

靳泽旭 / Zexu Jin

AI Agent Harness Engineer · Agent Runtime / Tool Use / Context / State / Evals · 后端与 AI 基础设施

电话: 18158522238 · 邮箱: jackjin1997@gmail.com · GitHub: github.com/jackjin1997 · 个人站点: jzxwiki.com · 学历: 本科 / 软件工程

个人简介

近 6 年服务端架构与生产系统经验, 近 3 年聚焦 LLM 应用与 AI Agent 系统。曾作为服务端技术负责人带领 5 人团队, 从 0 到 1 搭建承载 Agent 的工程骨架: Agent Loop、Tool Use 调度、Context 构造、State / Checkpoint、多 Agent 编排、MCP 接入、可观测性与评测闭环。

2026 年持续向 LangChain / LangGraph / LangChainJS 等项目提交并合并 Agent runtime、checkpoint、streaming、tool-use 边界问题修复; 也短期参与钉钉 AI Agent 方向, 补充企业协同产品中的 Agent / CLI / 工具链场景理解。

关键词: Agent Harness, Agent Runtime, ReAct Agent Loop, Tool Calling, MCP, LangGraph Checkpoint, LangSmith, LLM-as-Judge, Context Engineering, Multi-Agent, Claude Code, Cursor, Python, TypeScript, FastAPI, Hono, Kubernetes, Redis, etcd.

开源证据

- **LangChain / LangGraph 生态:** 已合并 11 个上游 PR, 覆盖 langgraphjs、langchainjs、langchain; 代表问题包括 checkpoint 后端一致性、特殊通道写入、metadata filter、Gemini built-in/function tools 混用、streaming function_call chunks、边界崩溃预防。
- **代表 PR:** langgraphjs #2516 / #2517 / #2518 / #2208 / #2186, langchainjs #10676, langchain #36108 / #34943。
- **其他已合并 PR:** OpenViking #800 Kubernetes Helm Chart, openclaw #47465 Markdown 解析修复, Antigravity Awesome Skills #61 / #69, Cosmos SDK #25863。

工作经历

Open Source / Agent Infrastructure · 上游贡献者 · 钉钉 AI Agent 短期经历

2026.03 — 2026.06 · LangChain / LangGraph / LangChainJS / OpenViking · 钉钉约 2 周

- 提交并合并 checkpoint 后端一致性、特殊通道写入、streaming function_call、Gemini mixed tools、部署基础设施等 Agent runtime 修复。
- 通过真实框架 failure mode 建立对 Agent state、tool invocation protocol、context persistence、developer experience 可靠性的工程判断。
- 短期参与钉钉 AI Agent 方向, 快速熟悉企业协同产品中的 Agent / CLI / 工具链业务边界、权限约束与工程协作方式; 作为补充经历呈现, 不展开为核心项目。

跨越星空科技有限公司 · 服务端开发 → 服务端技术负责人

2022.09 — 2026.02 · AI Agent 与智能自动化产品 · 带 5 人服务端团队

- 负责 AI Agent 产品线运行时骨架与后端基础设施的架构、选型、评审与关键实现方向。
- 主导 IMEANAI-Coyage Agent Harness 0 到 1, 包括多 Agent 编排、ReAct Loop、Dynamic Tool Selection、MCP 工具接入、Context / State 处理、流式推理、可观测性与可靠性机制。
- 设计 TypeScript + Hono 的 AI 在线服务框架, 作为团队 AI 微服务统一运行时底座, 支持 typed actions、streaming RPC、Zod 校验、etcd 选主、动态配置与强类型 SDK 生成。
- 推广 AI-native 工程流程: Claude Code / Cursor 主力编码, 人控架构与质量门禁, Issue 到实现到 PR Review 流程, 维护 CLAUDE.md / rule files 沉淀项目上下文。
- 早期主导类 Dify / N8n 的 AI 原生工作流编排平台后端交付, 服务银行、电商 B 端客户并获得续约。

FTL Group / 法翔物流 · 全栈开发工程师

2020.07 — 2022.07 · 法国电商物流 SaaS, 面向欧洲 B2B 市场

- 参与 esendeo 物流 SaaS 核心开发: SQL 与对账优化、AWS Lambda 异步 PDF 渲染、定价引擎拆表与 Redis 缓存。
- 从 0 到 1 搭建 esendeo-shopify 小程序, 基于 Next.js + Koa 打通 Shopify 订单同步与物流状态回填。
- 封装内部 UI / ORM 组件, 构建 Puppeteer / Playwright 爬虫用于地址与海关编码校验流程。

