

nlp 简历

个人信息	联系方式
姓名：崔建新	电话：13223549116
年龄：26 岁	邮箱： 13223549116@163.com
籍贯：山西	

教育背景

2021.09~2021.07

山西农业大学

专业：软件工程

工作经历

- 2020.11~2025.06 北京易联达商务服务有限公司 岗位：算法工程师
- 数据分析与处理：对大规模医疗文本数据进行分析处理，为模型训练提供优质数据。
 - 模型设计与实现：基于深度学习框架与前沿算法，根据公司业务需求，设计并实现高效的 NLP 模型。
 - 性能测试与优化：针对现有 NLP 模型，通过算法优化及架构调整，全面提升模型性能与推理速度。
 - Java 微服务高可用实现：使用 springcloud 生态圈实现高可用的银联网联服务 , 实现动态切换数据

技术能力栈

- 熟悉大型语言模型应用开发，包括基于 LangChain 等框架进行大模型的应用构建；
- 熟悉迁移学习与 LoRA 等高效微调技术，能够针对不同任务对预训练模型进行轻量级优化；

- **熟练掌握深度学习框架**，如 PyTorch、TensorFlow、Keras，具备从模型搭建、训练到调优的全流程经验；
 - **熟练使用主流 NLP 工具与库**，包括 Word2Vec、FastText、Transformers、HuggingFace 等，能够快速实现文本特征提取、语义理解等任务；
 - **熟悉 BERT、GPT 等主流预训练模型的应用与优化**，能根据具体业务场景选择合适的模型结构；
 - **具备 Python 编程能力**，熟悉 Pandas、NumPy 等数据处理工具，能够高效完成数据清洗、预处理和分析工作；
 - **具有 Java 微服务开发经验**，熟练使用 Spring Cloud 生态圈（包括但不限于 Eureka, Ribbon, Feign, Hystrix, Zuul, Gateway），擅长构建高可用性的分布式系统；
-

工作经历

项目一：智能阳标识别系统

项目背景：

随着医院体检业务的不断增长，体检项目数量和体检报告种类日益繁多，导致医生在审阅大量纸质或电子报告时效率低下，且容易出现漏诊、误诊的情况。为了解决这一问题，提升体检结果的处理效率与准确率，我们开发了一套基于人工智能技术的“体检报告智能阳标识别系统”。

技术架构：

- **自然语言处理（NLP）**：采用 Sentence-BERT (SBERT) 等预训练语言模型对非结构化体检报告文本进行语义理解。
- **OCR 图像识别**：针对扫描版 PDF 体检报告，marker-pdf 将图片转化成 md, 使用 Tesseract OCR 进行文字提取，转换为可分析文本。
- **信息抽取与规则引擎**：构建体检指标关键词库与单位区间规则，结合正则匹配与实体识别（NER）提取关键指标值。
- **异常判断逻辑**：根据医学标准制定阳性判断规则（如血糖 > 6.1 mmol/L 为阳性），自动标注异常项。
- **部署方式**：Docker 容器化部署，后端服务采用 Flask 框架搭建。

职责描述：

- 参与需求分析，深入理解体检流程及医生审阅痛点，协助定义阳标规则体系。
- 负责 OCR 模块的选型、训练与优化，提高复杂格式体检报告的文字识别准确率。
- 搭建 NLP 处理流程，完成从原始报告到结构化指标提取的全流程自动化。
- 设计并实现异常判断引擎，支持灵活配置不同科室、检查类型的阳性标准。
- 配合前后端团队完成接口联调与性能优化，确保系统在高并发场景下的稳定性。
- 对模型输出结果进行评估与迭代优化，持续提升系统的召回率与准确率。

项目成果：

- 系统上线后，体检报告的阳标识别平均耗时由原来的 5 分钟 / 份缩短至 10 秒 / 份。
- 在测试集上达到：召回率：92.4%，精确率：89.7%，F1 值：91.0%
- 医生审核效率提升约 80%，有效降低漏诊率，减少重复性人工劳动。
- 支持多模态输入（PDF、图片、Word 等格式），兼容性强。

项目难点与挑战：

- 体检报告格式多样、术语不统一，需构建灵活的信息抽取机制。
- 不同医院使用的参考范围不一致，需设计动态配置规则系统。
- 数据格式不规范，处理数据整合成统一格式

项目二：医疗智慧助手

项目背景：

随着企业对知识服务自动化的需求日益增长，为了提升用户交互体验与问题解决效率，我们构建了一套融合大语言模型（LLM）、检索增强生成（RAG）、意图识别、Agent 流程编排于一体的智能问答与任务执行系统，支持从自然语言提问到任务自动完成的全流程智能化处理。

技术架构：

- **大语言模型（LLM）**：基于 ChatGLM3-6B 大模型，进行微调适配，用于理解复杂语义并生成自然语言回复。
- **检索增强生成（RAG）**：使用 LangChain 框架，结合 Elasticsearch 与 Faiss 向量库，实现高效的知识召回与上下文注入，提升回答准确率。
- **意图识别与实体抽取**：采用 sentence-BERT 模型，识别用户问题中的意图与关键参数，用于路由不同 Agent 模块。

