

## 个人简历

**经历摘要:** 9 年前端与全栈工程经验, 近期聚焦 AI Agent 工程与应用落地, 实践覆盖上下文工程、RAG 检索、Tool / Skill / Sandbox 设计、Agentic UI、Tracing 及评测闭环等关键环节。持续使用 AI Coding 提升开发效率, 能够以 Owner 或核心贡献者身份完成从方案设计到全栈交付。

## 工作经历

美团 企业平台 · 高级 Web 开发工程师

2021.11 - 今

### ASR 产品评测与模型选型 · 2026.06 · 评测 / ASR / 后处理

模块 Owner

**挑战** 语音输入法需要在多个 ASR 模型和多版本后处理方案间持续选型与迭代, 但缺乏统一评测口径: ASR 评测既要覆盖 CER、句准等客观指标, 也要衡量中英混合、场景泛化和主观体验; 基准集与评测体系建设复杂, 依赖主观判断会放大迭代风险。

**动作** 搭建 Python + React 统一评测平台; 从 App 真实语料中筛选并标注 1,100 条通用用例和 200 条中英混合用例, 完成 18 个 ASR 模型在 CER、句准和场景表现上的横向对比; 针对后处理质量难以客观衡量的问题, 设计端到端 LLM Judge 评测框架, 按准确性、纠错能力、格式、流畅度等维度加权评分, 并与 Typeless 竞品进行胜 / 平 / 负对比, 将主观体验转化为可量化的迭代依据。

**结果** 评测结论支撑产品采用“流式快速响应 + 段式最终校准”的双模型方案; 后处理方案在 900 余条评测用例上持续迭代, 综合评分由 8.75 升至 9.26, 优秀率由 64.1% 升至 86.3%, 相对竞品胜率由 57.0% 升至 69.1%。

### Agent Skill 生态建设 · 2026.03-2026.05 · 评测体系 / Sandbox

核心贡献者

**挑战** 公司级 SkillHub 汇聚数万 Skill, 在规模化分发至各 Agent 平台前面临两类瓶颈: 缺乏可量化的质量准入机制, Skill 质量参差不齐, 难以建立统一信任标准; 各平台执行环境割裂, 接入成本高且基础能力难以复用, 制约了 Skill 生态的横向扩展。

**动作** 质量侧, 建设触发率、CLI 独立可用性、端到端功能用例三类评测, 并将结果接入发布流水线, 形成 Skill 上架前的质量门禁; 执行侧, 建设统一多租户 Sandbox 管理服务, 将 E2B 沙箱能力抽象为 HTTP / RPC API, 覆盖 Skill 发现、安装与加载、鉴权、依赖管理和工作区持久化, 降低各业务 Agent 重复建设执行环境的成本, 支撑 Skill 跨平台接入; 可观测侧, 为 Skill 评测引入 Trace, 提供 Skill CLI 进程与请求的可视化瀑布流, 辅助开发者定位执行过程中的性能瓶颈和调用链问题。

**结果** 完成约 90 个企业平台核心 Skill 的评测并接入发布准入; 评测与执行能力可覆盖 SkillHub 内 3 万至 4 万个 Skill, 已支撑约 700 个认证 Skill 跨业务 Agent 接入; 沙箱启动 TP90 由 3 秒降至 0.5 秒, 热沙箱命中率达 99%, 冷启动占比低于 1%。

### 工单平台 Agent 初版架构 · 2026.03-2026.05 · 工具路由 / RAG / 意图识别

架构负责人

**挑战** 工单平台涵盖上千团队服务目录, 每个目录绑定不同的知识库工具与办事技能, 全量工具直接暴露给 LLM 既受限于上下文容量, 也难以保障工具选择的准确性; 在大规模异构候选空间中实现高效意图识别与工具路由, 是 Agent 可用性的核心瓶颈。

**动作** 设计“意图识别 + RAG 目录检索”两阶段路由: 先通过目录级路由锁定团队服务范围, 将候选工具规模收敛至常量级; 基于精选的真实问答用例, 离线为每个服务目录生成语义画像, 并设计关键词标签体系, 通过向量检索与关键词搜索混合召回, 结合目录名、描述进行多路匹配, 兼顾语义覆盖与精准命中。

