Yu-Ting Lee

📞 +886+9+06782385 | ✉ yutinglee.nlp@gmail.com | in LinkedIn | ⓞ GitHub | 🌐 Website

NLP Engineer with 5 years of experience, encompassing with 6 years of machine learning expertise.

Skills

Languages: Python, Mojo, Julia, SQL, R, C/C++, Java, MATLAB

Software & Tools: PyTorch, PyTorch Lightning, Jupyter, GCP, GNU/Linux, Pandas, Scikit-Learn, NLTK, Hugging Face, Gradio, Streamlit, LangChain, Git, Docker, FastAPI

NLP Techniques: RAG, LLMOps, Fine-tune LLMs, Text Generation, Multi-Document Summarization, Recommendation System, Text Classification, Named Entity Recognition

Work Experience

Research Assistant — Oct 2023 – Present
Institute of Information Science, Academia Sinica — *Taipei City, Taiwan*

- Developing a system for Research Methods Recommendation.
- Applying algorithms to time-series contents generation settings and further evaluate with state-of-the-art text generation models.

Graduate Research Assistant — Oct 2021 – Oct 2023
Institute of Information Science, Academia Sinica — *Taipei City, Taiwan*

- Developed a state-of-the-art Knowledge Updating System in Full-Text, enabled collaborated content generation with GPT-4, performs 30% better than existed models.
- Fine-tuned LLaMA based models (Alpaca, Vicuna) with LoRA and Adapter.
- Algorithm development for knowledge updating utilizing state-of-the-art NLP techniques.

Graduate Research Assistant — Sep 2021 – Sep 2023
Department of Journalism, National Chengchi University — *Taipei City, Taiwan*

- Collaborated and developed knowledge updating system with expertise of journalism.
- Articulated research ideas to technical/non-technical audience.

Education

National Chengchi University — Taipei City, Taiwan
Department of Computer Science, Master's Degree — *Sep 2021 – Oct 2023*

- Advised by Tsai-Yen Li and Hen-Hsen Huang.
- Essay: Knowledge Update in Full Text Triggered by a News Event. Developed LLM for knowledge updating, outperformed state-of-the-art LLMs (e.g. GPT-4, LLaMA, Vicuna, etc.).
- Relevant coursework: Developed Wordle-based game with optimization method.

Chung Shan Medical University — Taichung City, Taiwan
Department of Medical Informatics, Bachelor's Degree — *Sep 2017 – June 2021*

- Honor: Supported by the Ministry of Science and Technology Research Fellowship for NLP research.

Publications

A Multi-grained Dataset for News Event Triggered Knowledge Update

Yu-Ting Lee, Ying-Jhe Tang, Yu-Chung Cheng, Pai-Lin Chen, Tsai-Yen Li and Hen-Hsen Huang. 2022. *In Proceedings of the 31st ACM International Conference on Information and Knowledge Management (CIKM 2022).* (Oral Presentation)

A Sentiment Analysis Method based on a Transformer and a Recurrent Neural Network

Yu-Ting Lee, Yen-Ching Chang. 2020. *National Taiwan University Management Reviews:* 1st *Management and Medical Sciences Interdisciplinary Conference.* (Poster Presentation)

Academic Services & Invited Talks

Reviewer, International Committee on Computational Linguistics (**LREC-COLING**), **2024**
Invited Talk, Dense Information Retrieval with LLMs, Chung Shan Medical University, **March 2024**

Teaching and Outreach

Founder – PyTorch Taiwan, Facebook Public Group   Jan 2024
- Promoting implementations and SOTA models with PyTorch and PyTorch-based framework.

Contribute to Open-Source Project – Lit-LLaMA, Lightning AI Inc.   June 2023
- [Lit-LLaMA](): Independent implementation of LLaMA pretraining, finetuning, and inference code that is fully open source. Fixed partial functions for fine-tuning LLaMA.

TA – Natural Language Processing — Dept. of CS, NCCU   Feb 2023 – June 2023
Supervised by Hen-Hsen Huang   *Taipei City, Taiwan*
- Designed concepts explanations for coursework, assisted with homework and projects to support 70 students.

TA – Practical NLP — Taiwan AI Academy   Aug 2022 – Sep 2022
Supervised by Hen-Hsen Huang   *New Taipei City, Taiwan*
- Designed introduction lectures and taught a two-week NLP project to support 20 intermediate practitioners.
- Taught Project: [Summary Generation Tool for News during Important Time Durations]().

Projects

WebChatRAG
Input a website URL, and start chatting with language model. [[GitHub]]()

Dynamic README Generation
Automatically update READMEs with each commit, highlighting new features and changes, saving time and enhancing documentation accuracy and clarity. [[GitHub]]()

Automated Generation and Evaluation of Research Methods in NLP
Developing an innovative system designed to automatically generate research methodologies based on specified objectives and a review of related work in the field.   [[GitHub]]()[[Demo]]()

News Event Triggered Article Updating System
Developed a system utilizing advanced LLMs, focusing on full-text knowledge updating in response to real-time news events, enhancing article accuracy. [[GitHub]]()[[Demo]]()

Summary Generation Tool for News during Important Time Durations (SGIT)
Developed NLP tool for summarizing COVID-19 social media content, countering misinformation, and improving public response through objective news analysis. [[GitHub]]()

Self-learning Certificates

Machine Learning Specialization – DeepLearning.AI	Dec 2022
Natural Language Processing Specialization – DeepLearning.AI	July 2021
Fine Tune BERT for Text Classification with TensorFlow – Coursera	July 2021