

冉熙

ranxi@88.com · ranxi.top · 19143656913 · GitHub @ranxi2001

个人信息

- 求职意向: 云原生 / AI Infra 研发工程师 (容器与大模型推理调度)
- 基本信息: 2001.09 出生 | 现居住地: 浙江宁波 | 研二在读 | 2027 年 6 月毕业
- 社区身份: 中国科学院大学开源与人工智能协会 (OpenCAS) 成员 (2024.09 – 至今)

教育背景

中国科学院大学 2024.09 – 2027.06

硕士, 计算机技术 (杭州高等研究院智能科学与技术学院, 杭州) | 中科院计算所 + 浙大联合培养

- 学业成绩: GPA: 3.75 技术方向: Agent Runtime、云原生调度与沙箱、分布式系统性能评测
- 核心课程: 并行计算, 计算机体系结构, 高级人工智能, 自然语言处理, 机器人与智能计算

浙江工商大学 2020.09 – 2024.06

学士, 信息管理与信息系统 (杭州) | GPA: 3.77 | 专业排名第一 + 保研至中科院 | 英语四六级

专业技能

- 编程语言: 以 Go (AgentCube/Karmada 上游协作)、Python (Agent/CLI/Benchmark/自动化) 为主; 熟悉 Java、SQL、JavaScript/TypeScript
- 云原生与系统: 熟悉 Kubernetes CRD/Controller、Pod 资源管理与 E2E 测试, 掌握 Docker、Linux、Git/GitHub Actions; 了解 cgroups/namespaces、RuntimeClass/CRI、KVM/MicroVM 与 Warm Pool
- Agent Runtime: 熟悉 Agent Loop、Tool Use、上下文与 Session 状态、沙箱生命周期 (创建/预热/执行/回收); 熟悉 OpenAI/Claude API、MCP 及 LangChain/LangGraph
- 开源工程: 具备上游 Issue 定位、PR 拆分、Code Review 与 CI 故障归因经验; 能够设计可复现实验并使用 p50/p95、成功率与资源状态分析性能瓶颈

工作与实习经历

华为, Agent Infra / 云原生开源工程师 (实习) 2026.06 – 至今

- [Agent Sandbox 适配]: 开发 agent-sandbox v0.5 适配分支, 覆盖 Runtime Adapter、CRD 与对象身份、Controller/Handler 及 E2E 链路; 完成 unit/build/lint/gen-check、Ubuntu 24 kind 集群与默认 AgentRuntime mTLS 验证, 明确旧 CRD 迁移等剩余风险后再推进上游。
- [沙箱性能与调度]: 设计 create session/run code/delete session 分阶段 Benchmark, 在本地 k3s 10 并发实验中定位 Warm Pool 容量瓶颈; 将池容量由 2 调至 10 后, 两轮端到端 p50 由 7.315s 降至 436/565ms, p95 为 804/933ms, 10/10 请求成功。
- [运行时架构研究]: 系统对比 Agent Sandbox、OpenSandbox、Agent Substrate、E2B、forkd、CubeSandbox、cage-bro 等方案, 围绕 K8s CRD 慢路径、节点本地生命周期、MicroVM 快照与 Sleep/Resume 形成架构分析和可验证演进路线。

Bybit, 风控 Agent 算法工程师 (实习) 2025.12 – 2026.06

- [Agent 任务编排]: 从零设计 5 阶段 OI 排查 Pipeline, 开发 12 个 Tool 对接 Hive/Presto, 支持条件路由、5 路并行、CDP 自动恢复与断点续传, 将 40 分钟人工流程压缩至 4 分钟端到端执行。
- [异构计算任务调度]: 构建批量计算 Agent, 自动编排 356 个 Spark/Presto 任务并按数据量切换执行引擎; 将单次扫描量从 10 亿 + 行降至 1%, 全币种 7 分钟完成计算并支撑 9 轮算法迭代。

浙江大学, 量化 Agent 系统全栈研发 (实习) 2025.06 – 2026.06

- [Agent 研究与风控]: 构建 Agent 驱动的 Web3 研究流水线, 自动编排新闻抓取、情感分析、信号生成、异常检测与交易执行, 形成从研究到风控处置的自动化闭环。
- [生产平台工程]: 使用 FastAPI/MySQL/Redis/Vue 实现多账户风险监控、Agent 状态可视化与策略配置平台, 接入持续运行环境并支撑策略自动化执行。

