

教育背景

帝国理工学院	应用计算科学与工程 硕士 Distinction（最高等级）毕业	2025.09—2026.12
宁波诺丁汉大学	计算机科学 学士 一等荣誉，优秀学生赴英交换两年	2021.09—2025.06

实习经历

微软（英国） CoreAI RAG / Agent 系统开发工程师（实习）	2026.06—2026.09
---	-----------------

项目背景：面向银行反洗钱合规审核，在本地部署约束下，独立从 0 到 1 搭建可审计 RAG Copilot，支持客户案例与监管政策的检索、问答及引用核验。

技术栈：Foundry Local、Phi-4、Dense Retrieval、BM25、RRF

- 混合检索优化：分别索引客户案例与监管政策，融合向量检索、BM25、实体保留及 RRF 排序，在 100 个可回答样本上将 Hit@5 从 90% 提升至 95%。
- 引用校验：基于稳定 Chunk ID、引用白名单及容错修复治理引用，结合句级支持度评测核验答案依据，引用召回率达 97.3%。
- 本地推理链路：设计可插拔 LLM 后端并接入 Foundry Local，统一 CLI、UI 和评测入口，支持模型按需加载、卸载及离线降级。
- 评测归因：建立检索、答案、引用、拒答、延迟、成本六维评测，完成 11 组单变量消融及 5 模型对比；在 120 题冻结基准上，任务成功率达 98.3%，未出现幻觉引用。

快手 电商广告团队 AI Agent开发（实习） 销帮帮2.0 (Sales Copilot)	2025.06—2025.09
---	-----------------

项目背景：面向电商广告销售与运营，提供自然语言查数和知识问答；主要负责爆品 Agent 查询链路和工作流，参与多 Agent 调度及 RAG 优化。

技术栈：Python、LlamaIndex Workflow、Google A2A、gRPC、Kafka、Milvus、Elasticsearch、BGE-M3/Reranker

- 多智能体协同架构：基于 Google A2A 开发主 Agent 调度，结合意图置信度与多轮上下文消歧；通过 Handoff 回传上下文和失败原因，排除失败 Agent 后重新路由，支持知识库兜底。
- 爆品数据查询Agent：主导爆品查询链路，以 LlamaIndex Workflow 编排参数抽取、归一化及校验；经 gRPC 完成短视频、直播、女蜗三类场景接入，支持权限过滤、脱敏及 Excel 导出，以 OdinAI 事件提供补参与流式进度。
- RAG知识问答引擎：通过 Kafka 消费文档变更并更新索引，检索前执行文档权限过滤；融合向量、全文、图谱召回及 BGE Reranker 精排，支持平台规则与操作手册问答。

美团 美团平台 服务端开发工程师（实习） 组队PK得红包	2023.06—2023.09
----------------------------------	-----------------

项目背景：面向红包营销引流场景，负责钱包模块开发与技术方案设计，覆盖余额记账、红包发放及异常补偿。

技术栈：Java、Spring Boot、Squirrel (Redis)、Mafka (Kafka)、Zebra (ShardingSphere)、Crane (XXL-JOB)、MTthrift (gRPC)、MySQL

- 分片并发记账：基于 Zebra 按 userId 对余额表、流水表分库分表，分散写入压力；以行锁控制同一钱包的并发更新，本地事务保证余额与流水一致；资金校验以 MySQL 为准，Squirrel 用于余额展示。
- 分片幂等控制：按活动、用户及业务事件构造稳定幂等键，结合 userId 分片路由与数据库唯一索引，使重复请求命中同一物理表；跨实例并发请求和消息重投复用原流水，避免重复创建发放任务。
- 发放状态机：区分待发放、结果待确认、已发放待落账等状态，将 RPC 超时视为结果未知；对已发放但本地更新失败的流水仅补记账，分离远程发放与本地记账的失败处理。
- 异步补偿机制：基于 Mafka、本地消息表与 Crane 定时扫描，持久化补偿任务并异步重试；消费时读取最新流水状态，跳过已完成步骤，多次重试失败后告警并转人工核查。

