

王佳明

随时到岗

男 | 汉 | 邯郸

17769881449 | 17769881449@163.com

专业技能

编程和操作系统：

熟练的运用python,数据挖掘、数据分析、简单的机器学习任务,对linux有所了解；

Transformers，Stable Diffusion原理：

Transformers：深入理解Transformer架构，例如多头注意力机制、位置编码、归一化等核心概念。解析自注意力机制的数学原理等；

Stable Diffusion：熟悉SD中扩散模型核心原理，Unet架构理解，文本-图像关联原理，并熟悉风格微调等技术；

LLM模型运用技术与微调：

LLM模型：熟悉Qwen2.5，ChatGLM等开源大模型；

Lora：熟练掌握 lora 微调技术，通过量化低秩适应高效的微调大模型，同时采用双量化技术对常数进一步量化，降低内存提升计算效率。

计算算法：

计算机算法：熟悉掌握多种计算机算法，如KNN, 线性回归，SVM，KL散度等多种计算机算法；

数据库的应用：

Neo4j：熟练使用Neo4j图数据库的基本操作，了解复杂关系的数据处理；

GraphRAG：将图结构进入 RAG 框架中，实现结构化数据，在处理具有复杂关系的数据和需要进行推理的任务时表现出色；

大模型应用框架：

LangChain：集成多种大预言模型构建高级nlp应用，例如问答系统，文本生成和对话系统,能用langchain的模块。

VLLM：VLLM 是一个专注于大规模语言模型推理的高效框架，旨在提供高性能的推理服务，支持多种大模型，如 GPT、BERT、T5 等。

教育背景

河北工程大学科信学院 - 计算机科学与技术 - 本科

2017-09 ~ 2021-07

工作经历

北京科百宏业科技有限公司

2021-09 ~ 至今

- 负责实体识别、情感分析、文本分类、信息抽取等基础算法模型的开发和优化；
- 利用NLP技术解决实际工作中的问题，如用户查询的语义理解、知识图谱构建和文本挖掘等；
- 关注行业内的最新技术和研究成果，推动创新算法和技术在特定行业的应用；
- 确保项目的长期稳定性和适应性，定期评估、自动化数据更新、用户反馈整合、性能优化等。

项目经历

金融智能决策平台

2025-02 ~ 2025-05

需求背景：

项目基于千问2.5-72B大模型深度定制，连接自主搭建的金融级数据平台，实现高效，安全的金融数据交互与智能决策。通过行业定向优化技术，系统精准适配银行、证券、保险等场景需求，提供覆盖风控、投研、客服、合规的全链条智能服务。结合自主金融级数据平台，可实现数据交互与智能决策的深度融合。

技术栈：

Qwen、dify、vllm、xinference、pytorch , docker

项目流程：

- 1、**模型部署**：基于vLLM框架部署千问2.5 72B模型，并爆出，对接dify平台;
- 2、**平台集成**：在Dify中配置 workflow;
- 3、**性能优化**：部署Embedding和Rework模型，通过检索增强与结果润色，提升问答准确率与流畅度。
- 4、**安全可靠**：设置API访问控制与数据加密，配置容错机制和缓存策略，保障系统稳定运行。

语管家

2023-10 ~ 2024-06

需求背景：

语管家是一款创新型家居语音问答助手，依托于近年来快速发展的人工智能技术，能够精准理解并响应用户的语音指令，为用户提供智能化的家居问答服务。通过先进的自然语言处理和深度学习算法，语管家能够适应不同用户的表达习惯和需求，实现个性化的交互体验。无论是家居设备控制、生活信息查询，还是日常问题解答，语管家都能以高效、准确的方式为用户提供帮助，让智能家居生活更加便捷和舒适。

技术栈：

Whisper、LangChain、RAG , LLama1

项目流程：

- 1、**构建数据管道**：收集家居数据对数据进行聚合、清洗等操作；
- 2、**文本处理**：转化成文本输入使用 E5 进行语义嵌入生成并转化为高维语义向量存储至 并进行比较；
- 3、**RAG 集成**：将LangChain 进行集成检索 E5 生成的语义嵌入进行检索，通过 RAG 和 LLama1 的集成生成回答；
- 4、**语音集成与合成**：部署 Whisper 模型，并将生成的文本通过 TTS 转化为语音输出。

