薛凤翔

AI 产品经理作品集

聚焦 Agent 产品、AI 评测与复杂知识工作流

Agent 产品设计　多 Agent 协作　AI Evaluation　模型路由

手机/微信 15618179903 · xuefx123@163.com · github.com/ChloeXue00

PROFILE

把 AI 能力变成可理解、可控制、可验证的产品体验

不是堆功能入口，而是围绕真实任务建立从输入、理解、执行到交付的完整闭环。

3	103	4	3 层
核心产品模块	评测题与标准答案	主对话关键阶段	用户 / Agent / 项目记忆

核心能力

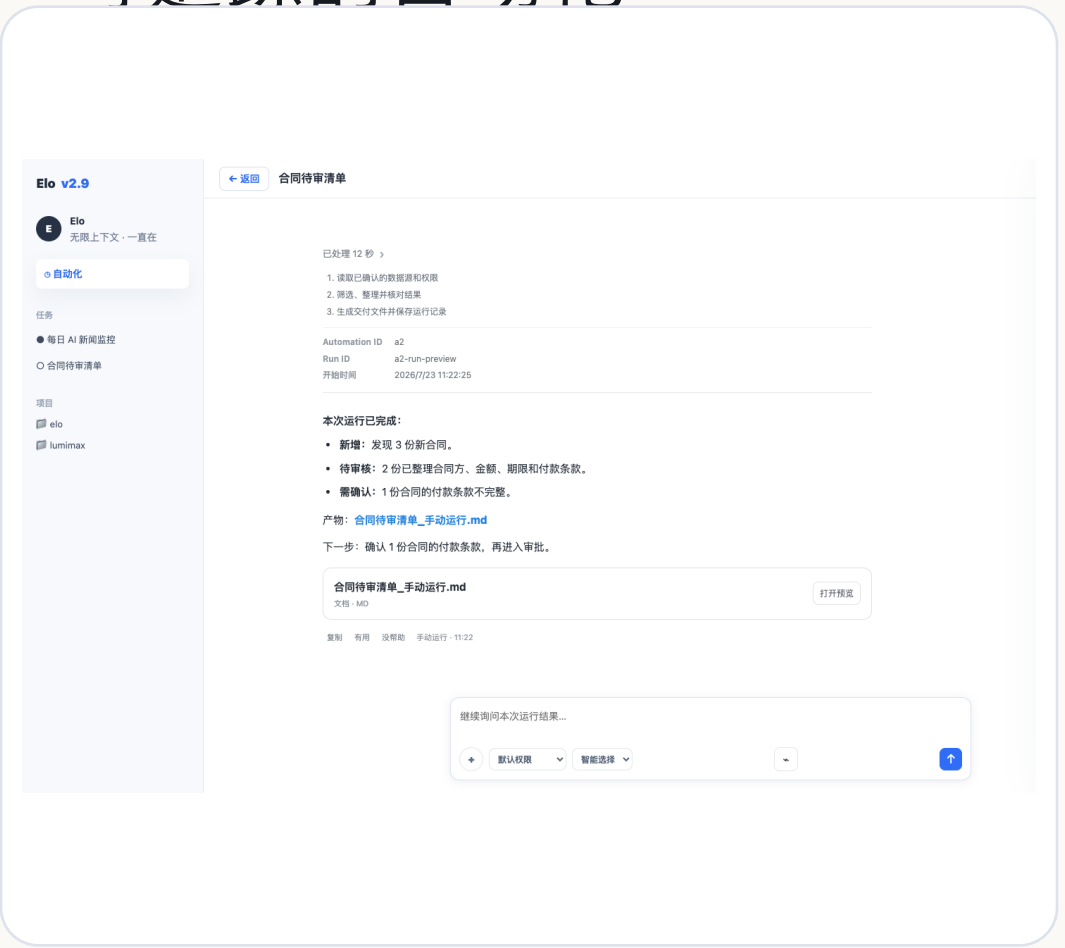
需求抽象　交互原型　PRD 与验收　Agent 机制　评测体系　成本权衡

状态边界：作品包含产品方案、交互 Demo 与本地评测数据；未把未接通的生产 Runtime 描述为已上线结果。

把一句自然语言需求，转成可确认、可追踪的自动化

用对话澄清目标、触发条件、输入输出和权限，再形成可运行定义与结果反馈。

- 先确认关键约束，再生成 AutomationSpec
- 运行状态、结果和失败原因对用户可见
- 敏感操作与外部副作用需要明确确认

