



年龄：27岁

性别：男

工作经验：5年经验

Tel: 13840551725

Email: 1941597171@qq.com

籍贯：安徽蚌埠

身高：168cm

爱好：动漫，爬山

个人评价：

- 1.关注魔搭社区，及时更新大模型知识，不断丰富自我，具有高效的学习能力；
- 2.性格随和，合作能力强，沟通能力良好；
- 3.乐观的生活态度，积极抗压，能够适应高强度的工作。
- 4.积极思考问题；有实用的AI应用开发经验和丰富的解决问题经验。

## 教育背景

2016-09 ~ 2020-06

东北大学 (985)

机械工程与自动化 (本科)

专业成绩： 概率论满分，理论力学年级第一（600人），学信网可查

荣誉证书：蚌埠市数学竞赛一等奖，安徽省物理竞赛三等奖，大学优秀奖学金三等奖

## 个人优势:

- 1.编程语言：精通 Python，熟练使用 fastapi 开发后端接口，掌握Linux、shell脚本编写，熟悉 Numpy、Pandas、Matplotlib等第三方库。
2. 熟练使用 LangChain、Rag、搭建 RAG 应用，掌握提示词工程、结构化输出、对话记忆、多模态应用、多场景知识库构建。
- 3.熟练使用 Dify等低代码平台进行 demo 验证与技术实现验证。
- 4.熟悉 Milvus 等向量检索引擎，熟悉产品级 RAG 实现方案的完整生命周期，基于 RAG 和实际业务需求完成 Agent 流程编排的经验丰富。
- 5.熟悉人工智能产品应用的主要路线 NL2API 的 NL2SQL 的实现细节，具备应用级技术实现经验，熟悉 RAG 标注增强方案实现细节。
6. 熟悉 MCP、A2A 等 Agent 交互协议，有 Multi-agent 协同系统开发经验。
7. 能够熟练使用 Vanna 库+dify 工作流，实现高质量的自然语言转 sql，再转 echarts 图表。

## 工作经历:

2024-05 ~ 2025-11

嘉兴智瞳科技有限公司

大模型应用开发工程师

工作职责：

1. 需求对接与业务解构：深度联动业务侧，精准拆解核心诉求，将抽象需求转化为可落地、可量化的技术实现目标，确保技术方案与业务价值对齐。
2. 核心技术开发与功能落地：聚焦 RAG 架构、多模态处理等关键技术模块的工程化实现，搭建知识匹配、多源数据解析、内容检索等核心业务逻辑；研发定制化 Agent 智能体，支撑个性化方案生成、自动化校验、智能交互等场景化功能落地。
3. 低代码 AI 应用创新探索：探索 Coze、Dify 等低代码平台的技术特性与应用场景，结合智能体开发技术，构建低门槛、高复用的 AI 应用落地路径，推动 AI 技术普惠化赋能业务场景。

## 项目经历:

### 《智法通系统》项目研发 (2024.07~2025.09)

项目背景：

随着法治社会建设的深入推进与市场经济的复杂化发展，法律风险已成为企业经营与个人生活中的核心风险要素。当前法律风险管控领域普遍存在的痛点：识别难、成本高、响应慢。

**核心职责：**开发了“法律咨询”、“合同审查”、“案件检索”“风险测评”四个智能体用到的技术栈：

langchain、fastapi、MioIO、Jieba、Milvus、Redis、RAG、Docker、mysql、Dify、text2sql、Nginx、ECharts

**工作职责：**

- 1、法律知识底座构建与优化，基于Langchain框架设计并落地全流程法律文档处理体系。
- 2、牵头封装Minio对象存储服务，定制权限校验机制，实现合同、法规、案例等多类型法律文件的加密上传、分类归档与合规下载，保障文档数据安全性与可追溯性。
- 3、针对法律文本专业属性，优化Jieba分词策略，设计智能切片规则，将文本切片经 Embedding模型向量化后存入Milvus向量数据库，检索准确率较基础方案提升30%以上；

- 4、基于RAG技术架构搭建智能法律问答系统，负责Ollama本地化法律大模型的部署与调优，通过langchain实现大模型与Milvus向量数据库的联动对接。
- 5、设计“相似度检索+法律关键词检索+法条关联检索”的多路召回机制，构建混合检索策略，解决单一检索模式下的漏检、误检问题，确保法律咨询响应的完整性。
- 6、主导Vanna.AI结合Dify设计可视化工作流，基于text2sql技术开发智能数据查询模块，支持用户通过自然语言查询合规数据、风险统计等信息，自动转换为SQL语句执行并返回结果。

