

张跃 Zhang Yue

求职意向：架构师 / AI Agent 开发工程师

男 | 33岁 | 12年工作经验（3.5年 AI 应用与 Agent 落地） | 期望城市：深圳

18679223924 okeeper.com@gmail.com okeeper.com github.com/okeeper

个人微信: 长风 公众号: 神笔君

> 对 AI 时代软件开发的判断与思考：从某种意义上讲，AI 已经对软件开发行业带来了深刻的变革，未来不再有传统的前端、后端，或 Java / Python / PHP 等具体实现语言分工的团队协作方式和组织形式，只有一种角色叫 **AI 工程师（他更像是一位“指挥家”）**。越到后面，越看重个人的**单兵作战能力，以及团队协作、需求描述、逻辑描述和执行力**等软实力。

个人优势

Core Strengths & Technical Summary

- **入局早、自驱力强的 AI 应用实践者（2023年初入局）：**早在 2023 年年初就开始关注 AI 发展并全身心投入 AI 的使用及相关技术研究，意识到这是一次伟大的技术革命；自发提议在公司内部构建统一 Token Hub（KuGPT），方便各业务线快速开展 AI 业务创新。3 年多以来，除日常工作外持续维护个人技术公众号**「神笔君」**，不定期分享前沿技术和个人 AI 看法，引起不少关注。
- **企业级 AI 研发效能平台 / 云端 AI IDE 从 0 到 1 架构与规模化落地：**基于 **VS Code 深度自研**适配 Akulaku 及东南亚金融企业的 AI 研发效能平台，担任**底层架构设计及核心研发**，攻克**底层沙箱隔离、云端开发环境、外部模型内部分发与计量统计、高并发 Worker 计算调度**等难点；打通需求输入→代码产出→个人 **Vibe Coding** 及跨团队协作并沉淀 **Harness 资产**。已在**电商（约100人）与金融平台部（约300人）共 400+ 研发大规模落地**，并获**泰国银行及东南亚本地团队**青睐，助力客户实现**“数据不出网”**研发提效。
- **大模型底层原理、微调（SFT/RLHF）与智能客服 Harness 落地：**了解 LLM 底层基本原理，熟练掌握 **Pre-train、Fine-tuning、SFT、RLHF** 等相关微调技术；熟练使用 OpenAI、Anthropic 等平台及向量数据库、**Dify、LlamaIndex、LangChain** 等框架。在电商和金融智能客服场景通过真实对话记录与 Wiki 语料清洗预处理及开源 **Qwen 模型微调**，使**人工介入率从 90% 降到 40% 左右**；并迭代 **Harness 工程**将下单、授信、订单查询、退款等核心流程融入对话式交互。
- **12 年大型后端及分布式架构设计经验：**曾任职**阿里巴巴菜鸟网络**，现任 **Akulaku 架构师及公司架构委员会成员**。精通 Spring Cloud、Dubbo 微服务架构、领域驱动设计（DDD）、JVM 性能调优与分布式系统一致性设计；开发过压测工具、可重试框架、分布式 ID 生成器等多种通用组件。
- **团队管理与高效沟通协调能力：**长期带领一线开发团队（7人）快速实现业务需求，结合成员特点帮助其适应性成长，打造了富有战斗力的核心团队，在需求分析、架构设计、跨部门协调到上线及应急处理中实现了良好的工程口碑。

核心项目经历（AI Agent 与大模型落地）

Key AI Agent & LLM Projects (2023 – 2026)

1. 企业级团队研发效能 AI 平台 / 云端 AI IDE 架构与研发 底层架构设计及核心研发（从 0 到 1 主导核心架构） 2024.06 – 至今

项目概述与规模：基于 VS Code 深度自研适配 Akulaku 及东南亚金融企业的研发效能平台。内部已覆盖**电商团队（约100人）与金融平台部（约300人）** 400+ 研发规模化使用 ；对外获**泰国银行及东南亚本地团队**青睐，助力客户在 数据不出网 前提下实现团队研发提效。