法律慧通

2023-02 ~ 2023-08

需求背景：

法律文本的复杂性、多样性和更新频率不断增加，传统法律检索工具已难以满足现代法律实践的需求。法律专业人士需要更高效、准确且智能化的工具来处理海量的法律文件、法规、判例和合同等，以提升工作效率、减少错误并提供更优质的法律服务。因此，我负责开发了一款基于传统RAG技术的法律智能系统，旨在通过自然语言处(NLP) 和机器学习技术，为法律领域提供高效、精准的解决方案；

技术栈：

BERT, GPT, Pandas , Numpy

项目流程：

- 1、**实体提取**：从非结构化文本中识别出关键实体；
- 2、**图谱构建**：利用LangChain和Neo4j将抽取的实体和关系转化为图结构，并存储到Neo4j图数据库中；
- 3、**向量索引与优化**：如果需要支持语义搜索，负责构建向量索引，并优化图谱的存储结构以提高查询效率；
- 4、**模型优化**：根据实际使用情况，持续优化NLP模型和知识图谱的构建流程，提高实体和关系提取的准确性；

智评助手

2022-09 ~ 2022-12

需求背景：

用户评价已成为企业洞察客户需求、优化产品和服务的核心资源。随着互联网的普及和社交媒体的快速发展，用户通过电商平台、社交媒体、在线论坛、客服系统等多渠道发表评价的行为日益频繁。这些评价不仅提供了用户对产品或服务的直接反馈，还蕴含了丰富的情感信息，例如满意、不满、惊喜、失望等。这些情感信息是企业了解用户真实需求、发现潜在问题、优化产品和服务的重要依据。然而，面对海量的用户评价数据，传统的分析方式效率低下，难以快速提取有价值的信息，也无法满足企业对实时性和精准性的要求。因此，企业迫切需要借助先进的技术手段，实现对用户评价情感倾向的自动化分析和分类，从而高效挖掘数据中的价值，为企业的决策提供有力支持；

技术栈：

TensorFlow , BERT , jieba , Docker

项目流程：

- 1、数据收集：**收集并整理海量法律文本数据，包括法律法规、判例、合同等，构建高质量的法律知识库。
- 2、数据预处理：**使用Python对法律文本进行清洗、分词、词性标注和句法分析，去除噪声数据并提取关键信息；
- 3、数据分块：**采用基于语义的分块方式（如按句号、段落分割）将法律文档分割成小块，便于后续处理；
- 4、模型优化：**使用LSH等技术对嵌入向量进行降维，提升检索效率；

智能翻译

2021-10 ~ 2022-04

项目背景：

信息社会中，跨语言交流的需求呈爆发式增长。无论是国际商务沟通、学术研究、新闻传播还是个人学习，高效且准确的翻译工具已成为不可或缺的工具。然而，传统的人工翻译方式存在诸多局限性，如耗时费力、成本高昂，且难以满足市场对快速响应的需求。这使得开发一款智能化的翻译助手系统变得尤为迫切。智能化翻译助手系统不仅能够显著提升翻译效率，还能通过先进的技术手段保证翻译质量。其核心价值在于打破语言障碍，促进国际间的文化交流和商业合作。通过引入自然语言处理（NLP）和机器学习技术，该系统能够实现多语言之间的即时翻译，同时提供高质量的翻译结果，满足不同场景下的多样化需求；

技术栈：

Transformers、MongoDB、RAG、NumPy、NumPy;

项目流程：

- 1、数据预处理：**清洗数据，去除无关信息，如广告、注释等，统一文本格式，构建高质量的平行语料库；
- 2、模型训练：**训练模型进行多种语言对的翻译任务。通过交叉验证等方法评估模型性能，持续优化模型参数，提高翻译质量和流畅度；
- 3、结果评估：**对系统输出的翻译结果进行人工校验，确保准确性和流畅度；
- 4、模型测试：**对系统进行多语言测试，确保翻译准确性和响应速度。

自我评价

我是一个认真负责且富有责任心的人，对待工作和学习都一丝不苟，注重细节，力求做到最好。性格开朗乐观，善于与人沟通交流，能够积极融入团队并发挥自己的作用。同时，我也具备较强的适应能力和学习能力，能够快速接受新知识、新事物，不断自我提升，努力朝着目标前进。