实习经历（续）

宁波诺丁汉大学 | IT Service 部门 | 实习生 校园主数据查询系统重构

2022.11—2023.05

项目背景：面向校内师生及实验室，负责主数据查询服务的分层重构、接口治理、缓存优化与新旧接口比对。

技术栈：Java、Spring Boot、MyBatis、Dubbo、Redis、Guava、线程池、Loggly

- **接口治理：**按 DDD 分层组织业务逻辑，通过策略模式与 routeKey 统一查询路由，将对外接口从 **30 个收敛至 4 个**，减少重复实现。
- **接口结果比对：**通过自定义注解、Spring AOP 与 Dubbo 泛化调用比对新旧接口结果，使用 Loggly 定位字段缺失及返回差异。
- **缓存模板：**封装本地缓存与 Redis 访问，统一刷新入口；通过随机过期时间分散集中失效，支持同步刷新及 MQ 异步刷新。
- **多租户扩展：**通过拦截器与 ThreadLocal 维护租户上下文，为多租户查询及跨校服务扩展提供支持。

项目经历

求职邮箱自动化 Agent | 开源项目（独立开发） github.com/ada-zl5825/Agenda-Agent

2026.07—2026.08

项目背景：独立开发招聘邮件 Agent，提取岗位与日程，更新申请看板并生成日历、待办及简报。

技术栈：LangGraph、FastAPI、Azure Functions、PostgreSQL、Microsoft Graph、Pydantic v2、AES-256-GCM

- **可中断恢复的工作流：**基于 LangGraph 编排 **15 节点**流程，以 PostgreSQL 检查点支持中断及跨进程恢复；检查点与业务数据分离，仅保存执行位置和安全状态。
- **隐私安全设计：**原始邮件、OAuth 凭据及明文动作 URL 不进入模型或检查点；链接采用 **AES-256-GCM 加密**，绑定邮件关联数据并校验目标主机名。
- **幂等申请状态机：**以语义事件指纹和稳定动作键避免重复操作；对时区缺失等 **5 类场景**触发人工评审，经校验与乐观并发检查后恢复，支持节点安全重放。
- **评测验证：**建立抽取、流水线重放与云端遥测三层评测；在 **60 例合成数据集**的一次 live 运行中，相关性 **F1 97.0%**，事件时间匹配 **98.3%**。

跨境购物 Agent | 个人项目 github.com/ada-zl5825/global-shopping-Agent

2026.06—2026.08

项目背景：面向跨平台比价与个性化购物，基于 AgentScope 构建 Supervisor-Workers 多 Agent 系统，按读写属性拆分子 Agent 并行检索，跨平台检索延迟较串行降低 **65%**。

技术栈：AgentScope、OpenSearch、LangFuse、Qwen3-35B-A3B

- **Cache Breakpoint 上下文治理：**以“早期摘要 + 近期原文”压缩上下文，10+ 轮对话控制在 **30k tokens** 内，Prompt Cache 命中率达 **80%**；支持长期记忆检索、删除及偏好黑名单。
- **OpenSearch 混合召回：**构建 Query / Item 双塔编码，融合语义、全文及标签召回；在 **200 条人工标注 Query** 上，Top-100 商品召回相关率较单塔基线提升 **22%**，跨语言场景提升 **35%**。
- **AgentScope Harness：**实现 LoopDetector、90s 异步超时与工具熔断，支持主循环恢复；接入 LangFuse Trace 归因耗时与 Token，任务完成率 **71%→89%**，bad case 排查 **30min→5min**。
- **Agentic RL 训练闭环：**基于 LLM-as-Judge 筛选轨迹，完成 SFT 并接入 GSPO；在 **100 条难度分层冻结基准**上，Qwen3-35B-A3B 归一化 RaR 得分为 **62（基座）→79（SFT）→88（GSPO）**，工具调用成功率 **90%+**。

专业技能

编程开发：熟练使用 Python，具备 Java 后端实践；熟悉 Linux、Git、Pytest、GitHub Actions。

Agent 工程：具备 AgentScope、LangGraph、LlamaIndex、Google A2A 实践，涵盖多 Agent 协同、上下文/记忆管理、断点恢复与运行时治理。

RAG 检索：具备混合检索、RRF、重排及引用治理实践；使用过 Milvus、Elasticsearch、Neo4j、OpenSearch、Foundry Local。

后端工程：熟练使用 FastAPI、MySQL、Redis；熟悉 Spring Boot、MyBatis、Dubbo/gRPC、PostgreSQL、Guava、Kafka；掌握事务、幂等、缓存一致性及 MQ 补偿机制。

训练评测：具备 SFT/GSPO、LLM-as-Judge 与消融评测实践；使用 LangFuse 进行链路归因。

英语能力：雅思 **6.5 分**，可独立阅读英文技术文档并开展技术沟通。