Yen-Hsiang Wang

+886 928-666-218 heliart_sho@hotmail.com https://heliart.me/ github.com/kenny75557

Research Interest

Natural Language Processing

Education

National Chung Hsing University **Sep. 2020 – Sep. 2022**
M.S. in Computer Science. Advisor: Yao-Chung, Fan *Taichung City, Taiwan*

Fu Jen Catholic University **Sep. 2016 – Jun. 2020**
B.S. Computer Science and Information Engineering. *New Taipei City, Taiwan*

Publications

Conference/Journal Papers

(† indicates equal contribution.)

1 **Yen-Hsiang Wang**, Feng-Dian Su, Tzu-Yu Yeh, Yao-Chung Fan. A Cross-Lingual Statutory Article Retrieval Dataset for Taiwan Legal Studies *In ROCLING 2024.*

2 You-Cheng Liao, Chen-Jui Yu, Chi-Yi Lin, He-Feng Yun, **Yen-Hsiang Wang**, Hsiao-Min Li, Yao-Chung Fan. Learning-From-Mistakes Prompting for Indigenous Language Translation *In ACL LoResMT 2024.*

3 Xingdi Yuan, Tong Wang, **Yen-Hsiang Wang**†, Emery Fine, Rania Abdelghani, Pauline Lucas, Hélène Sauzéon, Pierre-Yves Oudeyer. Selecting better samples from pre-trained LLMs: a case study on question generation *In Findings of ACL 2023.*

4 Rania Abdelghani, **Yen-Hsiang Wang**, Xingdi Yuan, Tong Wang, Pauline Lucas, Hélène Sauzéon, Pierre-Yves Oudeyer. GPT-3-driven pedagogical agents to train children's curious question-asking skills *In International Journal of Artificial Intelligence in Education.*

Tutorial

- ROCLING 2024 "Mastering Retrieval-Augmented Generation: Unlocking Advanced Retrieval Techniques," [Website and Slide]

Work Experience

National Chung Hsing University Natural Language Processing Lab **Jan. 2024 – Recent**
Research Assistant *Taichung, Taiwan*

- Researching on extremely low-resource machine translation, exploring how LLMs can enhance translation for Indigenous languages by leveraging Chain-of-Thought reasoning by giving relevant pattern by sentence retrieval. [LoResMT ACL'24]
- Proposed a universal method for generating a synthetic legal QA dataset and validating the performance of Sparse embedding, Dense embedding, and LLM query expansion in legal domain retrieval. [Github link]
- Building a retrieval augmentation generation application for agricultural domain question-answering, focusing on retrieval and re-ranking methods to enhance accuracy. [Shennong-TAIDE].

EmotiBot **Jun. 2023 – Dec. 2023**
Machine Learning Engineer *Taipei, Taiwan*

- Responsible for the AI applications development of intelligent customer support utilizing Large Language Models (LLM) and Retrieval-Augmented Generation (RAG) architecture.

Inria, National Institute for Research in Digital Science and Technology **Mar. 2022 – Sep. 2022**
Research Engineer Intern *France*

- Exploring the Educational application of large language models (LLMs) to automate the generation of diverse content to foster curiosity-driven questions among elementary students, reducing the need for manual intervention. [IJAIED]
- Researching on Question Generation Sampling and Evaluation of large language models. [ACL'23]

Industrial Technology Research Institute **Jul. 2021 – Aug. 2021**
Software Engineer Intern *Hsinchu, Taiwan*

- Training Transformer-based model for Legal Judgement Question-Answering.
- Development of Manual Annotation System for Data Labeling.

Awards

Best Paper Nominee | *ACLCLP* **Nov. 2024**

- Construct a Cross-Lingual Retrieval Dataset in the Legal domain and provide multiple baseline methods.