- **底层沙箱环境隔离 & 云端开发环境（金融级安全）：**从底层攻克 Sandbox 沙箱隔离 与 云端开发环境 架构研发，实现多租户强隔离、一致的云端研发环境及安全受控的 Agent 代码执行，严守银行与金融业务“核心代码与数据不出网”红线。
- **外部模型内部分发统计 & 高并发 Worker 调度：**构建外部大模型统一代理、 内部分发与计量统计 体系（多模型路由、配额审计与成本核算）；面向 400+ 研发并发及异步 Agent 任务设计可水平扩展的 Worker 计算调度架构 ，保障高并发场景下的稳定与低延迟响应。
- **全流程 Vibe Coding 衔接 & Harness 资产沉淀：**打通需求输入→代码产出→个人 Vibe Coding 及 Bug 修复、需求变更、跨团队协作闭环；将团队研发规范与个人 Harness 经验 沉淀为与模型解耦的公司级资产，伴随基座模型演进驱动团队效率持续跃升。

2. 电商与金融智能客服大模型微调 & Harness 对话式业务 Agent 核心架构设计 & 模型微调与 Agent 研发 2023.09 – 2024.08

项目概述：针对电商和金融场景开展智能客服 AI 创新应用落地，从早期检索编排到深入大语言模型底层微调提升准确率，并演进为融合核心交易链路的对话式业务 Agent。

- **多框架检索与编排打磨：**早期从提示词（Prompt）设计、RAG（检索增强生成），到 Dify LlamaIndex LangChain 等主流框架都做了大量尝试和工程优化。
- **真实语料清洗预处理与开源 Qwen 模型微调：**深入研究大语言模型底层基本原理，熟练运用 Pre-train Fine-tuning SFT RLHF 等相关微调技术；针对电商和金融常见问题，使用公司内部沉淀的**真实线上聊天客服记录、内部资料及 Wiki** 做数据语料收集与预处理，对开源 **Qwen** 模型进行微调，从各个维度优化回复准确率，取得显著成功—— 人工客服介入率从 90% 降到了 40% 左右 。
- **Harness 工程智能客服迭代（对话式电商交互）：**后续迭代使用 Harness 工程 改造智能客服，把电商核心的如**下单、授信、订单查询、退款**等核心业务流程融入到智能客服中，方便用户直接进行对话式电商业务交互，迎合当前个人 Agent 操作一切的使用习惯。

3. KuGPT 平台建设（公司内部统一 Token Hub） 提议发起人 & 架构设计及开发 2023.03 – 2023.08

- **自发提议与能力集成：**2023 年初随着 ChatGPT 爆火，自发提议在公司内部构建自己的 Token Hub（KuGPT），使公司各业务线能够低成本快速接入文本生成、文档向量化、文档向量搜索等能力，屏蔽底层各大模型差异，做到零成本切换。
- **异构架构与安全计费管控：**接口兼容主流标准并采用 JWT 鉴权；调研主流向量数据库，落地「业务逻辑入口为 Java + 底层核心 AI 交互为 Python」的可水平扩缩容分布式架构；统一管控敏感词出网安全、付费 API 路由与 Token 计费管理。成功支撑 **8 个业务应用**对接放量，日均 Query 量 **2w+**。

