Chia-Wei, Wu

☎ +886+9+82162688 | GitHub | ✉ chia.wei.nlp@gmail.com

NLP Engineer, Prompt Engineer

EDUCATION

National Chengchi University Sep 2022—July 2024

- Master of Computer Science. Research interests: Natural Language Processing, Prompt Engineering, Information Retrieval, Text Classification and Sentiment Analysis.
- Thesis: Unsupervised Multi-granularity LLM-based Reranking for Long Text Retrieval. *advised by Prof. Tsai-Yen Li and Prof. Hen-Hsen Huang.*
- NCCU Floodfire Project Research Team Member, *advised by Prof. Yu-Chung Cheng.*

National Chiayi University Sep 2019—June 2022

- Bachelor of Applied Mathematics. Key competencies: Linear Algebra, Discrete Mathematics, Linear Programming and Operational Research.
- Thesis: Using the least squares method to investigate the growth of honey crystallization. *advised by Prof. Jia-Wen Chen.*

WORK EXPERIENCE

AI Research Engineer Sep 2024—Dec 2024

National Center for High-Performance Computing *Taipei City, Taiwan*

- Developed a contextual retrieval-based retrieval-augmented generation model to address semantic mismatches between queries and documents.
- Used Selenium to automate web scraping through simulated user interactions.
- Successfully converted text into nearly 10 million high-quality OCR data, enhancing the model's ability to recognize Traditional Chinese in real-world scenarios.

Graduate Research Assistant July 2023—July 2024

Institute of Information Science, Academia Sinica *Taipei City, Taiwan*

- Developed a hybrid text reranking algorithm that overcomes the limitations of pointwise and listwise, enabling relative position adjustments and reducing the time complexity of pairwise.
- Developed text compression techniques to overcome length limitations in unsupervised text reranking tasks, preserving semantic integrity for large language model.
- Designed the term stack approach for query rewriting, effectively mitigating AI hallucination caused by an overabundance of key phrases.

Graduate Research Assistant Sep 2022—Aug 2024

Integrative Strategic Value Management System *Taipei City, Taiwan*

- Assist the IT product development department in drafting product operation manuals.
- Administrative Support: Word processing, event coordination assistance for large-scale events.

SKILLS

Software: Python, R, Git, Maple, Latex and Excel.

Language: Chinese, English and Taiwanese.

PROJECTS

Explore the Emotional Intentions of Social Media Users towards ChatGPT [Paper]

- Defined thresholds by post volume to identify active PTT boards, and used the Scrapy to automate data collection (Jan-Apr 2023) on public sentiment towards ChatGPT.
- Utilized Jieba for Chinese text segmentation and using TF-IDF algorithm to identify key information, ranking the most relevant information based on its significance within the text.

OTTO - Multi-Objective Recommender System competition, 576th [Kaggle]

- Developed a session-based recommendation system by transforming user consumption behavior with Word2Vec and integrating the ANNOY algorithm to reduce time complexity.
- Used the Polars library and ensemble learning algorithms to consolidate the results of multiple models (Co-Visitation and Word2Vec).

Multiple Data Science Projects in the R Programming Environment

- Developed an interactive visualization system for Principal components analysis and correspondence analysis in Shiny, allowing users to upload specific datasets. [Demo]
- Implemented K-Fold Cross-Validation to enhance the prediction stability for the protein subcellular localization problem.
- Predicted Santander customer satisfaction using feature engineering and classification models (XGBoost).

AWARDS AND CERTIFICATIONS

- Academic Research Award, National Chengchi University, 2024
- Invitee, Chulalongkorn University, Thailand, 2024
- Banking Apps Reviews Classification competition, 9th [Kaggle]
- TOEIC Green Certificate, 2024
- SQL for Data Science, Coursera, 2022

Autobiography

My name is Chia-Wei Wu, and I graduated with a degree in Computer Science from National Chengchi University, where I specialized in Natural Language Processing, including generative AI, information retrieval, and sentiment analysis. Currently, I work at the National Center for High-Performance Computing, where I focus on developing retrieval-augmented generation models to address public concerns related to government petitions. Throughout my academic journey, I maintained an active and inquisitive mindset, participating in various competitions and development projects that provided practical insights into real-world challenges. To expand my knowledge further, I worked as the graduate research assistant at both the Integrative Strategic Value Management System and the Institute of Information Science at Academia Sinica. Specifically, I designed user manuals for database systems and explored ways to enhance existing NLP techniques through prompt engineering, aiming to improve the performance of downstream tasks with large language models. Guided by my background in mathematics, I adhere to the principle of "Bold Assumptions, Cautious Verification," approaching challenges with logical precision and creative innovation.