

李星蒴

- **性别：男**
- **年龄：23**
- **电话：16631932625**
- **户籍：河北**
- **邮箱：2382990465@qq.com**
- **年限：1年**

教育经历

2021.9-2025.6 **(211,双一流) 河海大学**

数学与应用数学 | 全日制本科 GPA:3.80/5.00 专业排名：15

主修课程：数学分析、高等代数、实变泛函、复变函数、概率论、数理统计、数学建模、MATLAB 科学计算、Python、C++程序设计、数据结构、数据库、离散数学、机器学习等

通过英语四级，有一定英语阅读能力，熟悉 Python,MATLAB,C++等编程语言

个人技能

1.熟练使用Python语言, 使用**Numpy、Pandas、Matplotlib**做数据清洗和可视化等工作;

2.掌握SQL语言, 熟练使用**MySQL等数据库**进行数据处理工作;

3.熟悉**BERT、LSTM、GRU**等深度学习模型及其在自然语言处理中的任务;

4.掌握**模型量化、模型蒸馏、模型剪枝**等常见的模型压缩技术;

5.了解随机森林、GBDT、FASTTEXT等常见的机器学习算法;

6.熟练使用**PyTorch深度学习框架**, 熟悉**Transformer架构**以及Attention机制;

7.了解DeepSeek、千问等主流大模型，熟悉LLM大模型的API调用;

8.基于**LangGraph、MCP**等构建Agent系统，结合ReAct范式，实现多Agent复杂系统;

9.了解Linux系统，进行团队的高效协作开发;

10.了解**LoRA、Prefix-Tuning**等高效微调方法，针对性提升模型表现.

工作项目经历

2025.6-2026.2 南京阿贝斯信息科技有限公司 **AI算法工程师**

项目名称：某制冷设备企业销售售后多Agent-RAG智能辅助系统

项目介绍：面向某大型制冷设备企业的销售与售后业务痛点，本项目构建了一个**融合多Agent架构与RAG**（检索增强生成）技术的智能辅助系统，旨在解决产品信息分散、响应效率低、人工查阅易出错等问题。系统通过统一接入非结构化文档（PDF/DOCX等）与结构化业务数据（ERP/CRM），实现对技术参数、认证文件、安装规范、配件库存及客户特殊要求等知识的精准检索与智能生成，显著提升标书制作效率与售后服务准确性。

技术栈：BGE-M3 向量模型、Milvus 向量数据库、RapidOCR、HyDE 检索策略、RRF 融合算法、bge-reranker-large、FastAPI、Redis 缓存、Qwen2.5-7B、RagAS 评估框架

主要工作职责：

1.参与多源文档预处理流程：使用 **RapidOCR** 提取扫描件文字，结合 ChineseRecursiveTextSplitter 实现**“父块+子块”分层切分**，保留技术参数与工艺说明的上下文完整性；针对表格区域引入结构化识别策略，将 OCR 准确率从 79% 提升至 94%。

2.参与设计并实现混合向量检索架构：基于 **BGE-M3** 同时生成稠密与稀疏向量，分别建立 IVF_FLAT 与

SPARSE_INVERTED_INDEX 索引，支持语义匹配与关键词精确检索，并通过 **RRF 算法**融合结果，使 Top-2 相关性提升 22%。

3.开发多Agent协同逻辑：基于**微调 bert-base-chinese 模型**实现意图识别（销售/售后），并分别构建销售Agent（自动生成标书技术章节）与售后Agent（整合配件清单、安全规范、服务排期和安装维修问题解答），通过结构化 Prompt 引导 LLM 严格基于检索内容生成，确保输出忠实度达 0.93。

4.集成动态业务数据：对接 ERP （企业资源计划系统）获取产线状态与交付周期，对接 CRM（客户关系管理系统） 获取客户历史工单与现场安全要求，实现静态知识与动态业务状态的深度融合，支撑“该型号30天内可交付”等关键业务判断。

项目流程:

1.需求与数据梳理: 针对销售与售后团队在查阅产品手册、技术参数、认证文件及动态业务数据（如产线状态、配件库存）时效率低、易出错的痛点，明确系统需统一接入**非结构化文档**（PDF/DOCX等）与**结构化业务数据**（ERP/CRM）。

2.数据预处理：使用**RapidOCR**提取扫描件文字，结合ChineseRecursiveTextSplitter进行“父块+子块”分层切分以保留上下文；针对表格区域采用结构化识别策略提升OCR准确率，并对接ERP/CRM接口同步实时业务数据。

3.基线方案构建：基于**BGE-M3模型**生成稠密与稀疏向量，存入**Milvus数据库**并建立IVF_FLAT与SPARSE_INVERTED_INDEX索引，初步实现混合检索能力，并通过**Redis缓存**高频查询结果。

