

## 庄娘喜

☎ 联系电话: 15016534572    ✉ 邮箱: zhuangnx210087@stu.nfu.edu.cn  
🌐 技术博客: [https://blog.csdn.net/weixin\\_66397563](https://blog.csdn.net/weixin_66397563) (共收获 24 万访问量)



## 实习经历

### SHEIN(全球独角兽企业 Top5) — 数据分析师

2025.01 - 2025.06

- **数据清洗**: 结合 **Qwen-Plus 模型与 Prompt 思维链 (COT) 技术**, 自动化处理 1000+ 个格式复杂且不一致的 Excel 表格; 使用 Python 和正则表达式提取物料商品的名称、材质、工艺分类等信息。
- **数据分析**: 分析供应链采购流程中各环节月度总金额趋势、周期和转换率; 计算并评估商品单位成本的波动情况; 分析订单履行周期与审批流程等时效指标。此外, 按类别分析商品的金额占比, 并监控供应商订单履约率和交货准时率, 为降低成本和提升时效提供数据支持。编写复杂 Trino SQL 语句实现指标自动化计算, 并结合 BI 看板展示数据结果。
- **Agent 问答系统(供应商规划与物料需求最优分配)**: 针对供应链采购人工分配低效、供应商组合优化不足问题。基于 **Agent+Text2SQL** 智能问答系统, 通过意图识别分流处理物料需求分配与价格调控两类问题, 实现采购决策自动化。
  1. **SQL 数据库构建**: 设计 MySQL 表存储供应商基础信息和历史交付数据, 包括供应商名称、物料名称、历史交付绩效指标、月最大供应量、单价、运输单价、质量合格率、交付准时率、生产周期及调价规则等字段。
  2. **意图识别模块**: 基于公司内部的 Claude 模型 API, 利用 Prompt 工程优化技术, 对物料需求分配与价格调控等意图分类。
  3. **物料需求分配意图应答**: 根据用户问题, 使用 Text2SQL 技术, 自动从数据库中筛选出能够提供目标商品的供应商。对于多供应场景, 定义多目标规划函数(成本最小化、质量最大化、交付风险最小化)对总量进行最优分配; 最终将分配结果通过 Prompt 模板输入 LLM, 生成简明专业的采购方案。
  4. **价格调控应答**: 在用户询价或调价场景下, Text2SQL 抽取供应商调价规则, 结合实时市场均价计算价格波动幅度是否超过合同阈值, 动态生成调价方案。最终通过 Prompt 将市场均价波动、调价区间与风险提示整合, 输出清晰可执行的价格调整建议。

### 一路绿灯智能科技有限公司 — 算法工程师

2024.11 - 2025.01

- **项目概述**: AI Agent 代账助理, 使用 **RPA+Agent** 自动化接管微信账号, 识别群聊中的客户是否了解公司税务情况以及开发票需求。
- **项目职责**:
  1. **发票信息知识库构建**: 使用 OCR 技术对用户历史发票数据进行内容识别并存储在 MySQL 数据库。
  2. **微信自动化交互**: 开发基于 Python 的 RPA 自动化程序(使用 PyAutoGUI、ntchat 库), 实现微信群聊消息的实时获取与智能回复。
  3. **意图识别模块**: 基于 Qwen 模型 API, 利用 Prompt 工程优化技术, 实现客户税务咨询、发票开具等多种意图分类。
  4. **RAG 检索**: 对用户提供的发票内容进行提取, 并使用模型进行语义检索, 从数据库中检索相关数据以补全缺失字段。
  5. **Agent 代理模块**: 设计多轮对话策略, 结合知识库获取完整发票信息, 并使用 Selenium 实现开发票意图的代理工具, 自动登录国家电子税务局开发票并发送回微信群聊。
- **项目结果**: 自动开票准确率达到 95%, 减少人工干预时间超 60%。

## 项目经历

### RAG 应用优化项目 — 基于 RAG 的金融制度问答系统

2025.04 - 2025.6

- **项目背景**: 基于大模型的文档问答, 根据用户输入的金融制度类问题, 基于给定的金融文档库, 生成准确、合规的答案。
- **项目职责**:
  1. **知识库构建**: 对中国银监会/人民银行/外汇局发布的金融制度文档(800 份)提取文本内容, 使用 **递归分块技术**对文本内容进行多层级文本分块, 使用 BAAI/bge-large-zh-v1.5 将文本块转换成向量存入 Faiss 向量数据库。
  2. **意图识别**: 使用 Qwen3-8B-nothink 设计用户意图识别功能, 用户问题归类金融类、联网检索、不联网三类, 分别进行分流应答。
  3. **金融类意图应答**: (1) **多路召回**, 使用 Qwen3-8B-nothink 对用户问题进行改写、拆解、扩展成多个子问题 (2) 对每个子问题使用 **混合检索算法**(向量检索 + BM25 关键词算法), 分别召回 TOP5, 去重合并后输入到重排序模块, 对比基线答案准确率提升 4%。(3) **重排序与父文本块索引**, 使用 bge-reranker-v2-m3 模型对多路召回的文本块进行重排序筛选获取 TOP10 的子文本块, 根据父子索引获取其父文本块输入到答案总结模型, 对比基线答案准确率提升 3%。(3) **相关知识引用答案总结优化**, 使用 Qwen3-think 模式设计相关知识引用总结 Prompt, 对重排序后的文本块进行显示相关知识序号引用再进行答案总结, 对比基线答案满意度提升 2%。
  4. **通用问答意图应答**: (1) 联网意图, 通过 Tavily 搜索引擎 API 实现联网检索, 定制联网应答 Prompt 调用 Qwen-plus 回复用户问题, (2) 非联网意图, 定制非联网应答 Prompt, 调用 Qwen-plus 回复用户问题。
  5. **问题推荐模块**: 定制问题推荐 Prompt, 使用 Qwen-plus 模型对用户最后一轮对话进行三个问题推荐, 增加用户点击率。

