# 王博

求职意向 : NLP算法工程师



年龄 : 25 | 电话 : 15091437996 | 工作年限 : 2年

性别 : 男 | 邮箱 : bryan_0026@163.com

## 教育经历

2021.09 - 2024.06 西安建筑科技大学 硕士

2017.09 - 2021.06 西安建筑科技大学 本科

- **主修课程** : 自然语言处理 , 模式识别 , 机器学习 , 深度学习 , 图像处理等 ;
- **学术论文** : 在**多模态,图像识别**领域发表《多尺度上下文引导特征消除算法》等**3篇EI期刊**论文, 通过挖掘类别文本模态信息、多模态特征融合等方法构建判别性特征,显著提高细粒度场景下的图像分类效果 ;
- **专利** : 在**图像识别**领域发表《一种基于关键鉴定部位的商品真伪鉴定方法及系统》等**2篇专利** ;
- **荣誉奖项** : 获得研究生**一等学业奖学金**1次 , 二等奖学金1次 , “**服创赛**”二等奖 , “**华为杯**”数学建模二等奖 , “**蓝桥杯**”Python程序设计省级二等奖 , “**互联网+**”省级铜奖 , 陕西省第七届研究生创新成果展二等奖等多项国家级、省级、校级创新创业奖项 ;
- **语言水平** : 已通过**CET-6** , 能够阅读人工智能领域的英文文献 .

## 专业技能

- **大模型与生成式AI** : 精通BERT、GPT、T5等模型的微调与优化 , 掌握**LoRA、P-Tuning**等参数高效微调技术; 精通基于**LangChain/LangGraph**的**RAG系统设计**，具备从文档处理→向量索引→检索优化→生成的完整链路经验；拥有丰富的**Prompt Engineering**指令微调技巧 ;
- **模型部署与性能优化** : 熟练使用**Docker、FastAPI**实现主流大模型的本地化部署与API服务 ; 具备**vLLM、TensorRT**等推理引擎优化经验 ; 掌握量化、知识蒸馏、剪枝等压缩方法 ;
- **机器学习与深度学习** : 深入了解多种词嵌入算法 , 熟练使用随机森林、XGBoost等算法解决分类、回归问题; 熟悉CNN、**RNN**、**Transformer**等架构 ;
- 多模态与跨模态学习 : 熟悉ViT等视觉架构 , 具备图像分类、目标检测等任务的落地经验 ; 掌握CLIP、BLIP等跨模态表示方法 , 了解LLaVA、QWen-VL等端到端多模态大模型原理 ;
- **数据处理与工程能力** : 精通**MySQL** , 熟练使用Pandas、NumPy进行大规模数据处理 , 拥有丰富的数据增广、特征提取、数据清洗、异常检测实战经验 ;
- **关键框架与编程** : 熟练掌握Python , 精通PyTorch , 熟悉HuggingFace Transformers生态，具备自定义模型层和训练循环开发能力 ; 熟练使用Git、Linux、Shell等开发工具，具备良好的工程实践

## 工作经历

2023.10 - 2026.2 南京数之赢科技有限公司 NLP算法工程师

- 全面负责从多模态数据处理、知识库构建、模型微调至系统部署的全流程研发，主导RAG（检索增强生成）技术在多个业务场景中的落地与优化。
- 主导设计并构建大规模多模态企业知识库，高效处理超40万条文档、图像及结构化数据，通过PyMuPDF、OCR与自动化清洗流水线实现信息提取与结构化存储，为智能问答系统提供精准、实时的数据支撑。
- 深入应用模型微调与优化技术，基于LoRA、知识蒸馏及量化等方法，针对业务场景对Qwen、BERT等模型进行领域适配与性能优化，关键项目模型准确率提升至90%以上，显著改善推理效率与响应速度。
- 基于LangChain与FastAPI搭建高可用系统架构，实现“意图识别-检索-生成”的自动化闭环。通过优化混

