

## 李卓豪

男 | 湖南 | 2年经验

19891811379 | 1415932874@qq.com

## 项目经验

## 基于模型微调的智能对话助手

**项目背景：**传统的人工客服或通用对话模型在金融科技平台存在明显不足比如对公司的货币政策不熟、不同的市场信息存在差异等，因此我们决定打造一款基于模型微调的金融科技智能对话助手。

**关键技术：**BLEU, ROUGE, 继续预训练 (CPT), 规则匹配数据清洗, 监督微调 (SFT), 直接偏好优化 (DPO)

**项目重点：**

**多维度评估体系设计：**

- 构建基于**BLEU-4/ROUGE-L**的自动化评估框架，设计覆盖语义匹配度（bertscore、bleu、rouge加权）、流畅性、专业度的三维评分标准，建立含5,000+样本的黄金测试集。
- 主导人工评估流程标准化，通过K-Alpha系数验证评估者之间一致性 ( $\alpha=0.86$ )，筛选出最优基线模型（HumanEval达标率74.2%，优秀率22.5%）。

**领域自适应继续预训练 (CPT)：**

**设计多阶段数据清洗Pipeline：**

- 规则匹配：构建专业术语正则库（覆盖90%领域实体）。
- 分类器过滤：基于**RoBERTa**训练质量判别模型（准确率达92.3%）。
- MinHash去重：多线程去重，降低语义重复率至8%以下；

基于13B token高质量语料（9k+专业书籍/80k+论文/领域网页）使用**LoRA**在7B基座模型上实现参数高效微调，使专业问答的达标率提升了8个百分点（75.3%→83.4%）

**监督微调(SFT)与数据增强：**

- 制定企业级数据标注规范：构建超4000条种子数据（含单轮65%/多轮对话25%、长文本生成10%）；
- 开发数据增强Pipeline：基于T5的语义改写 扩增3.2万条训练样本，对抗生成（NLP Aug [词汇、字符扰动、回译、模拟打字错误]）提升模型鲁棒性；

最终人工评估达标率92%（+18% vs基线），优秀率达42.1%

**基于人类偏好的强化学习优化：**

- 基于DPO训练框架：采用Moving Window策略解决数据分布偏移问题且KL散度控制在 $1.5 \pm 0.3$ ；
- 设计Point-wise偏好数据采集机制，迭代优化1,200组对比样本；最终模型在真实业务场景中达到98.2%达标率，65.8%优秀率

## 基于多Agent协作的智能客服系统

**项目背景：**在金融科技行业，交易所存在复杂问题处理能力不足如用户咨询涉及交易规则、链上操作、合约安全等复杂场景且知识碎片化严重，导致回答准确率低。为此，我们研发了基于多Agent协作的智能客服系统，通过混合检索架构（RAG+BM25）与多Agent任务路由，实现复杂问题的精准拆解与高效响应。

**关键技术：**混合检索架构（RAG+BM25）, 多Agent workflow, LLM路由, 文本嵌入优化

**项目重点：**

**多Agent工作流设计：**

- 意图Agent：对用户的咨询意图对任务进行分流。提升召回精度、降低响应时间。
- 检索Agent：集成混合检索（BGE-m3+BM25），实现1W+知识条目的毫秒级召回。
- 生成Agent：采用DeepseekV3模型对生成文本风格微调适配。

**知识库增强与语义分块优化：**

- 创新动态分块算法：基于**Sentence-BERT**计算语义相似度（阈值 $>0.88$ ），实现上下文感知分块，对表格内容进行仅保留关联行列。
- 假设问题生成：为每个知识块自动生成3-5个衍生问题，提升知识召回准确率。

**多阶段检索系统架构设计：**

构建Retriever-Reranker双阶段框架，实现从粗排到精排的递进式知识筛选：

- 初筛层：采用BGE-M3生成1W+知识条目Embedding，召回Top100，Top10命中率92.4%；
- 精排层：引入BGE-Rerank-Large和BM25稀疏检索（ $k1=1.5$ ）对Top100结果重排序使命中率提升至97.8%，回答准确率提升至96.6%；

## 基于多Agent工作流的智能财经日报系统

**项目背景：** 传统推文的生成依赖人工收集、整理和排版，为了解决运营每日搜寻大量内容发布推文养号花费大量时间，我们设计开发了多Agent日报系统。

**关键技术：** 多Agent协同调度（Coze）,动态信息抽取与摘要生成,多模态文生图,知识库增强召回

**项目重点：**

### 多Agent任务分解与自动化流程设计：

- 构建 Agent协作框架，将复杂任务分解为 信息爬取→内容精炼→关键词提取→文生图→排版推送，5个独立Agent，降低单次任务复杂度，提升整体准确性。
- 采用 Coze平台 实现Agent调度，结合定时任务触发器，实现全自动化日报生成（从数据采集到推送全程无人干预）。

### 财经信息动态爬取与智能摘要：

- 开发 自适应爬虫系统，实时监控10+主流财经媒体（Bloomberg/金十/金色财经等），每日处理 200+条财经新闻。
- 基于 LLM摘要优化技术（Qwen-Max+规则过滤），生成 高信息密度 的财经要点，关键信息保留率 >98%
- 结合 关键词抽取算法（Qwen-Max+领域词典），提取核心术语用于后续多模态生成。

### 图片知识库构建：

- 对历史财经图片进行 分层标注（通用标签+专业标签，如「美联储加息」「某CEO人像」）
- 开发 双路召回系统（标注匹配+CLIP语义检索），配图精准率 达92%。

### 飞书自动化推送与运营增效：

- 集成飞书开放API，实现 定时推送，上线后每日运营时间从 2小时压缩至10分钟（效率提升91.6%）

## 工作经验

### 北京恒泰博远科技有限公司 - 大模型算法工程师

2023-06 ~ 至今

- 主导金融科技领域智能对话系统的研发，通过领域自适应预训练（CPT）和监督微调（SFT）优化7B基座模型，结合DPO偏好优化模型在对应业务场景的表现。
- 设计多Agent智能客服系统，负责知识分块及落库工作，创新使用动态语义分块技术，实现复杂问题精准拆解，提升了模型回答准确率。
- 开发多Agent财经日报系统，通过Coze平台协同调度爬虫、采用多AGENT协同工作，摘要生成及多模态配图，实现全自动化推送，提升了运营养号效率。

## 教育背景

### 湖南农业大学 - 智能科学与技术 - 本科

2020-09 ~ 2024-06

## 荣誉证书

- 中国软件杯全国三等奖
- 2022届高教社杯省级二等奖
- mathorcup、华东杯省级三等奖
- 国家级中级软件设计师

## 技能特长

- 熟悉python编程，linux开发环境，了解C/C++。
- 熟悉pytorch、huggingface等深度学习工具。
- 熟悉RAG、提示词工程，有项目落地经验。
- 熟悉LLaMA-Factory训练框架，掌握LoRA、QLoRA等高效微调手段。
- 掌握langchain框架的使用。