可一人完成全流程。

工作经历

杭州某公司 | AI 全栈架构师 & 核心算法研发

2025.06 - 至今

负责公司全部 AI 项目的架构设计与核心开发，涵盖防务情报、战术指挥、边海防、无人机四大业务线。

项目一：AEGI – 国际防务情报分析系统（已完成 50%+）

AI 驱动的 OSINT（开源情报）分析平台，实现多语言情报采集 → 结构化分析 → 假设验证 → 因果推理 → 态势预测的全链路自动化。

架构： FastAPI + Neo4j 知识图谱 + Qdrant 向量库 + MinIO + SearXNG 搜索聚合 + MCP 工具网关

核心技术亮点： - 7 阶段端到端分析管线：断言融合 → ACH 假设分析 → 对抗性评估（Defense/Prosecution/Judge 三角对抗）→ 叙事构建 → 知识图谱构建 → 多情景预测 → 元认知质量评分，每阶段可独立跳过/降级 - **LLM 治理框架（Grounding Gate）**：所有 AI 输出强制接地验证——有证据引用标记为 FACT，无证据强制降级为 HYPOTHESIS，杜绝情报分析中的 AI 幻觉；全链路审计追踪 - **知识融合引擎**：多语言声明提取（中/英/俄/阿）→ 实体消歧 → 跨语言实体对齐（规则候选 + LLM Rerank）→ 本体版本管理 - **因果推理**：时序一致性检查 → 因果链评分 → 反事实分析 → 信息协调检测 → 偏见/盲点分析

工程实践： OpenSpec 规范驱动架构（proposal → design → spec → tasks 四阶段文档），完整测试体系（单元/集成/契约/E2E/回归指标）

技术栈： Python, FastAPI, Neo4j, Qdrant, MinIO, SearXNG, MCP Protocol, Instructor, OpenClaw

项目二：某车企 AI 战术指挥大脑（进行中）

面向特种作战场景的车载战术指挥系统，在已交付的 AI 应急指挥大脑基础上全面升级。

架构创新 – 大模型 + 垂直小模型 + Rust 符号层： - **大模型层**：自然语言理解与战场态势语义分析 - **垂直小模型层**：SAM3 像素级分割 + YOLOv11 实时目标检测，专注垂直领域、推理快、资源低 - **Rust 符号层 9 步决策引擎**：意图解析 → 知识图谱校验 → HTN 任务分解 → 约束求解 → NSGA-II 多目标优化 → 资源分配 → 路径规划 → 冲突消解 → 行动序列输出，决策幻觉率 <1%、毫秒级推理

Multi-Agent 集群（LangGraph + CrewAI）：10+ 专业 Agent 分层协作——VoiceCommander（Hybrid Routing 意图分流）、StagingArea Agent（GIS 动态选址）、ReconAgent（MCTS 无人机探索规划）等

技术栈： LangGraph, CrewAI, Neo4j, Qdrant, Rust, SAM3, YOLOv11, NSGA-II, MCTS, FastAPI, PostGIS, Cesium

项目三：无人机飞控平台 + AI 识别算法（进行中）

无人机集群飞控管理与 AI 视觉识别一体化平台。

- 飞控管理：** 多机协同任务规划、实时状态监控（WebSocket 推送）、Cesium 三维航迹展示
- AI 识别算法矩阵：** 航拍目标检测（小目标/遮挡/尺度变化优化）、工业巡检（安全帽/烟火/入侵检测）、边海防多目标跟踪
- MCTS 自主探索：** 蒙特卡洛树搜索实现未知区域覆盖路径规划，结合 DEM 地形避障与能耗优化
- 端侧部署：** ONNX → TensorRT FP16 量化，Jetson Orin NX 推理延迟 <30ms/帧