合检索策略与排序模型，有效提升问答召回率与准确率，驱动多个项目成功上线并降低人工运营成本超30%，相关技术方案已形成标准化沉淀并在团队内推广。

## 项目经历

### 智能咨询与院端服务医疗Agent 2025.9-2026.2

面向院端客服人工压力大、患者导诊流程繁琐等痛点，主导研发一款服务于普通患者的医疗智能 Agent 系统。依托大模型与 Milvus 医疗知识向量库，结合 A2A 与 MCP 标准协议，构建覆盖“院前导诊-院中服务-院后随访”的全生命周期智能服务体系，大幅提升患者就医体验与医院运营效率。

**技术栈**： LangChain/LangGraph,DeepSeek, A2A协议, MCP协议, Milvus, Asyncio

**项目职责**：

- **A2A+MCP多智能体协同架构设计**: 搭建了负责“意图识别与任务拆解”的主控 Agent，以及分管“智能导诊、科普答疑、院内服务”等专项任务的功能 Agent. 引入 MCP (Model Context Protocol) 协议,将医院业务接口与 Milvus 知识检索能力封装为独立的 MCP Servers,以标准化、即插即用的方式动态挂载工具。
- **"全局状态机+增量投喂"上下文策略**:压缩复杂长对话中的多 Agent 上下文,主控 Agent 负责维护核心患者画像，并在 A2A 通信时只传递必要的增量 Context，避免了 Token 爆炸和功能 Agent 幻觉。
- **并发控制与多路会话聚合**： 针对多 Agent 协作带来的长链路延迟问题，重构意图处理逻辑为 Asyncio 异步协程模式，实现多子任务的并行调度；结合流式输出与工具级结果缓存，有效解决了多 Agent 调度拥塞，使系统整体响应速度显著提升。
- **口语化医疗意图与导诊算法优化**： 深度优化 NLP 意图识别模块，结合大模型的语义重写能力与 Milvus 医疗知识图谱/向量库，解决患者口语化症状描述匹配不准的问题。构建“多路召回”策略，显著提升了“症状-科室-医生”匹配的准确率与容错率。

**项目成果:**

- 替代60%以上人工导诊、客服咨询工作，导诊台接待压力降低38%；
- 优化Milvus向量检索策略，医疗知识检索准确率达91.5%，智能导诊科室匹配准确率93%，常见疾病科普、就医流程咨询回答准确率89%。

### 医疗临床辅助决策RAG系统 2025.4-2025.8

主导核心检索层设计，基于MySQL+Milvus双库架构，落地BM25与向量混合检索方案，覆盖诊断/用药/手术三大临床场景，检索效率提升70%、综合精准度93%，合规支撑医护临床决策

**技术栈**：BM25、BGE-M3 Embedding、BGE-Reranker、MySQL、Milvus、Redis、MC-BERT。

**项目职责**

- **数据架构设计**：MySQL存储病历关联的结构化数据，Redis缓存高频症状病历；Milvus构建4类向量集合，通过BGE-M3模型将非结构化文本为稀疏-稠密向量，IVF-FLAT索引优化存储，承载超100万条向量数据，实现百万级秒级检索。
- **检索方案搭建**：构建BM25+向量混合检索体系，jieba分词+定制停用词表优化全文检索，按疾病类型设置分级阈值；BM25检索不足时，触发Milvus多策略混合检索，结合BGE-Reranker重排序，保障语义匹配精准性。

**项目优化**

针对罕见病检索漏检、语义偏差等问题优化；构建病症加权评分规则，基于MC-Bert实现症状拓展与疾病预测；新增Milvus标量字段，标量过滤+向量匹配避免全量检索低效，拆分主次症状实施多轮检索+重排序，提升匹配质量。

