

Big Data Engineer · Full-Stack Builder · AI Agent Developer

# 鹿尧

深耕大数据多年，持续推进全栈开发与 AI Agent 工程实践

十年大数据开发工程师，数据分析师，长期在数据平台、数据仓库、实时计算、用户画像和业务分析场景里做工程落地。在微软持续拓展全栈开发和 AI Agent 工程能力，关注如何把数据、产品系统和 LLM 工程结合成可交付的软件。开源社区活跃者，GitHub 原创公开仓库收获 1.6K+ 星标，并在 OpenClaw、Hermes 等项目中有可核验合入贡献。

✉ 364939526@qq.com    📞 +86 136 9977 4962    📍 苏州 / 杭州 / 上海

## Experience

### 高级软件工程师

微软 (STCA · Bing / Skype / Copilot) 2022-04 → 至今 苏州

在 STCA WebXT/MAI 先后承担 Bing App / Start App、Skype、Copilot Mac 和 Surfaces Connectivity 的数据与质量工作；从业务数据 owner、指标体系和数据仓库，逐步扩展到跨端 telemetry、实验 scorecard、质量 RCA 和 AI-native 数据 workflow 建设。

- **2025.11 - 2026.04 Cross-Surface Connectivity 质量分析**

转入 Surfaces Connectivity 后，负责 Mac / CMC 等跨端 no-response、actionable-rate 和 chat funnel 质量分析；为 Mac 对齐 Mobile 的 KQL、Grafana dashboard、Geneva monitor 和 baseline，并定位 CMC bot filter、background-tab 1DS terminal-event 丢失等 telemetry artifact，把“真实失败”和“数据采集失败”拆开，帮助团队聚焦正确 root cause。

- **2025.02 - 2025.11 Copilot Mac 数据与质量体系**

在 Copilot Mac 数据基础薄弱的阶段，补齐核心指标、质量监控、实验 scorecard 和 feature dashboard；建立 ICM framework、daily / weekly OCV review 与 LT/PM 周会节奏，并主导 telemetry guidance、native rewrite backfill 和 voice funnel 分析，使留存、WAU、voice TDR 等核心指标能稳定跟踪、解释和优化。

- **2022.04 - 2025.02 Bing App MiniApp / Growth / Acquisition 数据**

覆盖 Bing App MiniApp (Rewards、Wallpaper、Weather) 以及后续 Growth / Acquisition 业务, 负责核心指标、dashboard、funnel 和 ROI / LTV 分析; 为 PM、运营和增长团队提供 feature 前后评估、Paid Ads 投放效率、Upsell funnel 与用户画像分析, 支撑预算和增长策略决策。

- **2022.04 - 2025.02 Skype Call Quality 与数据基础**

主导 Skype call quality funnel 建设, 从 telemetry 梳理、funnel dashboard 到优化点识别和 impact 计算, 支撑 1v1 call reachability / stability 等质量目标; 同时对分散的数据表做关键模块数仓建模和 ODS / DWS 表开发, 服务 PM dashboard、dev debug、AB scorecard 和质量 review。

## 高级软件工程师

**小米 (大数据 · 用户画像 · MiPush)** 2019-04 → 2022-03 北京 / 南京

在 MiPush 和用户画像方向负责 PB 级数据处理、实时/离线统计、竞品分析、DMP 标签和换机口径; 作为 6 人数据小组的技术 Owner, 除核心工程实现外, 还承担任务拆解与排期、代码 review、招聘面试、新人培养和跨团队沟通。

- **2021.04 - 2022.03 用户画像 · 小米换机数据**

初期主导集团级换机口径定义、数据建模、代码实现和质量验证, 解决多卡、多手、多设备流转下的序列识别问题, 产出准确率 90%+、月活覆盖 80% 的换机标签; 数据被全集团多个手机业务部门复用, 用于品牌忠诚度、拉新画像和手机业务战略分析。

