

徐丹颖

dx2143@nyu.edu | (+86) 135-2297-1325

[LinkedIn](#) | [Github](#) | [个人主页](#)



教育背景

纽约大学

纽约, 美国

电子与计算机工程学院 | 计算机工程专业 | GPA: 3.9/4.0 | 英语技能扎实 (托福109, GRE 320+3.0)

2023.09-2025.05

东南大学

南京, 中国

计算机科学与工程学院、软件学院与人工智能学院 | 人工智能专业

2019.09-2023.06

专业技能

- 语言: Python (Scikit-learn/Pytorch/Tensorflow/OpenAI), C++/C, Linux, SQL, Cypher, Java, Latex
- 工具: Hugging Face, Azure, Heroku, Dockers, MySQL, PostgreSQL, Git, Github, Gitee, Apache Spark/Hadoop, Hive
- 方法: 机器学习、深度学习、NLP、CV、LLM、数据库管理 (SQL/NoSQL)、概率统计、数据分析、预测建模、论文阅读写作

实习经历

Global AI

纽约, 美国

机器学习工程师

2024.05-2024.08

- 基于LangChain框架和Streamlit创建了在线新闻聊天机器人。
- 使用Git在Heroku上为了持续集成部署了该机器人。
- 利用Bucketeer提供的AWS S3兼容存储高效处理数据输入, 降低存储成本30%。
- 创建了一种存储优化方法, 支持自动更新以提高网页响应速度, 将网页后续处理时间减少95%。

Global AI

纽约, 美国

机器学习工程师

2024.01-2024.04

- 在30万份GDELT新闻数据上建立了PostgreSQL数据库, 提高时间效率20%。
- 通过并行处理自动升级数据库, 整体效率提升15%。
- 分析并可视化了100万条MSCI US指数时间序列股票数据, 识别趋势和异常。
- 开发了LSTM模型, MSE得分为0.4, 相比基线模型, 使股票预测效果提升19.7%。

华为南京研究所

南京, 中国

数据通信部门, 工具开发工程师

2022.08-2022.09

- 通过检查具有数千个C/C++函数的143个静态路径图来进行灰盒测试。
- 采用FUZZ测试技术对Linux下的872个代码文件进行白盒测试, 预计可提升产品性能30%。

项目经历

AI生成文本判别

2024.04-2024.05

- 使用Apache Spark处理和分析了50,000多条来自于人类和AI生成的句子。
- 研究了最先进的文本检测模型, 并训练了3个基线模型, 平均准确率为86%。
- 对BERT模型进行训练微调作为轻量语言模型代表, 准确率为93.2%。
- 在Google Colab和Hugging Face上使用了LlaMa2模型, 预测准确率为65.4%。
- 通过Azure OpenAI和Azure Notebook部署了ChatGPT3.5模型, 准确率为80.1%。

中文性别去偏重写器

2022.12-2023.06

- 设计了一个NLP框架, 通过模式转换、神经翻译和数据聚合来减少数据中的性别偏见。
- 在PyTorch上对14多万个句子训练Seq2Seq模型 (字符/单词级别) 和Seq2Seq注意力模型 (字符级别)。
- 设计了词嵌入关联测试 (WEAT) 并在CBOW模型上对其进行了中文评估, 将性别偏见减少了45.4%。
- 使用wvm-RoBERTa模型执行了共指解析下游任务, 在减少性别偏见后, 性能始终保持在92%左右。
- 在TextCNN模型上进行了情感分析, 在减少性别偏见后, 性能始终保持在80%。

基于深度神经网络的脑科学研究

2020.11-2022.05

- 使用Python(OpenCV)从1297个受过训练的猴子玩吃豆人游戏的视频中提取了120万张图像。
- 基于Python(TensorFlow)设计了一个结合AlexNet和LSTM的ConvRNN模型来预测玩家的下一步行动, 准确率达到84.6%。
- 执行了类别激活图 (CAM) 热分析, 用于激活层可视化。
- 调整Grad-Cam热分析模型以解释神经网络每层的决策机制, 并进行可视化。

获奖情况

- 2022年美国大学生数学建模比赛Finalist (特等奖入围)
- 第十七届全国大学生挑战杯揭榜挂帅专项赛特等奖