

孔琳凯

15535787215 | kk2804417203@163.com | 北京
AI产品经理/FDE工程师



个人总结

AI Native Builder, 计算机科学与技术专业出身, 先后从事过大模型算法、Agent产品经理、大模型运营等岗位, 擅长把模糊的业务问题拆成 Agent 可以执行的规则、上下文、Workflow、Tool 和评测体系, 并通过离线评估、Badcase分析与数据反馈持续推动Agent迭代, 形成数据-产品飞轮; 具备卓越的学习能力和实践能力, 热衷于与Agent交互, 具有敏锐的洞察力与好奇心。

核心能力

- Agent交付:** 擅长将模糊业务目标沉淀为规则、上下文、Workflow、Tool、Skill, 构建可复现、可规模化的 Agent 交付系统。
- Agent 评测:** 搭建Judge Agent、Trace、Badcase 闭环, 通过单变量优化、回归评测和数据反馈推动 Agent 持续 loop进化。
- Vibe Coding:** 熟练使用 Codex、Claude Code完成需求分析、原型开发、自动化编排与验证, 并将方法沉淀为可复用 Skill。
- Agent-first 产品设计:** 以 Agent 能力边界重构任务、流程、交互与人机协作, 推动方案从需求定义走向真实业务规模化交付。

工作经历

北京字节跳动科技有限公司-大模型运营 CQC

2025.12 - 至今

负责将内容审核、业务咨询等人工密集型场景转化为可运营的 Agent 交付链路。以万 Case 成本、自动处理率、质量与时效为共同经营指标, 协同上下游产品、策略与交付团队推进模型能力规模化落地。

核心指标一: 降低万case成本, 通过搭建审核Agent提升自动审核占比

- 面向抖音、汽水音乐、可颂等业务的审核需求, 完成规则重构、数据集建设、Prompt / Workflow 设计、离线评测、Badcase 归因与灰度迭代, 形成可复用的自动审核交付路径。
- 推动自动审核占比由 16.96% 提升至 39.03% (+22.07pp), 实现约 11 HC 的人力节降与 3.5 万元+/月 的成本优化; 持续以万 Case 成本和线上稳定性校验规模化收益。

核心指标二: 利用AI工具提效, 节降运营人员hc及工时

- 负责从需求调研、知识治理到交互与运营闭环的产品化设计, 搭建“意图识别 → Query 改写 → 语义+关键词混合检索 → LLM 总结 → 人工兜底”的智能客服Agent 链路, 并通过多轮版本迭代解决长尾咨询与知识更新问题。
- 将离线评测、用户反馈与知识库维护纳入同一运营闭环, 保障不同业务接入后的回答一致性、响应速度与风险可控性; 完成多个业务场景复用与 50+ 场业务培训/推广。
- 在已落地场景中, 实现 96% 回答准确率、15 秒内响应; 业务接入周期由 21 天缩短至 7 天, 月均支持 10 万+ 次查询, 节省 852 小时/月, 相当于约 12 HC 人力。

NodeskAI-AI产品经理 创新产品

2025.06 - 2025.11

面向电商大促期间高并发咨询、人工接待能力不足及商家知识维护成本高等问题, 建设融合多轮对话、RAG 知识检索与人工接管的客服 Agent, 使其同时承担咨询解答、导购推荐与转化承接。

负责内容:

- 产品链路:** 参与从 0 到 1 规划客服 Agent, 将 1000+ FAQ 归一为约 100 个结构化知识模板, 搭建“意图识别—槽位抽取—知识检索—工具调用—回答生成—安全校验—人工接管”完整链路。
- 评测底座:** 基于 DeepEval、Ragas 与 Langfuse 建设 Golden Set、RAG 指标和多轮会话 Trace, 固定模型、Prompt、知识库及工具版本, 支持批量评测与版本对比。
- 自动评测:** 构建“客服 Agent 执行任务、Judge Agent 按 Rubric 评分”的双 Agent 机制; 确定性任务采用规则/API 校验, 开放式回答评估正确性、完整性、帮助性、语气与安全性, 并通过人工抽检校准 Judge 一致性。
- 迭代闭环:** 将 Badcase 归因至意图、槽位、知识、工具、生成和安全兜底, 形成“评测—归因—优化—回归”Loop, 持续驱动 Agent 版本迭代。

