

PERSONAL RESUME

姓名：邓康

性别：男

年龄：24

电话：15201960205

邮箱：onesp123@163.com

意向岗位：AIGC 应用开发工程师

意向城市：上海

求职类型：社招



教育经历

2019.09-2023.06

江西农业大学

食品科学与工程 | 本科



项目经历

项目名称：医疗助手

项目描述：本项目基于开源 Qwen2.5-7B 大语言模型，在 A100 计算平台上利用 LlamaFactory 框架，对高质量的开源医疗数据集 Huatuo26M-Lite 进行针对性微调。项目核心目标是提升模型在医学领域的专业问答能力，特别是在中医诊断方面生成更准确、更符合医疗语境（包括专业术语运用、人性化表达及语气）的响应。项目工作涵盖从数据集处理、模型部署、高效微调到效果评估的全流程。

技术栈：

核心框架与工具：LlamaFactory (用于高效微调)

基础模型：Qwen2.5-7B (开源大语言模型，来自 Hugging Face)

数据处理：Python (数据清洗、格式化)

平台与资源：Hugging Face (模型与数据集来源), A100 GPU 服务器 (计算平台)

服务化部署：vllm, FastAPI

项目成果：

专业化能力显著提升：成功微调后的模型在回答医疗相关问题时，其专业性（如术语准确性）和人性化表达（语气、口吻）相比原始模型得到显著优化，有效减少了机械式回答。

中医诊断效果优化：针对 Huatuo26M-Lite 数据集在中医领域的特性进行微调，模型在该领域的诊断建议和解释能力得到针对性加强。

模型响应效率提升：通过 LlamaFactory 的优化技术和参数调整（如学习率、批次大小等），在保证效果的同时，提升了模型的训练效率和最终的推理响应速度。

项目名称：AI_CRM 系统

项目描述：本项目基于多模态 AI 技术构建智能 CRM 辅助系统，通过意图识别引擎动态选择任务处理路径（网络搜索、本地 RAG 检索、数据库交互），实现客户数据管理与业务场景的深度协同。系统核心流程包括：用户输入→意图分类（客户咨询、数据查询、业务操作）→调用对应工具链（搜索引擎、本地知识库、MySQL 数据库）→生成结构化响应。集成 RAG 技术实现 CRM 行业知识库的语义检索，支持从合同文档、客户沟通记录等非结构化数据中提取信息，结合 MySQL 存储的订单、客户画像等结构化数据，提供端到端智能决策支持。技术方案参考了企业级 CRM 的 AI 架构设计，并融合检索增强生成（RAG）技术优化知识响应精度

技术栈：

1. 编程语言：Python

2. 意图识别：HuggingFace Transformers (BERT 微调分类模型) + 规则引擎

3. 检索系统：RAG 框架：LangChain + Sentence-Transformers (构建本地知识向量库) 向量数据库：Chroma (存储 CRM 行业文档、产品手册等嵌入数据)

- 4.数据交互: MySQL 连接: PyMySQL (客户信息、订单记录查询)
- 5.大模型集成: DeepSeek-7B (文本摘要)、GPT-4o (对话生成)
- 6.业务工具链: CRM OpenAPI (对接企业客户管理系统)

● 项目成果:

- 1.动态意图处理体系: 实现 93%准确率的意图分类, 支持 6 类业务场景 (如客户信息查询、订单状态追踪、产品知识问答) 开发混合决策引擎, 根据输入内容自动选择最优处理路径 (例: "2024 年 Q3 华东区销售额"触发 MySQL 统计→RAG 补充行业对比数据→生成可视化报告)
- 2.多源数据协同检索: 构建包含 20 万+CRM 行业术语的 RAG 知识库, 检索响应速度<1.2 秒 (Chroma 索引优化) 实现跨数据源关联查询 (如通过客户 ID 同步显示 MySQL 中的交易记录+RAG 检索的沟通历史)
- 3.自动化业务流: 开发 12 个预置工作流 (如客户投诉自动创建工单→检索相似案例解决方案→推送服务建议) 通过 GPT-4o 生成符合企业话术规范的邮件/报告模板, 人工修改量减少 70%
- 4.企业级系统集成: 提供 RESTful API 接口, 已对接纷享销客、Salesforce 等主流 CRM 系统实现敏感数据脱敏处理, 通过 RBAC 权限控制保障企业数据安全



工作经历

2023.06-2024.08

江西神珠田园食品有限公司

质检员 QC

2024.08-2025.03

嘉善日善电脑有限公司

技术员



职业技能

深入理解 LLM 底层原理, 包括 Transformer 架构、自注意力机制与位置编码优化, 曾基于 LangChain 实现 RAG 系统, 使垂直领域问答准确率从 68%提升至 89%

熟练使用 LangChain 构建企业级应用, 开发过支持动态工具调用的客服机器人, 集成 SQL Agent 与文档解析模块, 解决 85%的常规咨询问题。

掌握 RAG 全流程优化技术, 包括文档分块策略 (RecursiveTokenSplitter)、向量数据库 (Milvus) 与重排序算法 (ColBERT), 召回准确率提升 52%

运用 LlamaFactory 实现大语言模型的高效微调, 深入理解 LoRA 高效微调算法, 在医疗垂直领域任务中使中医问答准确率提升 35%



自我评价

热衷前沿技术, 学习能力强, 关注潮流趋势, 适应行业快速发展。良好的沟通与团队协作能力, 执行力强, 积极参与团队讨论与决策