

我能做什么 WHAT I DO

从一个想法出发几小时搭出能跑的原型，也能设计评估方法判断它值不值得活下去。身份上在 Chase 定义的 Builder 和传统 PM 中间地带：能自己 Build 做验证，也能设计标准判断。

– 深度使用 **claude code**, **claude.ai**, **claude app**, **claude design**, 搭配 **graphify** 加强的 LLM wiki——把验证一个想法的周期 **从 3 个月压到 1 天**。

– 让评估维度从数据里长出来。AI 问答 trace 里浮现出 11 个维度（3 个是预设 rubric 想不到的）。

– 从判断到真上线。qibook 从查企业扩大到查“人”的新业务线，4 次碰壁每次用实验数据转向，反向蒸馏用户品味，**200 字三段论上线 qibook.com**。

教育 EDUCATION

蒙特雷国际研究院 MIIS

翻译与本地化硕士 (STEM)

2023.09 – 2024.06 GPA 3.9 / 4.0

2023–2024 走北外派出交换至 MIIS

北京外国语大学

会议口译硕士

2022.09 – 2025.06 GPA 3.8 / 4.0

厦门大学

英语文学学士 (Top 5%)

2018.09 – 2022.06 GPA 3.9 / 4.0

技能 SKILLS

AI 与 Agent

Claude Code / Claude App / Claude.ai 深度使用 · MCP 协议 · Multi-Agent 编排 · Prompt 工程 · 评估方法论设计 · Langfuse · Coze · LangGraph

产品

PRD · JTBD · A/B 测试 · 标注体系设计 · 意图分类 · 情景模拟 · 用户研究 · 跨团队推动 · 数据需求方法论

语言

CATTI 一级笔译 · 二级口译 · 中英双语原生

独立产品 INDEPENDENT

AI PMF Validator

基于 OASIS 开源平台搭 AI 虚拟焦点小组—验证产品想法商业可行性，为每一个人提供轻量商业咨询。

人生模拟器

平行人生模拟器让你探索那些“没走的路”。输入你的真实性格、经历和生活条件，观察自己平行人生的多种可能性。对话式人格采集 --> AI 人格档案生成 --> 5 层变量系统 --> 人生模拟与人生分支 --> 分层记忆与记忆可视化 --> 跟任意年龄的“世另我”对话/多版本互相对话

deep-decode 内容管线

选题到 四件套产出自动化: 骨架 → 草稿 → 信息图 → Word → TTS 播客。三源交叉验证防单一视角偏差。

完整故事、每个项目的 2000+ 字长文、设计决策复盘，都在 mingmingzi.com。

工作经历 EXPERIENCE

中数智汇 Chinadaas / AI 产品经理

JUL 2025 – PRESENT

企百科对话产品评估体系 自主发起

企百科上线了 AI 对话但没有上线标准——PM 说“答案不好”没数据依据，研发改完 prompt 无从验证变好变差。Rubric 保质期短：业务从“报告”变到“对话”就废一批维度。

用 **标注体系设计** (按工具调用链路分成简单商查 19 + 复杂商查 19 = 38 个涌现类目, 6–9 轮迭代冻结) + **三模型共识投票** (Qwen + GLM + Kimi 共识; Qwen3.6 + GLM-5.1 + DeepSeek-V3.2 多数票 2/3) + **LLM-as-a-judge** 自动化链路, 把评估从 PM 拍脑袋变成可复用流水线。Langfuse 上让评估维度从 10 条 Agent trace 中自然涌现出 11 个 (3 个是预设 rubric 想不到的)。

→ 发现 66.8% 满足率的瓶颈不在答案质量而在 68.2% 的误拒, 把产品团队的优化重心从答案质量翻到意图识别层, 单项潜力 +23%。

养一只虾: 从 chat bot 到会主动联系我的代理 OpenClaw + PM 分身 Amory

AI 工具在 terminal 和网页里、日常工作在飞书——两边不通, 灵感来了得切窗口找工具 (问题)。部署踩三层坑: API 限流、SSL 与权限、弱模型编排强模型会拆错任务 (编排者的天花板就是系统的天花板)。