技术栈： YOLOv11, TensorRT, MCTS, FastAPI, Cesium, PostGIS, Jetson Orin NX

项目四：边海防 AI 算法识别 & AI 辅助作战参谋（已交付）

包含两个独立子系统：**edge-vision 边缘视觉引擎**（端侧推理）和 **sentinel 战术参谋**（AI 辅助决策），个人独立完成。

子系统一：edge-vision 边缘视觉引擎

部署于 Jetson Orin NX (100 TOPS INT8)，实现端侧四模型级联实时推理。

- **推理管线：** YOLOv11 + ByteTrack 人车检测跟踪 → InsightFace 人脸识别 (ArcFace 512维嵌入) → PaddleOCR v4 车牌识别 (TensorRT FP16) → 车辆属性识别 (HSV 主色提取 + 车型分类)
- **五种推理后端自适配：** 自动检测硬件平台 (Jetson ARM / x86 CUDA / Intel iGPU) 选择最优后端
- **核心算法：** 多帧投票车牌聚合、套牌车检测 (指纹比对 + ReID 余弦距离 + 三级告警)、可插拔告警规则引擎 (5 类规则 / 8+ 种告警类型)
- **工程设计：** 异步多进程架构 (主进程检测 + 子进程识别，非阻塞推理)、环形帧缓冲证据录制、GStreamer RTMP/RTSP 推流

子系统二：sentinel 战术参谋系统

Neuro-Symbolic 混合架构：**LangGraph 12 步状态机 + Rust 符号推理 Sidecar** (Axum 微服务)，Python + Rust 混合 Monorepo。

- **LangGraph 12 步编排：** Ingest → Normalize → EventCorrelation → ContextEnhance → HypothesisBuilder → RiskScoring → CapabilityMatch → Routing → TaskingPlanner → DraftOutput → ValidateOutput → AuditLog
- **Rust 符号推理引擎：** JSON Logic 规则评估 (25+ 触发器确定性匹配)、多维风险评分 (硬性覆盖规则)、SOP 能力匹配与缺口分析、结构化输出校验
- **双 RAG：** GraphRAG (Neo4j) + VectorRAG (Qdrant, 法规/预案/案例检索)

技术栈： YOLOv11, ByteTrack, InsightFace, PaddleOCR, TensorRT, GStreamer, LangGraph, Neo4j, Qdrant, Rust (Axum), PostgreSQL, Jetson Orin NX

项目五：AI 应急指挥大脑（已交付）

面向应急救援的智能指挥系统，Neuro-Symbolic 混合架构，个人独立完成。该项目为某战术大脑的技术基座。

架构： FastAPI + Neo4j + Qdrant + WebSocket 后端 | React + TypeScript + Cesium + Mars3D 三维数字孪生前端 | Rust 混合 A* 路径规划引擎 (PyO3 桥接)

核心能力： - **20+ Agent 协同集群 (LangGraph + CrewAI)**：EmergencyAI 主决策 (17 步状态机 + HITL 人在回路 + 断点恢复)、ReconScheduler 侦察调度 (10 阶段 + 能量模型 + 三种覆盖路径)、OverallPlan 总体方案 (LangGraph + CrewAI + MetaGPT 三框架混合)、EarlyWarning 预警监测、VoiceCommander 语音指挥等 - **HTN 分层任务网络**：基于 Neo4j 的 Scene → TaskChain → MetaTask 三层分解，多场景组合 + 并行调度 - **运筹算法矩阵**：Rust 混合 A* (路网 + DEM 栅格统一代价搜索)、MCTS 任务序列优化、NSGA-II/III 多目标优化、VRP 多车路径规划 (OR-Tools)、CSP 能力-资源匹配 (CP-SAT)、离散事件仿真 (含生命体征仿真模型)、灾情评估 (地震烈度衰减/高斯烟羽/洪涝模型) - 语音交互：* ASR 多引擎适配 + TTS CosyVoice 语音合成

技术栈： LangGraph, CrewAI, MetaGPT, Neo4j, Qdrant, Rust (PyO3), NSGA-II, MCTS, OR-Tools, FastAPI, PostGIS, Cesium, React, TypeScript

