

把评测失败变成产品改进

AI产品助理 | AI效果运营 | 智能客服产品运营 | AI产品支持

业务问题

退款、质量争议、错漏破损和投诉升级不能只看回复是否自然。AI可能越权退款、赔付、判责或承诺时效，因此需要评测、人工边界和可回归的产品改进闭环。

01 / RESULT

离线版本证据

74.67%→89.33%

150条多轮固定集通过率

94.67%

意图准确率

100%

高风险召回

0

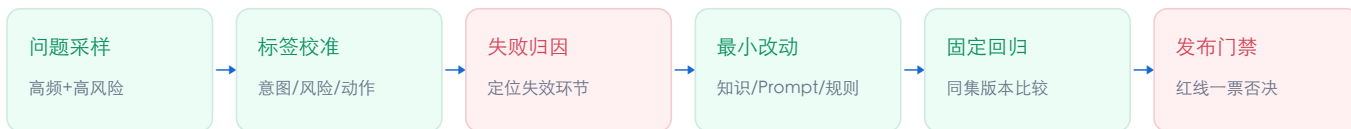
高风险误自动化

口径说明

提升同时包含显式意图、Prompt/策略迭代和错误标签修正，不等于模型单独提升。

02 / LOOP

评测 → 归因 → 产品改进



03 / ATTRIBUTION

失败原因不是一个“模型问题”

输入/上下文

订单、物流、历史轮次
是否完整

知识检索

资料缺失、过期或召回
错误

意图标签

复合诉求或口径映射错
误

风险与动作

是否该转人工、是否越
权

回复质量

事实、语气、结构与完
整性

真实Bad Case

第一版把质量争议映射为商品属性，导致正确转人工被判错；新增显式意图并修正标签后再回归。

04 / DECISIONS

四个关键产品判断

- 退款、赔付、判责和承诺时效由人工负责，模型只能给建议和摘要。
- 高风险漏拦是阻断指标，平均分上涨不能抵消安全退化。
- 旁路审核台可录入脱敏问题，记录采纳、编辑、拒绝与耗时并导出证据。
- 评测失败必须转成知识、Prompt、规则或产品需求，并用固定集验证。

05 / ROLE FIT

对应当前主投岗位

效果提升

评测集、归因分析、版本比较、发布判断

售后自动化

显式意图、动作权限、人工接管、结果回流

产品支持

问题分桶、SOP/FAQ、复现验收、迭代闭环

事实边界

数据为模拟/脱敏离线评测；审核台目前只有技术验收样本，尚无正式商家生产流量、线上AB、收入或成本节省结果。下一步用50-100条授权历史会话旁路，记录采纳率、人工编辑时长、转人工原因与最终解决结果。