项目经历

IMEANAI-Coyage · 多 Agent 旅行助手 / Agent Harness

2024.01 — 2026.02 · 服务端技术负责人 / Agent 架构师 · Python, FastAPI, LangChain, LangGraph, LangSmith, MCP, Claude, OpenAI, Bedrock, Qwen, Pinecone, Docker, Kubernetes

- **问题:** 自然语言旅行规划需要可靠 Tool Use 与实时校验; 纯规则难以处理复杂意图, 纯 LLM 容易生成不可预订或不可验证的行程内容。
- **Agent 编排:** 构建 Supervisor + 6 类专家 Agent, 覆盖酒店、机票、餐厅、景点、行程、汇率等领域; 子 Agent 内部跑 ReAct Loop, 完成推理、工具调用、观察与自检。
- **工具可靠性:** Dynamic Tool Selection 仅注入当前意图相关工具描述; MCP 标准化工具注册; Zod Schema 校验、超时处理、fallback 与降级保护主链路。
- **状态与上下文:** LangGraph Checkpoint + Time Travel 支持长对话恢复和节点级回滚; 按 token 阈值触发摘要压缩, 控制上下文膨胀。

- **评测与可观测性:** LangSmith 覆盖 Tool Call / Chain / LLM 三层 trace, 自定义 event dispatch 打标; 离线 LangSmith Dataset + LLM-as-Judge 驱动 prompt、harness、工具描述联合迭代。
- **结果:** 产品上线并持续迭代; 通过多 LLM fallback、工具熔断、Redis 削峰、prompt cache、SSE 流式体验和 MCP 声明式工具接入提升稳定性与迭代效率; 复杂意图下 TTFB 控制在 10 秒以内。

AI 在线服务框架 · imean-service-engine

2024.01 — 2026.02 · 独立设计 / 核心开发 · TypeScript, Hono, Zod, Etcd3, Redis, EJSON, Winston, LRU-Cache

- 搭建轻量、类型安全、可扩展的 Agent / AI 服务微服务框架, 支持装饰器 API、typed actions、RPC + AsyncIterator 流式传输、幂等标记和强类型客户端生成。
- 实现基于 etcd 的方法级分布式任务选主, 用于 Agent 评测回归、向量索引刷新等定时任务, 多实例下唯一执行并支持故障自动转移。
- 支持模型、温度、Prompt 模板等 Agent 参数动态热更新, 避免每次调参重新部署。

AI 原生工作流编排平台

2022.09 — 2024.01 · 服务端执行 → 技术负责人 · TypeScript, GraphQL, Prisma, React, Redis, MongoDB, RabbitMQ

- 构建面向 B 端客户的类 Dify / N8n 工作流编排平台, 将 LLM 调用、AI 决策路由、向量检索深度融入执行引擎。
- 实现 AtomFlow + CombinedFlow, 支持嵌套、并行工作流执行, 多租户 RBAC, 版本回滚, JSON 导入导出与审计追踪。

技术技能

Agent Harness: ReAct Agent Loop, Tool Use, Dynamic Tool Selection, MCP, Multi-Agent / Supervisor, Context Engineering, LangGraph Checkpoint, Time Travel, Agent State, Skills。

LLM / 评测: Claude, OpenAI, AWS Bedrock, Qwen, Streaming LLM APIs, Prompt Cache, 多模型路由, LangSmith, LLM-as-Judge, trace-based failure attribution, Multi-LLM Fallback。

后端 / 基础设施: TypeScript, Python, Go, FastAPI, Hono, GraphQL, NestJS, 微服务, etcd, Redis, RabbitMQ, PostgreSQL, MongoDB, Docker, Kubernetes, GitLab CI, Linux 性能分析。

AI 辅助开发: Claude Code, Cursor, repository rule files, CLAUDE.md, Issue 到实现到 PR Review 工作流。

教育与证书

浙江大学城市学院 · 软件工程 · 本科 · 2015 — 2019

软件设计师 (中级)