



谭一鹏

18521353005 | darcher005@outlook.com | 深圳
39岁 | 男 | 汉族

综合概要

- 1. 有不错的数理基础和逻辑思维能力，英语熟练。
- 2. Python开发9年经验，熟悉Python数据开发生态，包括NumPy，Scikit-learn，Pandas，Matplotlib，Jupyter Notebook等工具包。同时熟悉容器Docker技术和Python Web框架如FastAPI。另外在Elasticsearch搜索引擎方面有着丰富的开发经验。
- 3. 熟练掌握经典 ML/NLP 技术，如中文分词，TFIDF，文本分类/聚类，主题模型，SVM，随机森林，lightGBM 和 XGBoost。熟悉神经网络NLP技术，比如Word2Vec，Fasttext，BERT。
- 4. 熟悉大模型（LLM）训练和部署的相关技术：L0预训练，指令微调（SFT），价值对齐（RLHF），参数微调（Q/LoRA），以及KV Cache，GQA和Paged Attention等加速技术；还熟悉主流大模型应用技术：Prompt Engineering，RAG，Agent，以及Langchain框架的使用。

教育经历

中国科学技术大学	985 211 双一流	2008年09月 - 2011年05月
粒子物理与原子核物理 硕士 近代物理系		合肥
上海理工大学		2004年09月 - 2008年07月
应用物理 本科 应用物理系		上海

工作经历

城迈科技（南京）股份有限公司（OPPO外包）	2024年07月 - 2025年04月
后端工程师（NLP算法）云数能力中心	深圳
完成公司安排的各项舆情监控相关数据/算法开发：	
1. 训练迭代打标模型	
2. 开发崩溃日志快速聚类算法	
3. 使用LLM提取用户反馈的关键内容	
北京外企德科人力资源服务上海有限公司	2023年01月 - 2023年08月
python测试开发 华为云	深圳
1. 进行测试用例设计以及评审，手动测试以及自动化用例的Python脚本编写。	
2. 参与过两款 NoSQL 数据库 Redis 和 InfluxDB 的内核测试以及数据库管控层的测试，因此对这两款数据库的底层原理和管控流程都比较熟悉，对于华为云生态也有较深的了解。	
3. 参与过数据库云服务测试环境的搭建和运维。	
今日人才	2022年03月 - 2022年09月
算法工程师	深圳
1. HR 实体召回服务	
基于 Elasticsearch 开发了一个人力资源实体召回平台，为搜索引擎和人岗匹配的召回阶段提供统一的搜索服务。开发涉及简历和岗位数据的清洗、分析以及结构化，使用词嵌入模型向量化部分自由文本内容，进而利用余弦相似度等量度对这些文本进行向量召回，并作为相关性得分的一个因素提高搜索的准确度和覆盖率。	
2. 候选人简历去重引擎	
基于 Elasticsearch 开发了一个候选人简历去重引擎。利用候选人简历中的 ID 类信息和其他有区分度的信息，加上编码成 MinHash 的履历关键信息，可以准确高效地对海量简历进行基于候选人的识别和去重。工作的重点在于分析数据集合的完整性分布以及结合不同覆盖率不同可信度的多个字段组合进行可信的识别。	
深圳呗佬智能有限公司	2021年03月 - 2022年01月
算法工程师	深圳
1. 多维度智能人岗匹配引擎	
基于 Elasticsearch（ES）搜索引擎进行人岗匹配模块的Python开发。基于结构化的岗位画像和候选人简历，开发实现了简历的自定义多维度匹配算法。根据产品构思的业务模型（比如岗位模型和候选人的能力模型）以及功能需求，迭代研发设计相关实体的属性结构以及实现相应打分的ES Query，主要产出是实体在ES中的索引结构和相对应的 DSL 语句，以到达最相关简历的快速召回。另外使用了FastAPI工具来向外提供服务。	
平安金融管理学院（集团人力资源中心）	2018年08月 - 2021年02月
数据工程师 集团人力资源中心	深圳
1. 人才库搜索开发	
承担人才库搜索的开发。前期主要承担 Elasticsearch 的开发设计工作（索引规划和 Query 设计），后期主要是负责候选人	

数据（简历以及招聘流程业务数据）的 ETL 工作（Python），使用了MySQL和MongoDB等数据存储以及Kafka，RocketMQ等消息队列。之后还承担了人才库搜索项目 SA（需求分解）的职责。

2. HR 极速入职/审核

对各类（招聘和员工服务）证书进行关键字段的提取以节省人工审核成本和改善员工体验。在平台提供的通用 OCR 能力之上独立完成四类证书（学位证明、离职证明、结婚证和病假单）自动提取内容模块的开发，以及内外多个环境的 Web 服务部署工作（基于Docker）。其中版式较规范的学位证明和结婚证都达到了 90%以上的准确率。其中结构化的任务中涉及 NLP 的 NER（命名实体识别），文本分类，还有向量召回等子任务。

深圳市英威诺科技有限公司 2017年10月 - 2018年07月
算法工程师 深圳
负责给产品提供西班牙语的自然语言处理技术支持，主要包括新闻文类，关键词提取，相关新闻聚类 and 新闻推荐系统。

上海合合信息科技发展有限公司 2017年01月 - 2017年08月
算法工程师 上海
负责信息流推荐系统的搭建（主要使用 Spark SQL/Streaming 整合Redis，Elasticsearch），以及运用自然语言处理技术完成定制化的舆情监控任务。

上海态格信息技术有限公司 2014年07月 - 2016年12月
数据分析师 上海

1. 社媒文本预测股市大盘涨跌

基于文本分类探索网络文本对股票涨跌的预测能力。主要是基于词袋模型使用TF-IDF矢量化文本，探索了当时主流的分类器如SVM和随机森林（Random Forest），最后选择了前者。

2. 基于财经领域知识图谱的新闻推荐

在基于文本相似度的新闻推荐中，很难对关键词进行阅读兴趣的推测（比如从“K线”到“股市”），受启发于 Google 的 Knowledge Graph，构建了一个财经新闻领域的概念层级结构。构建方式主要是基于上下位词（上“股市”，下“K线”）在语料集中具有不同的分布方式，再基于关键词提取算法找到第三层级的具体关键词。

3. 其他舆情监控任务，比如热词分析，情感分析，关键词提取，内容抽取，新闻聚类等。