- **项目指标:** 评测集包含 1500 道题目，其中 1200 道选择题（含单选与多选）、300 道问答题。选项匹配（0/1）；问答题 embedding 相似度（0/1），最终计算平均得分作为评估指标。标准 RAG Pipeline 指标为 0.76，优化后为 0.85

LLM 微调应用项目 — 基于 LLM 和 RAG 的专属知识库校园问答系统

2024.02 - 2024.3

- **项目概述:** 基于 Qwen 模型，使用**LoRA 微调**与 RAG 检索增强策略提升模型在校园招生领域的问答精度与信息实时性。  
**项目职责:**
  1. 模型部署与服务化: 使用 Vllm 在 8\*4090 服务器上本地部署 Qwen1.5-32B 模型；通过 FastAPI 对模型服务进行封装成 API 接口。
  2. 数据集构建与微调: 使用爬虫技术采集学校招生政策、学校信息、学校概况和专业信息四个方面的数据，存储到本地文档，通过 ChatGPT 转化为 QA 数据集 (10k 条)。并使用 LoRA 技术对 Qwen 模型进行轻量化微调；
  3. 本地知识库构建: 将本地文档按固定大小分块，使用 BAAI/bge-small-zh-v1.5 模型进行向量化嵌入并存入 Faiss 向量数据库。
  4. 检索链路搭建: 对用户 Query 进行向量化，通过向量检索对本地知识库 +Bing 联网搜索分别召回 Top10 的文本块，使用 BAAI/bge-reranker-large 进行重排序筛选 TOP10 的文本块，将文本块注入 Prompt 模板，使用模型基于上下文语义整合信息，生成回复。
- **项目评估:** 对于问答库中已有的问题，模型回答正确率 94.6%，对于部分无对应文本的问题，模型通过联网搜索获取相关信息。

科研论文

大数据与人工智能研究中心 — 人工智能算法相关课题研究

2021.10-2024.09

- MultiverseAD: Enhancing Spatial-Temporal Synchronous Attention Networks with Causal Knowledge for Multivariate Time Series Anomaly Detection(**中科院二区 SCI 在审**)
- G-GLformer: Transformer with GRU Embedding and Global-Local Attention for Multivariate Time Series Forecasting(**CCF B**)
- Mamba or Transformer for Time Series Forecasting?Mamba and Transformer are All You Need(**CCF C 在审**)
- MSFNet: Exploiting Fourier Transform for Multi-Scale Electric Load Forecasting(**EI**)

个人技能

- **编程与框架能力:** 熟悉 Python 语言和 PyTorch 框架。
- **工具应用与开发:** 熟悉 Vllm 推理框架和 Docker 容器化部署技术，具备 FastAPI 后端服务架构设计与 API 接口开发经验。
- **数据采集与处理:** 熟练使用 Requests、Scrapy、Selenium 等爬虫技术；熟悉**RAG-图文多模态数据**的处理流程；掌握 MySQL 数据库。
- **大语言模型技术:** 掌握 Llama、ChatGLM、Qwen 等开源模型训练推理及 LoRA 等微调方法。熟悉 LangChain 的 RAG、Agent 开发框架。熟悉 Prompt Engineering，包括设计和优化提示词以提高模型性能，掌握 CoT 思维链和 Few-shot Prompting 少样本提示。

教育背景

广州南方学院 — 数据科学与大数据技术（大数据与人工智能系） 本科

2021.09 - 2025.06

- GPA: 89/100（3.91/5.00，Top 3%）
- 荣誉奖项: **国家奖学金**、优秀学生一等奖学金、2024 年励志榜样
- 相关课程: 人工智能原理、自然语言处理、数据挖掘与机器学习、大数据分析可视化、大数据智能分析实训

学科竞赛

- 🏆 国际数据科学竞赛平台 Kaggle 中获得 银牌
  - 🏆 MathorCup 高校数学建模挑战赛 全国一等奖
  - 🏆 数维杯国际大学生数学建模挑战赛 国际赛一等奖
  - 🏆 中国高校大数据挑战赛研究生组 全国一等奖
  - 🏆 软件著作权:《基于大语言模型的专属知识库问答系统》、《基于大数据和物联网的城市洪涝灾害预警系统》
  - 🏆 大学生创新创业训练计划项目 国家级 1 项、省级 1 项、校级 2 项

国家级: 智御洪涝: 基于大数据和物联网的城市洪涝灾害预警系统

省级: 基于计算机视觉与路径规划的无人仓储管理系统
- 🏆 MathorCup 大数据竞赛研究生组 全国二等奖
  - 🏆 中国高校计算机大赛-大数据挑战赛 全国优胜奖
  - 🏆 “泰迪杯”数据分析技能赛 全国二等奖
  - 🏆 “挑战杯”广东大学生课外学术科技作品竞赛 三等奖