

个人简历



姓 名：段孝杨

年 龄：24

电 话：19826578520

状 态：随时到岗

邮 箱：xyangduan@163.com

求 职：Product test intern

教育经历

2024.9-2027.6	上海海事大学	计算机科学与技术（研究生）
2020.9-2024.6	阜阳师范大学	电子信息科学与技术（本科）

实习经历

上海启链云智能科技有限公司

2026.01-2026.04

基于 LangGraph 餐饮运营场景 Agent

项目描述：面向餐饮运营场景的 AI Agent，目标是把餐厅代运营团队中大量依赖人工经验、分散文档以及重复沟通的工作流程，沉淀成 Agent 可调用的工具以及自动化的链路用来辅助餐厅老板和食堂管理人员做经营决策。

技术栈：Python / FastAPI / MySQL / PaddleOCR / MiniMax API / Agent Tools Calling / RAG

LangGraph 编排：设计基于 LangGraph 的多节点 Agent workflows，实现意图识别、工具规划、工具执行与结果汇总。

非结构化数据自动入库：构建非结构化数据自动入库流程，基于 PaddleOCR 与大模型解析菜单 SOP 图片，并将流程封装为标准 Skills，单图处理耗时由 60s 降至约 10s。

RAG 系统：搭建 RAG 检索链路，使用 BGE-M3 embedding 与 Milvus 向量库管理餐标、FAQ 等知识内容。

双层记忆机制：构建短期上下文记忆与长期用户记忆，支持用户偏好、业务规则和系统边界的持久化管理，提升多轮对话中的意图理解和响应一致性。

长期记忆优化：优化长期记忆更新机制，结合显式记忆信号与周期性会话扫描，降低单轮请求耗时约 5-10s。

项目经历

2026.02-2026.05

医疗助手 Agent 系统

项目描述：传统单 Agent 医疗问答系统存在响应效率低、复杂问题分析不全面、多轮对话上下文理解能力弱等问题。面对多症状、多维度的医疗咨询时，回答准确性与稳定性不足；同时缺乏长效记忆机制，无法连贯理解用户意图，难以满足实际医疗场景下高效、可靠、安全的智能问答需求。

技术栈：Python / ReAct / Redis / MiniMax API / Agent Swarm / Harness Engineering / Mem0 / RAG

Skills-Agent 两层架构：底层构建 7 个原子 Skills（知识检索、风险评估、症状分析、生活方式建议、ICD-10 编码、临床指南、深度研究）供 Agent 使用，上层 3 个专业 Agent（健康咨询、症状诊断、医学研究）根据任务需求自主选择调用，Skills 和 Agent 完全解耦。

Agent Loop 执行机制：基于 ReAct 框架实现 Think-Act-Observe 循环，LLM 自主决策工具调用并观察结果，设置硬性限制最多调用 2 次工具防止过度调用，单 Agent 平均响应时间控制在 15 秒以内。

Agent Swarm 协作系统：智能路由判断问题复杂度（70%单 Agent 快速通道/30%Swarm 协作模式），复杂问题自动分解任务并分配给多个 Agent 并行执行，Agent 间通过共享上下文间接共享任务状态和执行结果。

Harness Engineering 约束系统：通过 YAML 文件显式定义 Agent 能力边界并运行时自动验证，实现输出自动修复机制（缺少免责声明自动添加、检测到高危症状自动添加就医警告）。

专业技能

- **Python / FastAPI：**熟悉 Python、NumPy、Pandas。具备 RESTful API 与 FastAPI 后端开发经验。
- **MySQL / Redis：**熟悉表结构设计、索引优化及基础性能调优，了解 Redis 常用数据结构。
- **Agent 开发：**熟悉 ReAct、Plan-and-Execute、MCP、Skills、RAG，具备 LangGraph 实战经验。
- **开发工具：**熟悉 Git、Docker 与 Linux 等常用开发工具，具备项目协作与接口测试经验。
- **Codex / Claude Code：**熟练使用 Codex、Claude Code 辅助开发，提升编码、调试与重构效率。