- **前端交互与后端服务：**前端采用 React 实现交互界面；后端基于 FastAPI 提供服务，支持并发请求与异步处理。

职责描述：

- 主导 NLP 模块开发，包括意图识别、实体抽取、语义相似度计算等，提升系统对用户输入的理解能力。
- 设计并实现基于 LangChain 的 RAG 流程，优化知识召回与生成阶段的匹配效果，显著提升问答准确率。
- 构建 Agent 任务调度引擎，定义任务状态机与工具调用协议，实现多 Agent 间的协同与反馈机制。
- 参与大模型本地化部署与推理优化，使用 LoRA 微调技术降低资源消耗，提升响应速度。

项目成果：

- 系统上线后，用户咨询类问题的首问解决率提升至 87%，平均响应时间控制在 1.5 秒以内。
- 在测试集上达到：意图识别准确率：94.2%，问答准确率：89.6%，Agent 任务完成率：82.3%

项目难点与挑战：

- 用户提问多样性强，需结合规则与深度学习方法提升泛化能力。
- 多 Agent 之间信息共享与状态同步复杂，需设计统一的状态管理机制。
- 大模型输出不可控，需引入提示工程、约束解码与专家审核机制确保输出质量。
- 需对接多个业务系统接口，涉及权限认证、数据脱敏等安全合规问题。

项目三：直连银联网联商户历史退货系统开发与维护

项目背景：

由于历史交易数据量过大，采用传统数据库无法承受，因此开发了一个专门用于处理历史退货的功能模块。通过优化数据存储策略，确保系统能够高效处理大量历史数据查询请求。

技术架构：

- **数据存储：**采用 HBase 存储超过一个月的历史交易数据。
- **数据处理：**动态链接数据库，使用 Spark 读取并处理批量数据，HBase 库进行查询。
- **后端框架：**基于 Springcloud 框架进行后端开发，实现高效的数据处理和业务逻辑。

职责描述：

- 1. **数据迁移与存储优化：**
 - 设计交易数据存储策略，设置交易数据保留时间为一个月，之后迁移到 HBase 数据库。
 - 编写数据迁移脚本，从传统数据库中抽取历史数据并导入 HBase。
- 2. **数据可视化支持：**
 - 提供数据可视化支持，便于用户直观了解历史退货情况。

项目成果：

- 成功解决了历史数据存储问题，提高了系统的稳定性和查询效率。
 - 实现了高效的历史退货查询功能，显著提升了用户体验。
 - 确保了数据的安全性和完整性，满足了业务需求。
-

项目四：配合大数据的数据整合

项目背景：

为了保证数据的安全性、稳定性，每天需要从数据库中抽取数据进行验证，并发送到大数据平台进行进一步分析和处理。

技术架构：

- **数据抽取：**使用 Socket 短连接技术动态切换数据库，从指定的交易库中抽取当天增量数据。
- **数据压缩与转换：**对抽取的数据进行 gzip 压缩，并转换为 XML、CSV、verf 等多种格式以适应不同平台的需求。
- **数据同步与校验：**将处理后的数据同步给数据平台，并设计数据校验机制确保数据的一致性和完整性。

职责描述：

- 1. **数据配置管理：**
 - 负责维护和管理 MDB 管理库中的 big_data_para 配置项，确保数据配置的准确性和实时性。
- 2. **数据抽取与处理：**
 - 使用 Socket 短连接技术动态切库，从指定的交易库中抽取当天增量的数据。

- 编写了高效的数据抽取逻辑，确保数据的完整性和准确性，同时尽量减少对交易库的影响。

3. 数据压缩与格式转换：

- 对抽取的增量数据进行 gzip 压缩，减少了数据传输的带宽占用和存储成本。
- 将压缩后的数据转换为 XML、CSV、verf 等多种格式，以满足不同数据平台的需求。

4. 自动化脚本编写：

- 编写了自动化脚本，实现了数据压缩和格式转换的批量处理，大幅提高了工作效率。

5. 数据同步与校验：

- 设计并实现了数据校验机制，包括 verf 文件的生成和校验，确保数据的完整性和一致性。
- 监控数据同步过程，及时发现并解决数据同步中的问题和异常。

项目成果：

- 数据抽取过程得到了显著优化，提高了数据抽取效率和稳定性。
 - 数据压缩与格式转换流程实现了自动化，减少了人工干预，提高了工作效率。
 - 确保了数据的安全性、一致性和及时性，满足了大数据平台的数据需求。
-

自我评价

作为一名 NLP 算法工程师，我始终专注于将前沿的深度学习技术与实际业务场景相结合，具备扎实的理论基础和丰富的工程落地经验。熟悉主流预训练语言模型（如 BERT、ChatGLM、LLaMA 等），深入理解其原理及应用场景，并能够基于这些模型进行微调优化与部署上线。在大模型应用方面，熟练掌握 LoRA（Low-Rank Adaptation）等参数高效微调技术

我始终保持对新技术的敏感度与学习热情，紧跟行业动态，持续学习并实践 Agent 架构、多模型协同、Prompt Engineering、模型服务化等前沿方向。同时，在项目推进过程中，我注重团队协作与沟通效率，能够快速理解业务需求，并将其转化为可落地的算法解决方案。

在工作中，我始终坚持高标准、严要求，面对复杂问题时善于思考、乐于钻研，具备良好的抗压能力和执行力。未来我希望继续深耕 NLP 与大模型领域，推动人工智能技术在更多高价值场景中落地。