

于鹏逍

籍贯：河北邯郸 年龄：24岁

教育背景

纽约大学 

计算机工程硕士

美国 · 纽约

2025.09 – 2027.06 (预计)

吉林大学  吉林大学二等奖学金

计算机科学与技术学士

中国 · 长春

2021.09 – 2025.06

项目经历

多模态客服智能体设计 (具有多模态能力的客服智能体设计比赛)

2026.04 – 至今

技术栈：Python, LangGraph, Qdrant, BM25, RRF, Pydantic, VLM, 阿里云百炼

- **知识库构建**：基于产品说明书文本与插图构建多模态知识库；使用大模型清洗 OCR 乱码、错位与粘连文本，通过 VLM 生成图片结构化描述，并依据章节、步骤和图号完成结构化分块。
- **Multi-Agent 架构**：设计主 Agent 路由架构，将产品知识问答分流至 RAG 检索链路，将订单售后问题分流至客服链路，支持多轮对话、问题拆解、问题澄清及图片理解。
- **产品知识问答**：封装检索接口，实现向量召回与关键词召回、RRF 融合及模型重排；构建“Query 分析—工具检索—证据分析” Agent Loop，Chunk 证据不足时自动补采相关内容，提高检索质量。
- **电商客服 Agent**：基于公开电商客服数据集设计 Query2Query 与 Query2Answer 双检索策略，优先 Q2Q 检索，未命中时回退至 Q2A 检索，提高客服问答覆盖率。
- **项目效果**：LLM-as-Judge 评估由单一 RAG 的 0.63 提升至 0.69；在 DataFountain 算法竞赛中获得 34/456，排名前 7.5%。

智能旅行规划问答助手

2026.03 – 至今

技术栈：Python, FastAPI, LangGraph, MCP, Qdrant, Redis, MySQL, RabbitMQ, Docker

- **工具获取**：构建开放域旅行规划 Tool-Use 智能体，接入高德地图、12306 与网络搜索服务，支持路线规划、票价查询。
- **Task 机制**：基于 LangGraph 与 ReAct 实现 Agent Loop，完成模型推理、工具选择、参数生成、结果观察和终止判断；通过 Task 机制调度子 Agent，拆解并执行复杂多工具任务。
- **记忆管理**：实现短期对话缓存与长期用户偏好记忆，支持上下文动态压缩、历史向量检索；使用 RabbitMQ 完成长期记忆异步归档。
- **稳定性设计**：实现 MCP Client 自动重连、失败重试、权限隔离、重复工具过滤、参数缺失检测及异常恢复，提高 Agent 长周期稳定运行能力。
- **后端服务**：使用 FastAPI 提供 AI 服务，支持多用户并发调用、SSE 流式输出及会话管理；使用 Redis 保存短期状态、MySQL 持久化会话数据，并使用 Docker 完成部署。
- **评测体系**：基于 LangFuse 记录模型调用与工具轨迹，构建不同难度评测集，采用 LLM-as-Judge 对任务完成情况进行自动评估，任务完成准确率达到 85%。

专业技能

编程语言：Python, Java, SQL

后端开发：FastAPI, REST API, Pydantic, Redis, RabbitMQ, MySQL, Docker

大模型应用：LangGraph, LangChain, ReAct, Multi-Agent, MCP, RAG, Prompt Engineering, Tool Calling

检索与知识库：Qdrant, Embedding, BM25, RRF, Reranker, Query Rewrite, GraphRAG, Neo4j

工程与评测：LangFuse, LLM-as-Judge, 异步编程, Git, Linux, SSE