用 **心跳协议** (评分公式 $\text{impact} \times \text{confidence} - \text{cost}$) 让 Agent 自主扫 8 个数据源决定做不做 + 熔断机制防 token 烧穿; PM 分身用 **行为考古**——从 Claude Code 会话和 Redmine 日志归纳自己真实高频动作 (最高频是上传附件和更新需求字段, 不是写新需求), 筛出汇报 / 发需求 / 数据分析等场景做成 Skill。

→ 飞书 @ 一句话跑完以前 30 分钟手动汇报流程; 把我从被动切窗口改成被 Agent 主动推动。

qibook.com 个人工商简报 查“人”新业务线 · 0→上线

公司只能查企业, 但 7.8% 的用户需求是“拿人名查关联企业”——没人做过的新业务线 (问题)。纯 LLM 生成 100% 幻觉 (EO 实验验证), 22 版可视化无法收敛, 5 个银行部门定制版 PMF 5.25 分全 Dead End。

用 **反向蒸馏** (综合行业报告 + LLM 模拟银行客户经理视角, 把模糊认知翻成 6 类身份标签) 定义分类体系; 用 **data-req 四步数据需求方法论** (盘清 19 API / 1445 字段 → 从 1245 条会话做意图分类 → 缺口矩阵 → 数据需求文档) 打通数据到产品全链路。

→ 200 字三段论让银行客户经理 10 秒看懂一个人的商业角色和版图分布, 为公司从“查企业”扩出“查人”新业务线。

企百科 AI 搜索可见性 (GEO) 自主发起

在豆包搜“企百科”第一页只出应用商店——就绪度 F 级, 0.67/10。最根本障碍不是 SPA 空白渲染, 是网站对外关闭——说服 B2B 应用层团队“为什么要关心 AI 搜索”是组织问题不是技术问题。

写 SPA 语义注入方案 (JSON-LD + 语义 HTML + sitemap) 被研发采纳, 跨团队推动把网站对外打开; **搜索引擎逆向分析** (246 条 AI 搜索结果按域名统计) 发现外部占比悖论——千问引用中外部平台占 70%+; **autoresearch 方法** (Karpathy 框架, AI 自己跑 A/B 迭代评分) 优化软文 Skill, 把软文评分从 7.0 推到 12/12 满分。

→ SPA 语义注入 + 网站开放已落地, 让企百科从 AI 搜索完全不可见变成具备被索引的基础条件。

中数论坛 知识飞轮 + 方法论反思

公司内部知识散落在个人脑袋和飞书群聊里, 没有沉淀复用机制。冷启动是核心困难——上线后 8 用户 2 赞, 绝大多数帖子是我自己发。

设计 **Git 到 Agent 的语义协议** (普通虾 4 动作 push/comment/pull/star, 管理员虾 6 动作含 review/merge/rebase), Discourse 部署 + 论坛心跳 API 对接完成。

→ 真正的收获是一条方法论反思: 产品设计的正确顺序是用户真实行为 → 流程想象 → 产品设计, 从 UI 反推是死路。

标讯推荐 投标书 + 标讯 LLMwiki

同事要在 2 天内交一份 176 万投标项目的 9 章投标书初稿; 由此引出: 怎么找到合适的标? 现有剑鱼千里马都在做推荐, 纯拼“更准”没差异化空间。

用 Claude 一个 session 产出完整投标框架; 用 **情景模拟** (虚构用户老王的一天) 重新定位——用户不是“查标讯的人”而是“经营一家投标公司的人”; 设计 **画像自耕田机制** (决策因子 × 用户偏好双轴, AI bookkeeping 人 curation, 用户不断上传文件和提问, 标讯推荐越来越懂用户, lint 功能淘汰过期标/客户/竞对)。

→ 把标讯系统从信息推送重新定位为“经营助手”, PM 角色扮演用户从记忆出发设计产品。

Ming Mu

AI Product Manager

From idea to prototype · Let knowledge compound and grow

mingmingzi.com

sonikming@gmail.com / 119532205@qq.com

linkedin.com/in/amorymu

github.com/AmoryMing

WHAT I DO

I can go from an idea to a working prototype in a few hours, and I can design the evaluation method that tells us whether it deserves to live. My niche sits between Chase's "Builder" definition and a traditional PM: I build to validate, and I design the standards to judge.