- **2019.04 - 2021.04 MiPush · 安卓竞品数据**

接手并消化二十多个计算任务、数万行代码的竞品统计系统, 基于 PB 级 MiPush 数据构建手机厂商新增与活跃指标; 通过清洗、映射和 OAID 兼容, 将与 BCI 的统计误差从 50%+ 拉回到 10%, 并补齐 Kylin / Redash 可视化链路。

- **2019.04 - 2021.04 MiPush · 手机销售赋能与 DMP 标签**

作为一期主程设计文案挖掘数据仓库和项目框架, 开发三十余张表中的核心部分; 构建购机意向打分和意向品牌标签, 并为算法模型提供特征数据, 支撑小米商城留存、去库存和多业务线精准投放。

- **2019.04 - 2021.04 MiPush · 推送统计平台**

主导“米 Push 推送统计平台”重构, 面向集团内部业务方和外部第三方开发者提供全选人数、推送量、接收量、点击量、CTR、分时接收/点击等核心报表; 改造旧系统脉冲式写 HBase 和不合理数据结构, 提升稳定性并减少存储空间, 每年节省近百万元人民币成本。

## 大数据开发工程师

**美图 (风控与反作弊)** 2018-04 → 2019-03 深圳

在社区风控与指标监控方向负责实时流处理、规则引擎、异常检测和告警平台工程化, 把风控 PM / 运营规则配置、算法模型和生产告警串成可运营系统。

- **2018.04 - 2019.03 风控实时规则引擎**

面向美图社区全链路风控, 覆盖私信、关注、社区内容发布和文本创作等场景; 基于 Spark Streaming + Kafka 将实时日志抽象成可查询内存表, 用 QLEXPRESS 支持风控 PM 自助配置规则, 并

接入文本相似、文本分类模型和类 SQL 因子，把原先偏离线、上线成本高的规则流程改造成低延迟实时引擎，显著缩短规则上线周期并降低人工成本。

- **2018.04 - 2019.03 数据监控预警系统**

作为主要开发者参与公司级指标监控平台建设，服务美图内部数据团队的指标异常检测；接入数亿级指标并调用时间序列异常检测模型，异常命中准确率 98%+；负责 Spark 数据接入/重跑/查询、Redis 信号管理、Livy 服务化和邮件等告警链路的关键工程模块。

## 大数据开发工程师

**创维酷开 (数据平台)** 2016-07 → 2018-03 深圳

作为早期数据工程成员参与数据部从无到有建设，是多个模块的核心开发者 / Owner，覆盖 DMP、BI ETL、实时流处理、标签体系和数据挖掘能力。

- **2016.07 - 2018.03 DMP 平台**

作为从 0 到 1 核心开发之一，建设面向运营和营销团队的用户标签、人群圈选和画像分析能力；平台覆盖约 2000 万用户、近百个标签，使用 HyperLogLog 做人群规模快速估算，结合 Spark UDF / UDAF 计算画像指标，并通过任务优化将新增标签作业缩短约 1 小时。

- **2016.07 - 2018.03 BI 平台与用户行为实时计算**

开发并维护用户播放行为表和多张播放相关 ETL 表，支撑数据分析师在 BI 平台产出面向管理层的报表；沉淀数十个常用 Hive UDF 提升分析效率，使用 Redis sorted set + Spark Streaming 计算智能电视端用户行为实时数据，并参与 K-means 用户聚类、协同过滤推荐等早期数据挖掘任务。