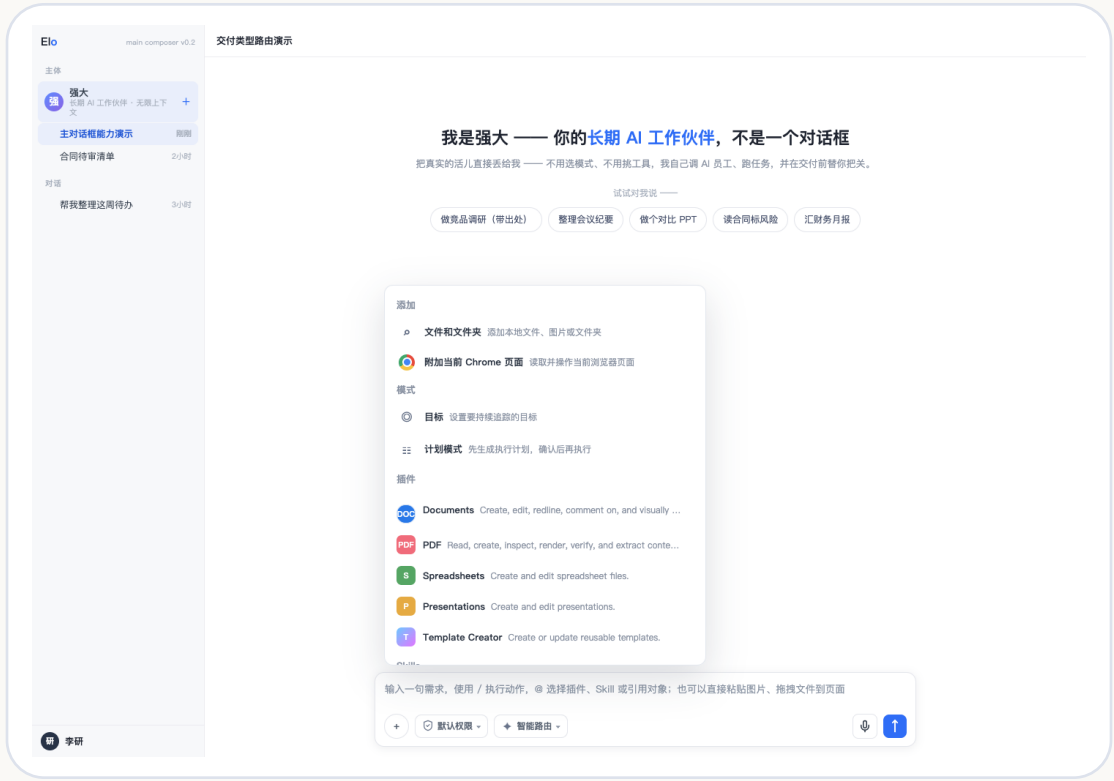


让对话成为 Agent 的任务控制台

统一任务入口与能力选择：输入上下文、选择模式和工具，然后进入执行。

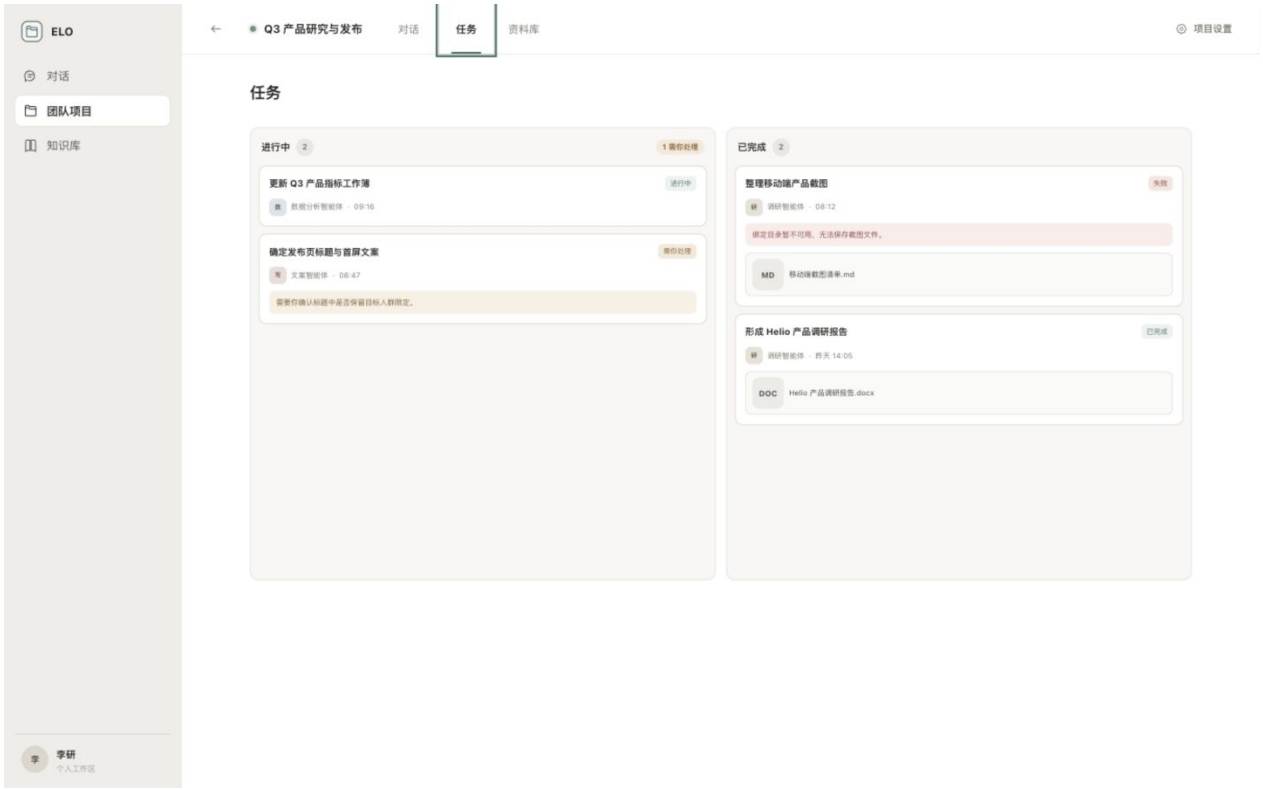


统一任务入口



文件 · 页面 · 目标 · 计划 · 插件

让每项任务只有一个 Owner，让协作边界清晰



关键机制

- Task Owner / Subagent 明确任务归属
- @指定响应，减少多 Agent 抢答与重复调用
- 共享状态，但项目上下文默认隔离
- 关键节点由用户确认，避免不可逆副作用