**结果** 以目录级路由替代全量工具暴露, 有效降低 LLM 上下文压力; 架构可随服务目录规模横向扩展, 为知识库问答和办事类 Skill 接入奠定基础。

### 数据分析 Agent 端到端建设 · 2025.06-2025.12 · Agent / Harness

方向 Owner

**挑战** 受数据合规约束, 方案只能使用内部或非 SOTA 开源模型, 模型能力受限; 数据分析又要求结果严谨且可复现。如何通过可执行代码、评测体系与反馈闭环约束 LLM 的概率性输出, 是项目落地的核心挑战。

**动作** 架构层面, 先以 SQL 作为 DSL 桥接 LLM 与数据源; 在遇到表达能力和输出稳定性瓶颈后, 演进为 Python Code Interpreter + Sandbox, 发挥 LLM 的代码生成能力。推理模式从 Plan + Execute 演进为 ReAct, 通过单步推理、行动与反思循环提升预测、归因等复杂场景的灵活性。独立建设 Agent Trace 与评测体系, 覆盖离线用例集、线上失败样例回收和 LLM Judge, 形成迭代验证与评测集持续增长的闭环。在输入处理侧解决百万行 Excel 性能瓶颈, 引入 LLM 启发式识别非标准表头, 并落地 Agent 安全护栏、Working Memory 管理与工具效率优化。

**结果** 能力落地到低代码表单记录分析、在线多维表格 AI 分析、公司级 IM 数据分析子 Agent 和审批工作流分析节点, 覆盖约 1,000 名用户; 评测体系积累约 200 条用例, LLM Judge 优秀率约 95%, 在非 SOTA 模型约束下实现了可迭代、可度量的 Agent 交付。

## 流程低代码平台架构升级与工程化 · 2022.09-2025.05 · 低代码引擎 / 工程化

核心贡献 / 模块 Owner

**挑战** 企业办公域低代码平台以 PC / H5 审批表单搭建为起点，需要在避免重复建设的前提下，将引擎能力扩展至 IM 消息卡片、AI 流式卡片、AI 助手技能卡片等异构场景，并支持跨平台渲染。同时，高流量 IM 触达产品的稳定性与体验指标口径尚未建立，需要在跨端、跨团队协作中从 0 到 1 完成体系建设。

**动作** 推动一套引擎统一支撑审批表单、IM 消息卡片、AI 流式卡片和 AI 助手技能卡片的搭建与跨平台渲染；发起属性设置器重构，改进面向 Schema 的架构设计，提升配置效率并降低维护成本；发起编辑器核心模型重构，通过交互层抽象使表单与卡片搭建共用同一套编辑器能力；作为模块 Owner，负责高流量 IM 触达产品稳定性与端到端体验指标体系的目标拆解、方案设计和跨团队落地。

**结果** 低代码平台由 PC / H5 审批表单扩展至 IM 富交互消息卡片等场景，支撑办公域产品交互形态演进；统一交互模型支持多形态搭建复用，并为后续业务扩展保留架构空间；可观测性体系从 0 到 1 落地，为高流量产品提供稳定性保障与持续的指标运营能力。

## 英语流利说 主 App 团队 · 高级 Web 开发工程师

2019.08 - 2021.10

**Web 阅读器架构:** 独立负责书籍阅读产品 Web 版，完成富文本处理与多屏精准分页，并在中低端设备上平衡分页精度与渲染性能，保障阅读内容在不同尺寸设备上的稳定呈现。

**低代码搭建工具:** 独立设计可插拔、可版本化的物料组件体系与可视化编辑器，累计支撑 1,000 余个单页搭建，显著缩短投放引流类页面的开发周期。

## 美团 到店综合业务部门 · Web 开发工程师

2018.02 - 2019.08

**业务与团队交付:** 负责点评丽人业务线商家端产品迭代，指导两位初级工程师完成项目交付与能力成长，并以业务成果和技术项目支撑团队成员晋升。

**运营提效工具 0-1:** 从 0 到 1 设计并实现 Banner 模板图生成工具和跨技术栈分享图生成工具，实现素材复用、程式化编排与多端出图，减少设计重复制图时间 60% 以上，并降低相关研发工作量约 50%；Banner 工具后续演进为公司级公共服务并移交设计工具团队。

## 早期经历

2014.08 - 2018.02

Ezbuy · 前端开发工程师

2017.05 - 2018.02

AspenTech · 非软件研发岗位

2016.05 - 2017.05

沪东中华造船集团 · 非软件研发岗位

2014.08 - 2016.05

## 教育背景

华东理工大学 · 硕士 化学工程与工艺(2011.09-2014.07) · 本科 油气储运工程(2007.09-2011.06)

上海

## 专业技能

**AI 应用工程** AI Agent、Agent Harness、Tool / Skill 调度、Sandbox、Tracing、Eval、LLM Judge、Prompt 迭代、RAG、MCP、AI Coding。

**全栈与前端架构** React、TypeScript、Node.js、Python、Vue、SSR、Express、前端架构、低代码引擎、可观测性、自动化测试、构建工具。

**平台与数据能力** Vercel AI SDK、Mastra、Langfuse、E2B Sandbox、关系型数据库、K8s、ECharts。