- **Deep user of Claude Code, Claude.ai, Claude App, and Claude Design**, paired with a Graphify-augmented LLM wiki — compressing idea validation from 3 months to 1 day .
- **Let evaluation dimensions emerge from the data.** Eleven dimensions surfaced from real Q&A agent traces — three of them impossible to anticipate with a pre-written rubric.
- **From judgment to shipped product.** Expanded qbook from company lookup to a new "people lookup" line: four dead ends, each pivot driven by experimental data, reverse-distilled user taste into a **200-word, three-paragraph brief that shipped on qbook.com** .

EDUCATION

Middlebury Institute of International Studies (MIS)

M.A. in Translation & Localization Management (STEM)

Sep 2023 – Jun 2024 GPA 3.9 / 4.0

Sponsored exchange from BFSU to MIS, 2023–2024.

Beijing Foreign Studies University (BFSU)

M.A. in Conference Interpreting

Sep 2022 – Jun 2025 GPA 3.8 / 4.0

Xiamen University

B.A. in English Literature (Top 5%)

Sep 2018 – Jun 2022 GPA 3.9 / 4.0

SKILLS

AI & Agents

Daily-driver use of Claude Code / App / Claude.ai · MCP protocol · multi-agent orchestration · prompt engineering · evaluation methodology · Langfuse · Coze · LangGraph

Product

PRDs · JTBD · A/B testing · annotation-taxonomy design · intent classification · scenario simulation · user research · cross-team influence · data-requirement methodology

Languages

CATTI Level 1 (Translation) & Level 2 (Interpreting) · Native-level Chinese and English

INDEPENDENT PROJECTS

AI PMF Validator

Built an AI virtual focus group on the open-source OASIS platform — stress-tests the commercial viability of a product idea, giving anyone lightweight business consulting on demand.

Life Simulator

Explore the paths you didn't take. Conversational personality capture → AI persona profile → 5-layer variable system → life simulation with branching → layered, visualizable memory → converse with a "parallel you" of any age, or let multiple versions talk to each other.

deep-decode Content Pipeline

Automates the whole path from topic selection to a **four-format output**: skeleton → draft → infographic → Word doc → TTS podcast. Three-source cross-verification prevents single-viewpoint bias.

Full stories, 2,000+ word deep-dives per project, and design post-mortems live at mingmingzi.com.

EXPERIENCE

Chinadaas / AI Product Manager

JUL 2025 – PRESENT

Evaluation System for Qibaikē Conversational Product Self-initiated

Qibaikē launched its AI chat experience with no acceptance bar — PMs said "the answers are bad" without evidence, engineers tweaked prompts with no way to tell if things got better or worse (**the problem**). Rubrics have short shelf lives: moving from "reports" to "dialogue" invalidated an entire batch of dimensions.

Combined **annotation-taxonomy design** (split by tool-call paths into 19 simple + 19 complex query types = 38 emergent categories, frozen after 6–9 iterations), **three-model consensus voting** (Qwen + GLM + Kimi for query verdicts; Qwen-3.6 + GLM-5.1 + DeepSeek-V3.2 majority vote 2/3 for coverage), and an **LLM-as-a-judge** automated pipeline — turning evaluation from a PM's gut call into a reusable assembly line. On Langfuse, 11 evaluation dimensions emerged naturally from 10 real agent traces (3 of them impossible to foresee in a pre-written rubric).

→ Revealed that the 66.8% satisfaction-rate bottleneck wasn't answer quality but a 68.2% false-rejection rate, redirecting the product team's optimization focus from answer quality to intent recognition — unlocking a +23% single-lever upside.

"Raising a Space Lobster": From Chatbot to an Agent That Reaches Out to Me

OpenClaw + PM alter-ego "Amory"

AI tools live in the terminal and browser; day-to-day work happens in Feishu — the two don't talk, and every flash of inspiration meant switching windows to find the right tool. Deployment hit three layers of pain: API rate limits, SSL and permissions, and a weaker model orchestrating a stronger one kept decomposing tasks wrong — the **ceiling of the orchestrator is the ceiling of the system**.