AGENT CREATION & MEMORY

轻量创建，分层记忆，默认隔离

首屏只要求名称和一句话目标；Profile 自动生成，Skill 推荐后确认，知识库按需绑定。

01

用户记忆

语言、表达方式、长期偏好

允许跨项目复用

02

Agent 记忆

身份、职责、长期协作经验

随 Agent 保留

03

项目记忆

文件、决策、任务与对话

默认不跨项目

本地 Agent 接入：默认只建立会话索引；用户确认后才将摘要去重并写入指定记忆作用域。

Grok Bot 多 Agent 产品实测

独立体验创建、配置、群聊协作与网页操作；目标不是罗列功能，
而是还原 Agent 如何接任务、找工具和交接工作。

我的工作

设计任务、逐项实测、记录问题并归纳规则

测试范围

创建、工具、群聊分工、任务交接与网页操作

核心发现

角色、能力与任务归属要同时清楚

证据边界

个人实测与调研记录；非研发参与

我得到的三个产品判断

01 先选角色，再补能力

先说明要做什么，
再选择需要访问的工具。

02 多人对话不等于协作

明确 @ 对象、负责人和顺序，
才能避免抢答。

03 工具可靠，网页操作补位

有接口时优先调用；
关键步骤交还用户。



群聊协作实测：用明确指派和任务描述完成交接

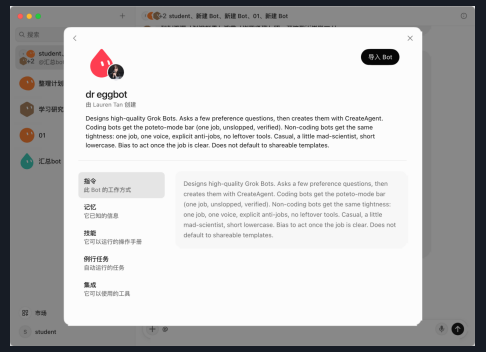
PRODUCT JUDGMENT

从竞品观察，到 Elo 产品判断

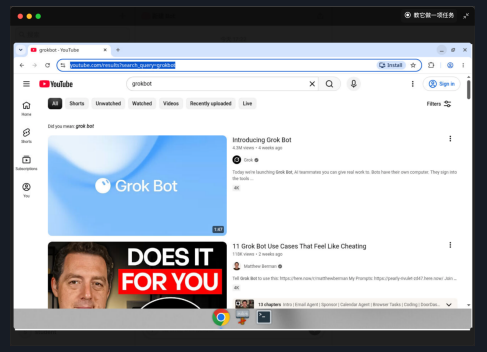
把观察落到入口、责任、执行与权限，而不是照搬界面。



创建：先明确工作类型



Bot：角色与能力一起配置



执行：过程对用户可见

转化为产品规则

入口

先说目标，不要求先配完所有能力

责任

一项任务只设一个负责人

执行

状态、产物和失败原因始终可见

权限

登录、付费、外发等步骤由用户确认

结论

角色、工具和自动执行不是三个孤立开关；任务归属清楚，协作才可控。

证据范围：产品实测与调研结论，不代表官方产品说明。

让复杂知识可探索、可追溯



产品判断

- 节点代表人物、事件、地点、组织与主题
