YUELYU JI

yueluji@gmail.com • 4126260932 • https://www.linkedin.com/in/yuelyuji/

Anticipated graduation dates: 04/13/2027

EDUCATION

University of Pittsburgh, Pittsburgh, USA
Ph.D. *2023 -2027*

University of Pittsburgh, Pittsburgh, USA
Master in Information Science December 2022

Nanjing Agricultural University (NJAU), Nanjing, China
Bachelor in Information System June 2020

RESEARCH EXPERIENCE

University of Pittsburgh Advised by Dr. Daqing He, Pittsburgh, USA
Graduate Student Researcher August 2023 - Present

- **Reranking for Document Retrieval**: Spearheaded the development of *ReasonRank*, a reranking framework that enhances interpretability by generating explicit and comparative reasoning for document rankings. Fine-tuned using ***LLaMA3.1 7B*, and *Mistral*** models, the *ReasonRank* framework improved NDCG@5 scores by **5%**, outperforming state-of-the-art models like **GPT-4** and achieving **73.37%** on the DL19 dataset. This work optimized retrieval results for large-scale search and compressed the model-sized information retrieval systems.
- **Bias Detection and Mitigation in Clinical Trial Matching**: Co-authored *BiasGuard*, a framework designed to reduce bias in clinical trial matching and medical question-answering tasks. Using *contrastive learning* and fine-tuning on ***LLaMA 3.1 7B* and *Mistral*** models, we reduced bias scores from **0.215 to 0.192** on patient-trial matching tasks and achieved similar improvements in error rates for MedQA tasks, ensuring fairness across sensitive demographic attributes.
- **Biomedical Text Summarization**: Developed the *RAG-RLRC-LaySum* framework for the *BioLaySumm* competition, fine-tuning the *Longformer Encoder-Decoder (LED)*, *Flan-T5*, and *Mistral* models to combine *Retrieval-Augmented Generation (RAG)* and *Reinforcement Learning for Readability Control (RLRC)*. The model improved readability by **20%** and factual accuracy by **10%** on the PLOS and eLife datasets, demonstrating significant advancements in generating accurate and readable layman summaries for complex biomedical research.

WORK EXPERIENCE

NetEase Cloud Music, Hangzhou, China
Machine Learning Intern March 2023 - August 2023

- **Developed CPCTG model**: Designed a controlled text generation model using LLMs, boosting user satisfaction from 4.3 to 4.6.**Improved accuracy with reinforcement learning**: Increased text generation Rouge-L score from 43% to 60% through reinforcement learning.**Cross-team collaboration**: Worked with engineers and product teams to integrate and refine the model for optimal performance.

PUBLICATIONS

- **Yuelyu Ji,** Zhuochun Li, Rui Meng, Daqing He ReasonTeach: Teaching Student Models to Rank through Reasoning-Based Knowledge Distillation. Arxiv 2024 (targeting NAACL2025)
- **Yuelyu Ji,** Zhuochun Li, Sonish Sivarajkumar, Hang Zhang, Yanshan Wang BiasGuard: A Unified Framework for Addressing Fairness in Large Language Models for Clinical Trial Matching and Medical Question Answering. Arxiv 2024 (targeting NAACL 2025)
- **Yuelyu Ji,** Zeshui Yu, Yanshan Wang, Assertion Detection Large Language Model In-context Learning LoRA Fine-tuning, Journal of the American Medical Informatics Association. (ICHI) 2024