大型分布式系统与高并发架构项目

Distributed Systems & High-Concurrency Architecture

4. Akulaku 电商交易中心微服务重构 & 业务监控违规风控体系 架构师及开发负责人

2019.09 – 2023.01

- **核心交易与计价系统重构：**主导电商核心交易系统从单体向微服务架构（Spring Cloud/Dubbo）的推翻重构；抽象通用计价中心与高扩展优惠库表模型（兼容优惠券、运费券、零元购等迭代）；解决分布式一致性事务问题，采用美团 Leaf 作为分布式 ID 生成器。
- **全链路压测与性能调优：**主导上线前系统压测与性能瓶颈分析，优化后**交易下单单机写接口性能达 300 TPS（TP99 耗时 400ms），订单读接口单机达 1000 QPS。**
- **业务监控体系与违规中心：**设计维护近 50 项业务和指标监控，结合 Grafana 实时微信/电话告警；针对印尼市场灰产刷单防控，设计开发「交易违规中心系统」，通过 **EL 表达式动态维护监控指标获取、计算规则与告警内容。**
- **通用组件封装共享：**开发压测工具 `ec-service-benchmark`、高可靠可重试框架 `consistent-invoker-spring-boot-starter`、分布式 ID 组件 `leaf-spring-boot-starter` 等，大幅提升团队开发效率与代码复用率。

5. 阿里菜鸟国际 WMS 库存对账系统

2018.04–2018.09

- 担任项目 PM / PG，针对复杂仓储系统仓库库存与上游不一致痛点，独立完成架构设计、工期评估与开发落地。
- 优化 WMS 库存技术架构，实现差异报表与智能调账系统，**将 WMS 出库库存上下游准确率从 99% 提升至 99.99%。**

6. 百倍现货交易系统 & 千米商品中心

2015.03–2017.08

- 独立设计百倍科技交易核心模块，基于 Redis 研发**高性能内存撮合引擎（撮合性能达 300 TPS）**，主导公司从 0 到 1 微服务化实践。
- 负责千米网核心商品服务中心重构、收银台开发与大规模商品数据库结构整改迁移。

工作经历

Work Experience (12 Years)

Akulaku 架构师 / AI Agent 开发工程师（公司架构委员会成员）

2019.09 – 至今

- **AI 研发平台与智能客服 Agent 落地：**从 0 到 1 主导基于 VS Code 深度自研的企业级 AI 研发效能平台 / 云端 AI IDE 底层架构与核心开发，攻克底层沙箱隔离、云端开发环境、外部模型内部分发统计及高并发 Worker 计算调度难点，已在电商（约100人）与金融平台部（约300人）共 400+ 研发大规模落地，并获泰国银行及东南亚本地团队青睐（保障数据不出网）；主导智能客服 Qwen 微调（SFT/RLHF，人工介入率 90%→40%）与 Harness 对话式交易改造；发起主导 KuGPT（Token Hub）。
- **核心系统架构与团队管理：**负责电商交易中心交易链路各系统稳定维护、微服务重构与性能调优（写 300 TPS / 读 1000 QPS）；带领 7 人一线开发团队，协调工作量、把控发版质量、拆解 KPI 进行前瞻性规划并帮助成员成长；作为公司架构委员会成员负责重要技术方案制定。

阿里巴巴菜鸟网络 (Alibaba Cainiao Network) 高级开发工程师

2017.09 – 2019.09

- 负责菜鸟网络仓储管理系统（WMS）的日常需求分析、开发与系统设计优化；主动跟踪解决线上问题并编写自动化脚本工具提升可靠性；独立设计并实现 WMS 与上下游链路库存差异报表及智能调账的库存对账系统（准确率 99%→99.99%）；在大部门组织《比特币实现原理》主题分享。

广州百倍网络科技有限公司 开发工程师

2015.09 – 2017.08

- 独立完成核心交易模块架构设计开发及 300 TPS Redis 撮合引擎系统；参与领域模型划分及 Spring Cloud / Dubbo 微服务架构从无到有落地，负责服务容灾与自动化测试方案；在公司内部组织《MySQL 读写分离》《Disruptor》《服务注册发现》等技术分享。

江苏千米网络科技有限公司 Java 开发工程师

2014.07 – 2015.09

- 独立完成私有云电商平台定制项目研发、千米网收银台项目开发及商品数据结构调整迁移工作；曾多次获得公司内部优秀员工称号。

教育经历与专业证书

Education, Certifications & Technical Output

🎓 **深圳大学** · 本科 · 计算机科学与技术 🏅 **资格证书：**软件设计师（国家中级职称） ✍️ **长期技术输出：**公众号「神笔君」 | 个人网站 okeeper.com