北京工成商通科技有限公司 | 核心架构师 / AI 研发负责人

2022.08 – 2025.06

作为公司核心技术领导，主导自研 AI 项目及大型国企（青岛城运、青岛航信）架构设计与实施。

项目一：DeepSeek 税务智能体系统（服务于青岛航天信息）

2025.03 – 2025.06

- 主导 Agentic RAG 框架落地，设计“自主检索-动态反思-工具调用”税务专家 Agent
- 引入 Re-ranking 重排序机制，长尾法规检索精度提升 40%
- 成果：7×24 无人化税务咨询，复杂问题一次性解决率从 52% → 85%

项目二：AI 驱动的教育平台（公司自研）

2024.11 – 2025.03

- 部署 Agentic RAG 框架，实现自主决策的知识检索与生成
- LORA 微调 Llama 2 / GPT 系列模型，生成个性化学习计划
- 成果：学习计划采纳率 78%，用户留存率提升 35%

项目三：青岛城运“e 生化”综合服务平台（Java）

2022.08 – 2024.10

- 微服务架构：基于 Spring Cloud Alibaba（Nacos/Sentinel/Gateway）构建微服务底座，支撑 7,000+ 车辆、百万级用户并发
- GIS 核心攻关：自研 GPS 轨迹校正（HMM + 卡尔曼滤波），年节省第三方服务费 30 万+
- 金融级安全：国密 SM2/SM4 双离线支付协议 + API 动态防御网关，通过国家保密局安全审计
- 智能调度：VRP 算法优化配送路线，车辆装载率提升 22%

技术栈：Java, Spring Cloud Alibaba, Flink, Kafka, TiDB, Redis, Netty, SM2/SM4, HMM, VRP

北京人瑞人力资源服务有限公司 | 后端研发工程师（Java）

2021.09 – 2022.07

智能家居 B2B2C 电商平台：基于 DDD 重构商品模型，Redis Cluster 多级缓存 + Elasticsearch 搜索引擎，秒杀 QPS 峰值提升 3 倍，RabbitMQ 异步削峰 + ELK 分布式日志

技术栈：Java, Spring Cloud, Redis Cluster, RabbitMQ, Elasticsearch, ELK, Docker, K8s

黑龙江溜光大道网络科技有限公司（创业） | 联合创始人 / CTO

2020.12 – 2021.07

溜光大道 APP：独立完成 Java 后端 + Android 原生（Kotlin）全栈交付，用户激励闭环 + 三级分销 + Kafka 实时结算与反作弊

技术栈：Java, Spring Cloud, Redis Cluster, Kafka, Android (Kotlin), Docker, K8s

北京新致软件有限公司 | 架构师 / 技术负责人（Java）

2016.10 – 2020.11

保险行业核心系统（中银/光大永明/人保健康）：主导旧系统 Spring Cloud 微服务化重构，千万级数据批处理从数小时缩短至分钟级，管理 50+ 人研发团队，负责基础架构建设（Redis/Kafka/Nginx）

技术栈：Java, Spring Cloud, Redis, Kafka, Nginx, Oracle, MyBatis

中民投集团 | Java 开发工程师

2015.04 - 2016.10

金融科技与智慧社区平台核心交易、O2O 订单及秒杀模块开发。

技术管理经验

- 管理 50+ 人研发团队（上市公司），制定技术规范与敏捷流程
- 管理 20+ 人跨部门项目组（青岛城运），协调产品、开发、测试与运维
- 创业公司 CTO 经验，擅长技术团队从 0 到 1 组建

个人特质

- **全流程独立交付：**从产品需求分析、架构设计、多语言开发（Java/Python/Rust/Go/TS）、AI 大模型集成、前端可视化到 CI/CD 部署运维，可一人完成全流程
- **AI Native 开发者：**深度使用 Claude Code、Codex CLI、Cursor 等 AI 开发工具构建生产级系统，持续探索人机协同最佳实践
- **极致热情：**日均工作至凌晨 2 点，全年持续学习，技术是生活方式