- 从全局图进入局部上下文
- 证据面板解释关系来源
- 基于节点继续提问，而不是只看静态图

完成产品逻辑与交互 Demo；真实文档解析和 GraphRAG 链路待接入。

按任务难度，自动选择合适模型

模型不是前台固定选项，而是由任务类型、质量要求、时延和成本共同决定。



路由原则

- 轻量解析优先走低成本模型
- 多模态和复杂推理按需升级
- 失败降级、预算上限与结果校验
- 用评测集决定路由，不依赖模型排行榜

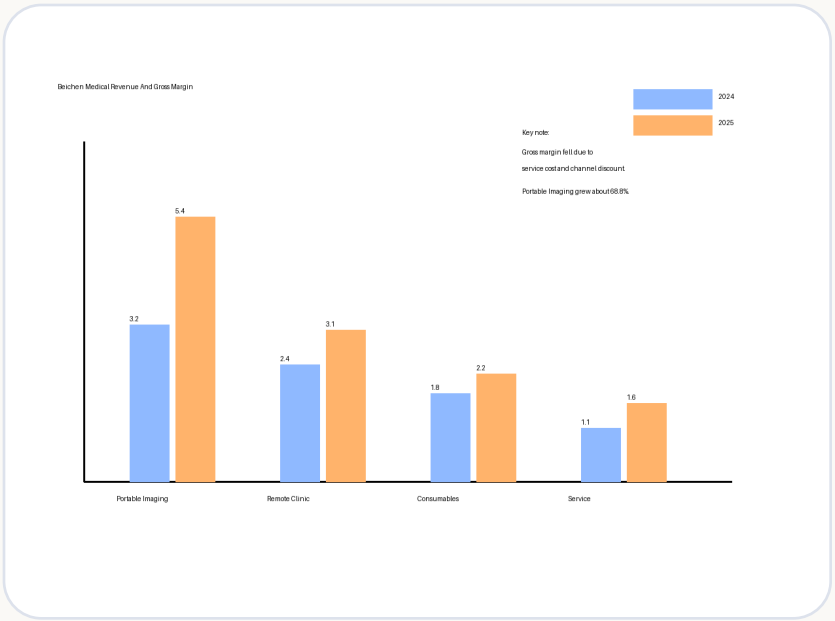
已形成分层方案；真实质量、时延和成本收益仍需基准验证。

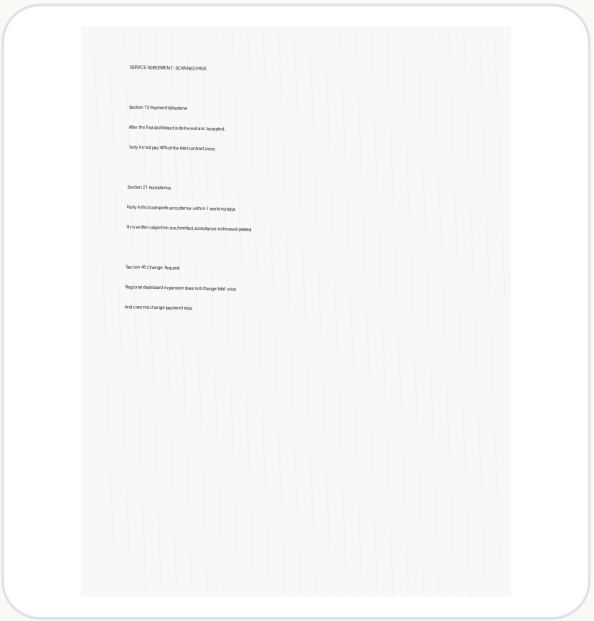
用 103 组题目，把 AI 能力变成可测指标



EVALUATION EVIDENCE

从考试材料到评分规则，形成可复现评测





评测结构

- Corpus：复杂中文资料与真实文件格式
- Questions：事实、跨文档、多跳、冲突和不可答题
- Gold：标准答案、标准证据和必要推理步骤
- Rubric：正确性、引用、证据链、反幻觉与稳定性

样本使用合成内容，避免业务数据和版权风险；文件本体覆盖 PNG、PDF、PPTX、WAV。

价值：支持模型选型、版本回归、Bad Case 分类和产品能力边界判断

PRODUCT METHOD

从场景出发，用证据收敛产品决策

01

找真实任务

明确用户要完成的工作、约束和失败代价。

02

定义闭环

把理解、执行、确认、交付和恢复放进同一流程。

03

做可交互验证

用关键页面验证路径，而不是用静态概念图代替产品。

04

建立评测

用数据集、标准答案和评分规则推动模型与产品迭代。

所有案例均区分：产品方案 / 可交互 Demo / 本地评测 / 工程接入 / 生产结果。

CONTACT

如果你在寻找能把 Agent 方向收敛成产品闭环的人，我们可以继续聊。

希望加入真正把 AI 能力用于复杂工作、并愿意用评测和用户结果验证产品价值的团队。

手机 / 微信	邮箱	GitHub
15618179903	xuefx123@163.com	github.com/ChloeXue00