4.核心系统搭建：构建多Agent协同架构，基于**微调bert-base-chinese模型**实现意图识别（销售/售后），分别开发销售Agent（自动生成标书技术章节）与售后Agent（整合配件清单、安全规范与服务排期），并通过结构化Prompt引导Qwen2.5-7B大模型生成忠实答案。

5.系统优化与策略增强：通过引入**HyDE策略**来处理模糊售后问题，采用**RRF算法**融合稠密与稀疏检索结果，并利用bge-reranker-large进行重排序；同时构建**RagAS评估框架**形成优化闭环。

6.系统部署：使用**FastAPI**异步框架部署系统服务,实现多Agent-RAG智能辅助系统在实际业务场景的高效运转。

7.效果评估与迭代：参考**RagAS评估指标**（上下文相关性、忠实度、答案相关性），通过在真实业务场景中测试，确保系统输出的准确性与可用性，并持续将高频新问题写入缓存以优化响应速度。

项目成果：销售技术文档自动生成准确率达 **91%**，人工修改率下降 **65%**；售后人员单次上门准备时间缩短 **40%**，配件携带错误率降至 **2%** 以下；客服咨询量减少 **58%**，内部知识利用率提升 **3.2 倍**。

2025.3-2025.6　　南京阿贝斯信息科技有限公司　　AI算法工程师

项目名称: 酒店客户体验评论五星评级系统 (基于BERT模型微调)

项目介绍:随着在线酒店预订的普及, 用户评论已成为消费者选择酒店的重要依据。本项目旨在解决现有的五星评价系统存在的一些明显局限性：评分与评论文本情感不一致，简单复制粘贴的无效评论干扰判断以及缺乏对评论内容的深度分析。为**自动校正不可靠评分星级**，该项目基于**BERT情感分析**结果调整数值评分，生成多维度体验报告（服务、卫生、设施等权重分析），并为消费者提供可视化决策支持，为管理者提供可操作改进建议。

技术栈: BERT、FASTTEXT、Pandas、jieba分词器、模型蒸馏、MLP、MySQL、FastAPI

主要工作职责：

1.参与训练FASTTEXT基线模型，将数据集通过**jieba分词器**处理成合适数据形式，通过基线模型训练测试结果，确定下面BERT模型微调方向。

2.参与构建**微调BERT模型加MLP多层感知器**修正的组合模型，并制定了将五分类任务分解成多个二分类任务的评论评级处理方案，显著提高了分类准确率。

3.参与训练微调BERT模型，通过**调节不同标签损失占比**权重来平衡标签的分布不均匀，用 '**高-低-高-低**' 的学习率调整模型，不断跳出模型在局部最优解处收敛的问题，经过训练最终模型的准确率由基线模型的**32%** 提升到最终模型的**82%**。

项目流程:

1.评论数据收集: 数据方面来自携程等平台的酒店评论数据，以及酒店官方平台评论数据。

2.数据预处理： 去除评论数据中的乱码字符、偏僻字，以及相似度很高的简单重复评论，最后对评论数据进行简单的初步分类和拓展性分析。

3.基线模型训练：对评论数据使用jieba分词器进行分词后，项目Baseline Model选择使用**FASTTEXT模型**进行初步评估和可行性分析。

4.BERT微调模型搭建：项目基于**BERT模型**进行微调和**MLP多层感知器**分类修正模型进行搭建。

5.模型训练和模型优化：面对数据集标签分布不均匀问题，使用**focal_loss损失**计算损失，并使用上下变化更大更灵活的学习率进行训练优化。

6.模型部署：使用**FastAPI**进行部署，并提供可视化界面，方便查看评论星级分类结果。

7.模型评估：参考内部测试评估指标，通过在测试数据集上进行测试，确保项目模型结果可用性。

项目模型成果:

模型通过多次训练后，在测试数据集上进行测试后准确率达**82%**，评分星级和评论文本情感不一致的问题在经过了该模型处理后得到了显著解决。

获奖经历

2023.9 **2023 年高教社杯全国大学生数学建模竞赛本科组全国二等奖**

担任建模手，编程手

在数学建模竞赛中，主要负责处理试题“多波束测线问题”中的海底测绘数据，对离散点进行梯度处理后，使用聚类分析等**数据分析**方法完成海底区域分块，并利用智能**退火算法**,针对不同区块建立相应的多波束测线路径模型，最后使用 MATLAB、**Python** 编写模型程序解决测线路径最短问题。同时作为**队长**协调参赛队内三人的分工合作情况，获得**高教社杯全国大学生数学建模竞赛本科组全国二等奖**。

自我评价

1本人性格乐观开朗，富有责任心，有耐心不怕困难。

2.同时拥有良好的沟通和协调能力，善于应变，能够快速融入团队并贡献价值。

3.具有良好的数学学科素养，思维敏捷，逻辑严谨。

4.对新技术充满热情,持续学习和探索AI领域的前沿技术。