

杜旭东 · 算法工程师 / 大模型工程师-2025

☎ 电话: 19520549015 ✉ 邮箱: 993820876@qq.com 🌐 CSDN: 橙子小哥的代码世界 (15w+ 浏览)

🎯 求职意向

目标岗位: AI Agent 算法工程师 / 大模型工程师

🚀 职业综述

拥有多年大数据与大模型工程经验, 熟练使用 **ChatGLM3-6B LoRA/P-tuning** 完成本地微调, 结合 RAG系统与百万级知识图谱, 实现问答系统准确率 $\geq 92\%$ 、响应延迟 $< 500\text{ ms}$ 。擅长 Spark/hive ETL 架构落地, 具备完整从模型训练、优化、部署到系统集成的全链条能力。

🔧 核心技术

- **LLM 微调 & RAG**: ChatGLM3-6B LoRA/P-tuning ($r=8-32$, $\alpha=32$, dropout=0.1, 参数占比 0.35%, 显存 $\leq 24\text{ GB}$)
- **知识图谱构建与检索**: 150w实体、600w关系, Neo4j + Cypher 查询优化, 结合向量检索构建 Graph-RAG 系统
- **NLP 算法实现**: BiLSTM-CRF, CasRel (SPO F1 93%), BERT, seq2seq, transform, 文本分类, NER, 意图识别
- **大数据流水线搭建**: Spark + Hive + tidb + tableau + 邮件架构, 日处理 ≥ 3 万条日志, 延迟由小时级降至10分钟
- **模型推理优化**: 蒸馏、量化、剪枝, 使响应延迟降低 $\geq 30\%$ 、分类精度保持 $\geq 91\%$
- **系统部署**: FastAPI + Flask + Docker + Git + Linux, 打包成微服务稳定运行
- **编程语言与框架**: Python · SQL · Java · PyTorch shell
- **大模型技术**: langchain, lora, 提示词工程, pturning, vllm, kv 缓存, milvus, bm25, rerank, glashattention

👛 工作经历

高级算法工程师 - 数据智能项目组 (佰聆数据股份有限公司)

*2022.06 - 至今

- 创建知识图谱平台 (150w实体、600w关系) 与 Spark/Kafka 流水线, Neo4j 提供毫秒级查询

- 使用 ChatGLM3-6B LoRA 微调模型，提升问答准确率至 91%，显存控制在24GB
- 构建 Hybrid Graph-RAG 系统，命中率达93%、MRR88%，客服介入次数下降20%
- 日处理 ≥3万条日志，流水线耗时从小时级缩至10分钟，使用 Git 管理并部署在 Linux + Docker 环境
- 搭建120+ 用户知识库与问答助手，有效提升部门效率20%，问答准确率达70%

算法工程师 - 创新孵化项目组

*2020.01 – 2022.05

- 文本分类模型精度由 78% 提升至 94%，适配缴费预测与客户情感分析
- 通过蒸馏与剪枝优化 BERT，延迟减少 30%、准确率保持 ≥91%
- 搭建 Spark + Hive ETL，支撑日处理平台，保障模型训练langchain框架

大数据开发工程师 - 算法策略组

*2019.08 – 2020.01

- 构建分钟级调度系统，日清洗日志达3亿条，支撑大规模数据处理
- 使用 hive SQL + Tableau 构建数据仓库与报表，提升运营与产品决策效率

大数据开发工程师 - 离线数据中心

*2017.08 – 2019.10

- 构建覆盖30+部门的 Tableau 实时报表系统
- 数据ETL

项目成果

政务 Graph-RAG 问答平台

整合 800w 政策文档，通过 LangChain + Embedding 构建知识库；ChatGLM3-6B LoRA 模型实现问答准确率92%、命中率93%、MRR88%；FastAPI+Docker 微服务部署。

部门数据服务助手 rag + chatglm3-6b + 企业微信

政务知识图谱抽取系统

采用 BiLSTM-CRF + CasRel 联合抽取，提取结构化数据至 Neo4j 图谱，SPO F1 达93%，支持 Cypher 查询调用。

法律文本多分类系统

Pipeline 模型集成：RF→FastText→BERT，准确率离婚类95%、诉讼类91%；蒸馏至 TextCNN，模型体积降低85%、准确率保持91%，稳定提供日常处理服务。

✅ 一站式数据平台的搭建

数据仓库+ Hdfs + hive + spark + tables + tidb + 内部邮件，稳定提供数据分析师的日常分析处理服务。



荣誉认证

- Coursera项目管理认证
 - 优秀员工奖
 - 全国大学生数学建模三等奖
-



自我评价

具备独立设计与落地 LLM 部署+ 知识图谱 + rag 系统的能力，具备大数据仓库ods -dwd -dm -app 搭建能力，以及大数据ETL报表平台的，擅长使用 LoRA 微调大模型并实现高效问答系统。拥有坚实的工程化背景与团队协作能力。未来期待深入 AI Agent 架构与 Tool-Calling +mcp 机制，实现上下游系统的全面融合。