开源贡献

AgentCube (Volcano 社区) — Kubernetes 原生 Agent 运行时 (10 个已合入 PR) 2026.06 – 至今

- 以 Go/Kubernetes 工程栈持续参与上游，合入 10 个 PR，覆盖 Agent Sandbox 适配、Go 工具链与多架构镜像、组件清理、依赖与 CI 治理等；另有 2 个 PR 进行中。
- 参与 SandboxPool Feature Proposal #431 评审，围绕 Agent 心跳失效、Static Pod 扩缩容、RuntimeClass/CRI 路由、资源一致性等提交 13 条公开评审意见，其中 3 项问题由作者更新解决。

Karmada — Kubernetes 多集群编排系统 (4 个已合入 PR) 2026.06 – 至今

- 合入 4 个上游 PR：更新 Kubernetes/etcd 版本基线，迁移 18 个 GitHub Actions Workflow 至 Ubuntu 24.04，修复 FlinkDeployment CRD 清理时序与 RemedyAction 状态更新问题；另有 1 个证书轮换特性 PR 进行中。

K8s 社区开源项目 — Kubernetes 生态与开发者工具 (5 个已合入 PR) 2026.06 – 至今

- 向 kubernetes-sigs/work-api、Agents365-ai/drawio-skill、Karmada Website/Dashboard 合入 5 个 PR，覆盖 GitHub Actions 升级与固定、Runner 迁移、分发元数据一致性与图表规范；连同 AgentCube/Karmada，累计在 6 个上游仓库合入 19 个 PR，另有 3 个进行中。

zero2Agent — 面向开发者的 Agent 工程开源体系 (GitHub ★181) 2026.03 – 至今

- 独立构建 80 篇教程 (8 模块)，从 60 行 Agent Loop 拆解到生产级架构，覆盖 Tools、Memory、Subagent、Teams、Worktree 隔离与 SDK 差异；完成 13 个主流 Agent 框架源码级横评，阅读量 30w+，阅读人数 6 万人。

开源项目与工程实践

Agent Infra 开放研究资产 — 可复现报告与实验记录 2026.06 – 至今

- 围绕 Agent Sandbox、Warm Pool、Sleep/Resume、K8s 慢路径/节点快路径与 E2B 协议持续公开研究过程，累计沉淀 50 份技术报告与 103 个原始 Benchmark/验证产物。
- 记录运行环境、原始日志、同 SHA CI 对照、历史 Flake 与未验证项，区分上游公开指标和本机实测结果，为性能结论、PR 拆分与 Code Review 提供可追溯证据。

OfferPilot — 自研 Agent Harness 面试诊断系统 (GitHub ★21) 2026.06 – 至今

- 手写 Agent Loop 与多子 Agent 协作链路，实现 Tool Use、上下文预算、会话状态、流式输出和断点续传；集成 ASR 完成语音录入、转写、诊断与 Markdown 报告导出闭环。

荣誉奖项

国家级:	2024 研究生数模竞赛二等奖 (1/3) 2023 挑战杯揭榜挂帅全国二等奖 (1/10)	2022 美国大学生数学建模 ICM H 奖 (1/3) 2021 国家励志奖学金
省部级:	2022 浙江省政府奖学金 (前 5%) 2022 省挑战杯铜奖 (1/10) 2022 省物流设计竞赛二等奖 (4/5)	2022 全国大学生数模竞赛省二等奖 (1/3) 2022 省电子商务竞赛二等奖 (3/5) 2021 省互联网 + 创新创业竞赛铜奖
校级:	2025 国科大学业二等奖学金 (前 20%) 2022 海平专项奖学金 (全校仅 30 名) 2023 综合能力突出奖学金	2021、2022 综合一等奖学金 (前 3%) 2022 勤工助学之星 (全校仅 10 名) 2022 三好学生 (前 5%)

发表论文

- Sun, S.; **Ran, X.**; et al. “DataR2E: Research and Prospects on the Value Release of Data Elements in Web 3.0”. *IEEE Global Blockchain Conference (GBC)*, 2025.
- Ju, C.; et al.; **Ran, X.**; et al. “Blockchain Traceability System...Image-Based Interactive Traceability Structure”. *Systems*, 2022. (JCR Q2)