---

## Capability Highlights

### 大数据

**数仓 / 实时计算 / 数据分析 / 指标体系**

- 大数据工程底盘：十年经验覆盖数据平台、数据仓库、实时计算、用户画像、数据埋点和数据分析；使用包括 Spark、Flink、Kafka、Hive、HBase、Redash、Grafana、Kusto、Geneva 等通用技术栈，以及微软内部的 Cosmos、Titan、Power BI 等数据工具。
- PB 级数据处理的 MiPush 数据系统：负责实时/离线推送统计、安卓竞品分析和面向内外部开发者的统计报表；接手竞品系统后，通过清洗、映射和 OAID 兼容，将与 BCI 的统计误差从 50%+ 拉回到 10%。
- 数仓重构与成本优化：主导 MiPush 推送统计平台重构，把旧系统脉冲式写 HBase 和不合理数据结构改成更稳定的模型，减少 HBase 存储，每年节省近百万元人民币。
- 能影响业务决策的数据分析：为 Skype、Bing App、Growth / Acquisition、Copilot Mac 和 Connectivity 建 funnel、dashboard、AB scorecard、LTV / ROI 分析与数据驱动优化；典型结果包括 Skype 1v1 call 成功率 88% → 98%、Copilot Mac 新用户留存 24.0% → 31.7%、Voice TDR 46.3% → 31.93%。

- 监控预警与数据质量体系：从美图公司级异常检测平台（异常命中准确率 98%+）到 Microsoft Geneva / Grafana monitor 和 CMC 1DS telemetry-loss，能把生产问题拆成可追踪、可告警、可复盘的数据系统。

## 全栈开发

### 云服务 / 客户端 / 内部工具 / 开源贡献

- 数据密集型产品里的端到端工程能力：把原始 telemetry、云服务、dashboard、脚本和轻量产品界面串成 PM、工程师和数据同学能直接使用的工作流。
- Microsoft Azure 云、telemetry 与 Mac 开发经验：熟悉 Azure 云服务使用，负责过 Copilot client telemetry 埋点与指标设计；在 Copilot Mac 中参与 Mac feature 开发。
- 内部工具与 dashboard 开发：近两年持续开发 Python 为主、少量 JS / TS 的数据可视化、Kusto 查询助手、MCP 工具和数据查询 agent，主要服务自己和团队同学的日常分析、排障等。
- 开源项目里的全栈贡献：向 [NousResearch/hermes-agent](#) 合入超过 15 个 commit，覆盖 ACP、agent runtime、config、tools、Discord / Telegram / Desktop / Gateway；向 [openclaw/openclaw](#) 合入超过 8 个 PR，覆盖 messaging adapter、agent dispatch 和 transport。
- 自建产品实践：Fawn（育儿 Agent，暂未开源，包括服务端、Android 与 iOS App）、[voice-buddy](#)、[code-while-shit](#)、[dashboard-gen-skill](#)、[golden-flower](#) 等项目覆盖 CLI、前端、后端、agent 集成和 dashboard 交付；熟练通过 Vibe Coding 快速进入新技术栈并做出可用版本。

## AI Agent

### RAG / MCP / AI Native Workflow

- FY25 Microsoft AI School China 一等奖：AI Agents 赛道独立提交 *Copilot Mac Data Query Agent*，从 Microsoft 中国 AI 培训项目多个赛道的决赛项目中胜出。
- Copilot Mac Data Agent：在没有成熟数仓、只有原始 telemetry 的情况下，构建 KG-RAG agent 提升 Kusto 查询效率；技术栈包括 OpenAI Agent SDK、MCP-style tool calling、Neo4j、Kusto connector、Azure Container Registry，核心难点是元数据理解、指标定义、业务理解等。
- Agentic data workflow：把 KQL / SQL、dashboard 迭代、OCV 报告、AB 读数、single-user telemetry 调查等高频数据工作逐步迁移到 AI 辅助流程，用数据经验约束 agent 输出质量。
- Agent 源码与开源实践：独立维护 [Claude-Code-Source-Study](#)（1.5K+ stars / 530+ forks），并把源码阅读沉淀转化为 Hermes Agent、OpenClaw 等项目的真实合入贡献。
- AI 辅助复杂排障：通过 LLM 多轮辅助阅读 Edge Chromium 与 Picasso 等 40GB+ 规模的源码，支撑 CMC bot filter 和 terminal-event 丢失 RCA，把 agent 能力用于生产级开发和 debug，而不只是简单的 demo。