核心成果: 618 期间实现 83% AI 接待率、95% 用户满意度、4.45 秒平均响应; AI 群接待使转化率提升 5.5pp, 冷启动配置周期由 30 人天缩短至 3 人天, 知识维护成本下降 90%, 支撑日均 1 万 UV 的咨询场景。

项目经历

Kongkong Harness(自研) | 让 Agent 稳定交付并持续进化的工程框架

背景: TCS 自动审核交付横跨队列、模板、数据送审、插件、Workflow、评测与发布: 业务同学认为 CLI 门槛高, 模型运营仍依赖手工搭建和个人 PE 经验。调研发现约 80% 同学未使用 TCS CLI, 已使用者也因缺少统一规范、版本管理和验证协议, 导致同类需求交付效果不一致、经验难复用。

动作：

- **方法设计：**提出“PE 1.0 解决模型理解规则—Harness 2.0 解决 Agent 稳定交付—Loop 3.0 解决 Agent 持续进化”的三阶段方法，并设计业务需求、方案选型、规则结构化、PE 翻译、Workflow 编排、工程管理、质量反馈七层架构。
- **工程实现：**将 TCS、Jimu、评测与发布能力沉淀为可复用 Skills，使 Agent 能完成需求审计、队列/模板创建、Workflow 搭建与修复、单/批量测试和 Trace 回溯，将分散操作组织为可追溯、可复现、可回滚的标准交付链路。
- **质量 Loop：**以人工 GT 为基准，驱动 Agent 执行 Badcase 归因、单变量优化、回归评测与发布门禁，形成“Agent 搭 Agent、Agent 评 Agent、Agent 迭代 Agent”的持续优化闭环。

成果：在团播话术场景中，1.5 小时完成从队列、模板到数字员工的端到端交付，处理 2.7 万条数据，实现 100% 准确率、98% 召回率，节省约 90 人天；另一场景 30 分钟完成 Agent 搭建与批测，处理 2.4 万条数据、节省约 30 人天。在可颂优化场景中，将原需 2—3 周的人力调优压缩至 1 天，精确率达到 93%，过滤率由 33% 提升至 57%。

V5-AI Native | 基于Harness与原子能力编排重构自动审核

背景：原有 V5 按 24 个业务垂类分别搭建 24 个 Agent，将人审 SOP 逐一翻译为 PE；但输入仅有 1 分钟随机切片、ASR、OCR 与标题，输出却要覆盖 418 个三级标签。相同能力被重复建设，跨垂类边界 Case 要么被强行判定、要么全部退回人工，Agent 难以复用，人仍是发现问题、归因和迭代的全链路瓶颈。

动作：

- **能力抽象：**从 29 个一级、122 个二级、418 个三级标签中抽象出电商直播、主播身份、语义理解、本地生活、非真人识别 5 类原子能力，把按垂类重复建设转为按能力复用。
- **多 Agent 编排：**分别搭建 5 类原子能力子 Agent，仅输出结构化能力结论；由主 Agent 负责能力路由、信号组合、规则决策与统一输出，实现子 Agent 独立迭代、跨垂类编排。
- **人机协作与演化：**将人审规则改写为面向 Agent 的证据要求、判定边界与冲突优先级；让 Agent 主动识别证据不足、规则冲突、垂类交叉和定义模糊四类边界 Case，确定 Case 自动提交，不确定 Case 携带证据与原因回流人工仲裁，仲裁结果再进入 Harness 完成规则、PE 与评测集迭代。

成果：完成从“24 个垂类分别建设 Agent”到“5 类原子能力子 Agent + 主 Agent 编排”的架构重构；新垂类接入由重新搭建 Agent 转为配置能力组合与规则包，原子能力可独立版本管理、评测和跨垂类复用。人机分工由逐 Case 审核转为“Agent 处理确定 Case、人工仲裁边界 Case”，并通过结构化回流将迭代周期由周级缩短至天级。

教育经历

长春理工大学-计算机科学与技术 本科 全日制 2021.09 - 2025.06

- **在校期间Rank：**3/289，获得本校推免直博资格，获得推免至天津大学、复旦大学保研资格
- **荣誉：**国家奖学金、国家励志奖学金、全国大学生数学竞赛国家级一等奖、中国大学生计算机设计大赛国家级二等奖、蓝桥杯全国软件和信息技术专业人才大赛国家级二等奖