MaoLin He

+86 18873266361   marine2231h@gmail.com
https://github.com/rinmy23

Educations

The University of Melbourne `QS Top100`

Master of Information Technology (with Distinction)
- Related Subjects: Research Project (88), Research Methods (82), Natural Language Processing (85), Advanced Studies in Computing (81), Statistical Machine Learning (77), Declarative Programming (85), Distributed Systems (85)
- GPA: 3.89/4.0 First Class Honours
- Supervisor: A/Prof Brian Chapman

Beijing Normal University `985` `211` `Double 1st-Class`

Bachelor of Science (Computer Science and Technology)
- Average Score of professional courses: 82/100
- Bronze Medal of China Collegiate Programming Contest (Twice)

Research Interest

My primary research interests are centered around natural language processing. My research is motivated by the ultimate goal of enhancing information access for individuals entering a new field. Using, understanding, and improving retrieval and large language models (LLMs) can provide us with powerful tools to enhance our research.

Research Experiences

Developing Optimal Grounding Resources for Consumer QA Systems   Feb 2024 - Aug 2024

Research paper of Advanced Studies in Computing (COMP90005)   Advisor: A/Prof Brian Chapman
- Information Retrieval Optimization: Developed an innovative **Hybrid Semantic Real-time Document Retrieval (HSRT)** method, pioneering the integration of term-based real-time search (by leveraging Boolean search of PubMed search API) with semantic similarity-based search for NCBI resources (by MedCPT retriever), including both PubMed and PMC.
- Chunk Optimization: Introduced **Semantic Enhanced Overlap Segmentation (SEOS)**, a novel text segmentation approach that combines sentence semantics with embedding model influences while leveraging chunk overlap for enriched context.
- This study, by combining the two optimizations with LlamaIndex, improved the response accuracy of the Claude-3-haiku model by **5%**. This work is now submitted to **the Journal of Biomedical Informatics (Impact Factor: 4.0)**.

Master Thesis: Creating a Cancer Patient-facing QA System using LLMs   Feb 2024 - Aug 2024

Student Research Assistant   Advisor: A/Prof Brian Chapman
- Developed the **Single-Step Comprehensive Retrieval with Evidence-Processing (SCREP)** method, integrating works of COMP90005. Central to SCREP is the **Hierarchical Evidence-Processing (Hier-EP)** which uses multi-factor filtering and adaptive granularity control to extract high-quality evidence, improving response accuracy by **8%** in the Claude-3-haiku model. This work is now submitted to **the IEEE Journal of Biomedical and Health Informatics (Impact Factor: 7.700).**
- Integrated source tracking, citations, and readability adjustments to enhance output understandability and verification.
- Proposed the **Multi-Criteria Adaptive RAG (MCA-RAG)**, applying medical QA criteria for optimal retrieval approaches (no retrieval, SCREP, Multi-Step Retrieval). A key component, the **Dual-LLM Collaborative ReAct-based Multi-Step Retrieval (DCRMR)**, combines LLaMA-3-8B for stepwise responses with LLaMA-3-70B for high-level reasoning. MCA-RAG enhanced **LLaMA-3-8B's accuracy to match GPT-4**.

Selected Assessment Projects

Automated Fact Checking For Climate Science Claims   Feb 2023 - Jun 2023

Individual research project of Natural Language Processing (COMP90042)
- Developed information retrieval(IR) function using Dense Passage Retrieval (DPR) to search for the most related evidence for the given claim in a dataset of over 1208k+ records
- Analysed data and results, pondered causes and investigated relevant papers; Adjusted hyperparameters and applied novel methods (Error-Prone Negative Sampling and Adaptive Retrieval top-n), improving IR model's F1 score from 0.12 to 0.16
- Implemented a classification function for claim verification using 4 pre-trained models (BERT, ALBERT, Roberta, Distilroberta) and an ensemble approach with the highest accuracy of 0.56

Reversi Autonomous Agent   Jul 2022 - Nov 2022

Group research work of AI Planning for Autonomy (COMP90054)
- Implemented 3 different algorithms (BFS, MCT and alpha-beta); Improvements are made to each of these three algorithms based on the rules of the game
- Collaborated with team-mates on developing a Wiki, documenting and describing solutions; Received 31.25/35 for the assignment

Services

Semester 2, 2023 Orientation Volunteering tour guide   Jul 2023 - Aug 2023

Showed 25 commencing students around campus during orientation week; Introduced the layout of University and Faculty facilities

Additional Informations

- **ProgrammingLanguages:** C/C++,Java,Python, R
- **DeepLearningFramework:** PyTorch,Keras

- **Tools:** LATEX, Llamaindex
- **Languages**: Chinese (native), English (PTE: 68)