## Open Source Projects

## Claude-Code-Source-Study

1.5k+ stars / 536+ forks

个人源码研究与学习项目，围绕Anthropic在2026年4月“开源”的Claude Code版本的代码，做系统性源码解读与学习。中文社区影响力最高的 Claude Code 内核研究参考工具书籍之一。

## Hermes Agent

15 个 main commit · ACP / agent runtime / config / tools / multi-channel

向 [NousResearch/hermes-agent](#) main 分支合入贡献，覆盖 ACP session、agent runtime、配置校验、工具调用，以及 Discord / Telegram / Desktop / Gateway 等多 channel 集成问题。

## OpenClaw

8 个 merged PR · messaging adapter / agent dispatch / transport

向 [openclaw/openclaw](#) 合入多项修复，集中在 GoogleChat / Matrix adapter、active memory deadline、ACP dispatch path、OpenRouter transport、embedded session takeover 等真实工程问题。

## AI-Native Productivity Tools

Python / TypeScript / HTML · CLI / dashboard / agent workflow

围绕AI coding workflow 持续做工具实验，包括 [voice-buddy](#)、[dashboard-gen-skill](#)、[skila](#)、[watch-claw](#)、[golden-flower](#) 等；重点是把想法快速推进到可运行、可复用的小工具。

## GitHub 公开工程信号

源码研究 / AI Agent 工具 / 数据可视化 / CLI / 小型 Web 应用

公开项目覆盖源码研究、AI Agent 工具、数据可视化、CLI、小型 Web 应用和游戏实验；主要公开语言包括 Python、TypeScript、JavaScript、HTML、Shell、C# 和 GDScript。

## Skills

### AI / Agents

LLM agent architectures

OpenAI Agent SDK

MCP (Model Context Protocol)

KG-RAG / Knowledge Graph + RAG Neo4j

Claude Code internals

Agentic workflows

### Data Engineering

Spark (batch & streaming) Flink Kafka

HBase Hive / HDFS Kylin / Doris

Kusto (KQL) Cosmos / Titan

PB-scale pipelines

Prompt engineering & token economics

### Analytics & Experimentation

A/B testing & scorecards

Funnel analysis (call / voice / chat)

LTV modeling

Attribution

OCV / user feedback analytics

Power BI · Grafana · Redash

### Observability & Telemetry

OneCollector / 1DS pipelines

Geneva / ICM

Cross-platform telemetry schema design

Event instrumentation & validation

SLO / quality dashboards

### Backend / Full-stack

Python

TypeScript / Node.js

Shell

Azure / Azure Container Registry

Docker

Web / CLI tooling

REST / async services

### Languages & Tools

Git

Markdown

Maven

PySpark

SQL

Liquid / Jekyll

Godot (GDScript, for hobby work)

## Awards

### 一等奖 · FY25 AI School China

Microsoft · 2026-02

AI Agents 赛道独立提交 *Copilot Mac Data Agent (KG-RAG)*，并在 Microsoft 中国 AI 培训项目决赛中胜出；决赛项目覆盖 AI Agents / RAG / Fine-tuning / Multimodal 等方向。

### Transformational Impact · FY25 最高等级绩效

Microsoft · 2025

因在 Copilot Mac 数据基础设施建设中的核心贡献获得 FY25 最高等级绩效评价；工作覆盖 telemetry、dashboard、scorecard 与质量 review 机制，成为团队产品质量和增长分析的重要基础。

### 破格晋升16级

小米 · 2019-2021

因小米换机数据和 Android 竞品数据体系建设连续获得 A 级绩效；一年内从 15 级晋升至 16 级，快于常规晋升节奏。

## Education

湖南大学

985 · 信息与计算科学 · 2012-09 → 2016-06 · 长沙

---

## Languages

---

🇨🇳 中文 — 母语

🇬🇧 英文 — 工作流利（日常技术读写）

---