Designed a **heartbeat protocol** (scoring formula: impact × confidence – cost) so the agent autonomously scans 8 data sources and decides whether to act, plus a circuit breaker to prevent token burn. The PM alter-ego was built through **behavior archaeology** — mining Claude Code sessions and Redmine logs for my actual high-frequency actions (the top ones were uploading attachments and updating ticket fields, not drafting new requirements), and packaging reporting / requirement-writing / data analysis into reusable Skills.

→ An @-mention in Feishu now runs in one sentence what used to be a 30-minute manual reporting flow; the agent has shifted me from reactively hunting for tools to being proactively nudged by the agent.

qbook.com — Personal Business Brief New "people lookup" line · 0 → shipped

The company could only look up *companies*, yet 7.8% of user queries were "give me a person's name, show me the companies behind them" — a greenfield line no one had built. Raw LLM output was 100% hallucination (verified in the EO experiment); 22 visualization iterations failed to converge; five bank-specific versions all dead-ended at PMF scores of 5.25.

Defined the taxonomy via **reverse distillation** — synthesizing industry reports with LLM simulations of bank relationship managers to translate fuzzy intuition into six role labels. Used my **four-step data-requirements methodology** (inventory 19 APIs / 1,445 fields → **intent classification** from 1,245 real conversations → gap matrix → data-requirement doc) to wire data end-to-end into the product.

→ The shipped 200-word, three-paragraph brief lets a bank relationship manager grasp a person's commercial role and portfolio in 10 seconds — opening an entirely new "people lookup" business line on top of the existing company product.

Qibaikē AI Search Visibility (GEO) Self-initiated

Searching "Qibaikē" on Doubao returned only app-store listings on the first page — AI readiness grade F, 0.67/10. After diagnosing, I pitched it up the chain. The deepest blocker wasn't the SPA's empty render — it was that the site wasn't publicly accessible. Convincing a B2B, application-layer team to care about AI search was an **organizational problem, not a technical one**.

Wrote the SPA semantic-injection spec (JSON-LD + semantic HTML + sitemap) and got it adopted by engineering; drove cross-team alignment to open the site to the public web. **Reverse search-engine analysis** (246 AI search results grouped by source domain) exposed the "external placeholder paradox" — external platforms made up 70%+ of Qwen citations. An **autoresearch loop** (Karpathy framework; the AI runs its own A/B scoring iterations) pushed advertorial scores from 7.0 to a perfect 12/12.

→ SPA semantic injection and public access both shipped — moving Qibaikē from invisible in AI search to fundamentally indexable.

Chinadaas Internal Forum Knowledge flywheel + methodology reflection

Internal knowledge lived in people's heads and scattered Feishu groups, with no mechanism for it to accumulate or be reused. Cold-start was the core difficulty — after launch: 8 users, 2 likes, and most of the posts were my own.

Designed a **Git-to-agent semantic protocol** (regular "shrimps" get 4 actions: push / comment / pull / star; admin "shrimps" get 6, including review / merge / rebase). Deployed on Discourse and integrated with the forum heartbeat API.

→ The real takeaway was a methodology lesson: the correct direction for product design is **real user behavior → imagined flow → product design** — reasoning backward from UI is a dead end.

Bid Recommendation Proposal drafts + bid-announcement LLM wiki

A colleague had 2 days to deliver the first draft of a 9-chapter proposal for a ¥1.76M bid; that surfaced the upstream problem — how do you find the right bids to chase? Incumbents (Jianyu, Qianlima) were all competing on "more accurate recommendations," leaving no differentiation headroom.

Used Claude in a single session to produce the full proposal skeleton. Used **scenario simulation** (a fictional day in the life of user "Lao Wang") to reposition the product — the user isn't "someone checking bid announcements" but "someone running a bidding company." Designed a **self-tending profile mechanism** (decision-factors × user-preferences on two axes; AI does bookkeeping while humans curate; users keep uploading docs and asking questions so recommendations grow sharper; a lint function retires stale bids, clients, and competitors).

→ Repositioned the bid system from "information push" to "business operations assistant" — designing the product through the PM role-playing the user's lived memory.