**项目成果**

- 诊断辅助：相似病例匹配精准度92%，罕见病指南覆盖度98%，诊断一致性Kappa>92%，决策时间缩短60%。
- 用药辅助：特殊人群用药禁忌识别准确率99%，药物相互作用召回率95%，用药风险降低80%。
- 手术辅助：操作规范匹配精准度94%，并发症预测有效性74%。

- 医院效益: 临床使用率70%+，误诊率降4%，医生满意度9.1分；单患者检查成本降20%

## 临床用药辅助知识图谱系统 2024.12-2025.3

面向严谨条件逻辑的临床循证用药辅助系统，通过医疗知识图谱+大模型结合锁定多跳逻辑关系，实现“疾病-特殊人群-禁忌症”多约束下的个体化用药推荐与安全风险预警，解决临床用药规范难、风险管控难的痛点。

**技术栈：**LangGraph、Neo4j、Faiss、LoRA、MC-BERT-LARGE、RoBERTa、Qwen

**项目职责：**

- **双路径实体与关系抽取**:搭建Neo4j医疗知识图谱，设计8类核心医疗实体与7类语义关系，基于微调MC-BERT-LARGE实现药品说明书的实体与关系批量抽取,基于qwen-max搭建Few-Shot框架完成长文本临床指南Chunk切分的知识提取，实体抽取F1值达92%，关系抽取F1值达88%；
- **归一化与向量索引构建**:基于医疗RoBERTa完成实体归一化与权威术语映射，构建Faiss向量索引，实现「Neo4j结构化检索+Faiss语义检索」的双路召回，提升知识召回的全面性与精准度；
- 基于 LangGraph 的 **Agentic 检索工作流**:实现用户输入实体提取、Cypher自动生成、用药安全校验、循证内容整理的全流程自动化，通过分支校验管控医疗输出风险；
- **LoRA微调结构化输出**:完成全链路数据合规设计，所有知识节点标注权威来源与循证等级，满足临床医疗可追溯要求。

**项目成果：**

- 构建5万+药品、症状、功效、疾病实体，及8万+知识关系；
- 双路检索召回率较关键词检索提升18%,系统用药推荐与禁忌症预警准确率达94%;
- 覆盖95%常见临床用药场景,医师整体满意度87%+;

## 医疗文本结构化与术语标准化系统 2024.5-2024.11

主导医疗文本结构化与术语标准化系统落地，解决非结构化医疗数据互通难、人工录入低效、医疗术语不统一的行业痛点。通过医疗专用预训练模型与渐进式增强策略，实现数据质量与临床效率的双重提升，为后续AI辅助诊疗筑牢标准化数据基础。

技术栈：MC-BERT-Large、RoBERTa、ICD-10编码、渐进式数据增强

**项目职责**：

- **数据建设：**主导电子病历、检验报告等医疗数据的脱敏、标准化处理，构建10w+条高质量私有医疗数据集，配套医疗标准化术语表与ICD-10映射词典，支撑模型落地的核心数据需求。
- **结构化提取：**基于**MC-BERT-Large**实现临床文本命名实体识别（NER），优化人工录入流程，实体识别精确率95%,人工录入时间缩短60%以上，显著降低医护人员人力成本。
- **术语标准化：**通过**“同义词表+RoBERTa”**分层协同完成医疗术语标准化（CTC），实现多源数据统一ICD-10编码转换，采用严格2分类语义相似性学习，无医学错误术语生成，高频医疗术语编码覆盖度90%以上，数据一致性提升80%+。
- **模型优化：**引入**渐进式增强与组合数据增广策略**，缓解医疗数据标注不足、分布不均痛点，提升模型泛化能力与落地稳定性，保障系统在临床场景的安全可控。

## 个人评价

热爱技术钻研，专注 NLP 与智能问答领域，在 RAG 检索增强、知识图谱构建等方向有丰富实践经验。

擅长用户需求分析与跨部门协作，具备数据驱动决策思维。

持续关注 AI 行业动态，追求以技术创 新提升产品体验，为业务场景提供高效解决方案。