M3 · 职业版 Agent · 第 8 讲

第 8 讲：项目——可运行工作流

客服退款助手的窄场景

阅读约 25 分钟

练习约 30 分钟

产出：工作流 + 演示脚本

本节目录

项目背景

一、需求与边界

二、五层架构

三、三个工具

四、评测集（12 条）

五、演示脚本

六、复盘四问

七、交付清单

八、结课标准

九、延伸阅读

项目背景

客户是一家小型电商，售后每天 80 条咨询，一半是“退款到哪了”。客服只有两个人。

目标不是“做一个全能客服”，而是做一个**能查、能答、退款必须确认**的窄流程。两周上线，先跑一个渠道。

一、需求与边界

项	定义
能做	查订单、查退款进度、解释政策
不能做	直接退款、改地址、承诺时效
缺信息	转人工，不猜
渠道	先接企业微信单渠道

边界写在系统提示第一段，不放在最后。

二、五层架构

1. **入口**：企业微信消息；
2. **编排**：识别意图 → 查订单 → 组织回答 → 判断是否转人工；
3. **工具**：订单查询、退款查询、政策知识库；
4. **记忆**：当前会话 + 用户历史工单（只读）；
5. **审核**：退款类请求一律人工确认。

三、三个工具

工具	输入	输出	备注
get_order	订单号/手机号后四位	状态、金额、时间	只读
get_refund	订单号	进度、预计到账	只读

工具	输入	输出	备注
search_policy	关键词	政策片段 + 出处	必须引用

工具全部只读，第一版不接写接口。这是最有效的安全设计。

四、评测集（12 条）

- 正常 5：查订单、查退款、问政策、混合问题、追问；
- 边界 3：订单不存在、手机号不匹配、多订单；
- 失败 2：接口超时、知识库无结果；
- 对抗 2：诱导退款、要求查他人订单。

通过线：正常 ⩾ 4/5，全部无越权。

五、演示脚本

1. 用户问“我的退款到哪了”；
2. Agent 要求订单号后四位；
3. 查到“已退款，3 个工作日到账”；
4. 用户说“太慢，直接退我”；
5. Agent 说明政策并转人工，不执行退款。

演示只演这五步，不要贪多。

六、复盘四问

1. 哪一步最容易被诱导?
2. 哪个工具最常失败?
3. 人工介入率是多少?
4. 单条咨询成本是多少?

七、交付清单

- [ ] 架构图与系统提示；
- [ ] 三个工具与错误处理；
- [ ] 12 条评测与结果；
- [ ] 权限矩阵与确认流程；
- [ ] 日志字段与看板；
- [ ] 一页上线说明（范围、限制、转人工规则）。

八、结课标准

能在一台干净环境里，按文档跑通演示脚本，且评测通过线达标，才算完成。

九、延伸阅读

- 前七讲全部；
- 知识库 `1.3 AI Agents（智能体）`、`4.4 商业化：行业解决方案`。

> 结课：把工作流、评测集、权限清单放进作品集，下一单就有参照。