

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

M.S. in Computer Science. Advisor: Yao-Chung, Fan

Sep. 2020 – Sep. 2022

Taichung City, Taiwan

Fu Jen Catholic University

B.S. Computer Science and Information Engineering.

Sep. 2016 – Jun. 2020

New Taipei City, Taiwan

Publications († 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*.

Experience

National Chung Hsing University Natural Language Processing Lab

Research Assistant

Jan. 2024 – Recent

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 sentences by retrieval. [\[LoResMT ACL'24\]](#)
- Proposed a 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

Machine Learning Engineer

Jun. 2023 – Dec. 2023

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

Research Engineer Intern

Mar. 2022 – Sep. 2022

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

Software Engineer Intern

Jul. 2021 – Aug. 2021

Hsinchu, Taiwan

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

Selected Project

PTT Search Engine | *Vue.js, Flask, Elasticsearch, Docker*

Oct. 2020 – Jan. 2021

- Built a search engine including all articles and commands crawled from popular boards on PTT(web forum).
- Implemented BM25 (Elastic Search) for retrieving comments from the database and processed the data using Object-Oriented Programming techniques.
- [\[Github Link\]](#)