项目难点：法律文本语义理解的歧义性与专业性难题

**解决方案：**采用“模型微调+知识增强”双策略，构建“法律术语-释义-适用场景”结构化知识库，通过LangChain工具链实现术语与上下文的关联解析，提升语义理解精准度。

2023-02 ~ 2024-05                      南京景曜智能科技有限公司                      应用开发工程师

工作职责：

- 1. 业务转需求：和业务同事紧密配合，把他们说的“要解决什么问题”拆明白，连潜在的需求都挖出来。把“想要个整理数据的智能体”这种模糊说法，变成“3天内实现数据自动匹配，准确率达95%”这种具体、能落地的技术任务，确保做的东西真能帮业务提效。
- 2. 做开发落地功能：专门搞RAG、多模态这些实用技术的开发，把知识匹配、多类数据解析、内容快速查找这些核心活儿做扎实。还要开发专属的AI智能助手，帮业务生成专属方案、自动检查内容对错、实现智能对话，让技术真能用到具体场景里。

项目经历：

《健康智餐系统》项目研发（2023.03~2024.04）

项目背景：

随着人口老龄化加剧、慢性病患率上升以及临床营养诊疗规范化需求的提升，营养风险筛查已成为疾病治疗与健康管理的关键前置环节。当前营养风险筛查与管理领域仍存在诸多痛点：

- 1、传统筛查流程高度依赖人工操作，还容易因人为因素出错；
- 2、营养知识体系更新迭代快，临床营养指南、疾病营养干预方案的频繁修订等。

项目架构：

本系统采用五层分层微服务架构，自上而下依次为：  
客户端层、API网关层、应用服务层、AI能力层、数据层。**各层职责清晰、解耦良好，支持高并发、可扩展与快速迭代。**

本人主要设计了“数据分析”、“智能对话”、“风险筛查”、“个性化食疗”四大智能体；

用到的技术栈：

langchain、fastapi、MioIO、Jieba、Milvus、Redis、Ollama、RAG、Docker、mysql、Dify、text2sql、Nginx、ECharts、Vanna

工作职责：

- 1.基于Langchain框架设计并落地“文档解析 - 内容提取 - 智能切片 - 向量入库”；
- 2.针对专业营养文本属性，优化Jieba分词策略，设计“指南章节 - 指标维度 - 案例要素”三级智能切片规则，将文本切片经Embedding模型向量化后存入Milvus向量数据库，减少大模型幻觉；

- 3.基于 RAG 技术架构搭建智能营养问答系统，负责 Ollama 本地化营养大模型的部署与调优，设计“相似度检索 + 营养关键词检索 + 指南关联检索”的多路召回机制，构建混合检索策略，确保召回响应的完整性和准确性；
- 4.通过Vanna.AI实现text2sql来返回检索结果，再通过Dify平台搭建工作流，通过多结构分支将查询结果转化为不同的ECharts图表；
- 5.引入线程池技术和消息队列，优化高并发下的测评请求处理，结合 Redis 缓存机制缓存高频查询结果，将系统响应速度提升 40%，有效解决峰值时段性能瓶颈；
- 6.设计 Nginx 反向代理架构，配置加权轮询为主、最少连接数为辅的负载均衡策略，搭配 IP 哈希机制保障会话粘性，同时设置请求限流，防止高并发导致的系统崩溃。

项目难点：

难点一：高并发问题

解决方案：

- ① 引入线程池技术和消息队列，优化高并发下的测评请求处理，结合 Redis 缓存机制缓存高频查询结果，将系统响应速度提升 40%，有效解决峰值时段性能瓶颈；
- ② 设计 Nginx 反向代理架构，配置加权轮询为主、最少连接数为辅的负载均衡策略，搭配 IP 哈希机制保障会话粘性，同时设置请求限流。

难点二：多维度个性化配餐的精准性平衡（业务核心痛点）

解决方案：

- ①采用“遗传算法（全局寻优）+线性规划（局部精准）”组合引擎，将营养达标率设为核心权重（占比 60%），偏好与成本设为辅助权重；
- ② 通过Drools引擎嵌入“优先级规则”（如“过敏食材绝对排除 > 营养需求 > 口味偏好”）。

2020-07 ~ 2022-12

鸣志电器（太仓）有限公司

结构研发工程师

主要项目经历：

《特斯拉车载小型无刷电机》设计开发

工作职责：

主要负责设计电机的全部流程，从需求审核，到可行性分析，竞品分析，方案设计，结构设计，结构仿真，绘制图纸，原型机测试，生产机测试，以及售后技术服务等。