Yihuai Hong

Email: yihuaihong@gmail.com Phone: +86 15815289871 Homepage: https://yihuaihong.github.io

Education

Bachelor of Engineering, Computer Science Sept 2020 –June 2024
South China University of Technology(**SCUT**), Guangdong, China
GPA: 3.59/4.00 (Top **5**)
Relevant Courses: Algorithm design and analysis(94), Probability & Mathematical Statistics(94), Discrete Mathematics(95), Data Structure(92), Database System(90), Java Programming(93), Advanced Language Program Design(97), Advanced Topics of Information Technology(98)

Publications

Intrinsic Evaluation of Unlearning Using Parametric Knowledge Traces June 2024
Yihuai Hong, Lei Yu, Shauli Ravfogel, Haiqin Yang, Mor Geva
Under review for ICLR 2024

Dissecting Fine-Tuning Unlearning in Large Language Models June 2024
Yihuai Hong, Yuelin Zou, Lijie Hu, Ziqian Zeng, Di Wang, Hainqin Yang
EMNLP 2024 Main

Interpretability-based Tailored Knowledge Editing in Transformers June 2024
Yihuai Hong, Aldo Lipani
EMNLP 2024 Main

ConsistentEE: A Consistent and Hardness-Guided Early Exiting Method for Accelerating Language Models Inference Aug 2023
Ziqian Zeng*, **Yihuai Hong***, HongLiang Dai, Huiping Zhuang, Cen Chen
AAAI 2024, Main Track**.**

Research Experience

Research Intern, **University of Toronto** Sept 2024 - Present
Supervisor: *Prof. Zhijin Jin **(Rising Star in ML, DS and EECS)***
- Mechanistic understanding of LLM Reasoning and Memorization

Research Intern, **Alibaba DAMO Academy** Aug 2024 - Present
Supervisor: *Dr. Lidong Bing and Dr. Wenxuan Zhang*
- Effective parametric Knowledge Erasing in LLM

Research Intern, **Tel Aviv University** Feb 2024 – Aug 2024
Supervisor: *Prof. Mor Geva Pipek from **Google Research***
- Internal Evaluation of LLM Unlearning and Mechanistic understanding of Jailbreaking

Research Intern, **China International Digital Economy Academy** Dec 2023 - Feb 2024
Supervisor: *Dr. Haiqin Yang*
- Dissecting the drawbacks of Finetuning-based Unlearning in LLM

Research Intern, **University College London** June 2023 - Dec 2023
Supervisor: *Prof. Aldo Lipani*
- Targeted Knowledge Editing in LLM mitigating Over-editing issues

Research Intern, **South China University of Technology** June 2022 - Aug 2023
Supervisor: *Prof. Ziqian Zeng*
- Efficient Inference in LLM through Early-exit and RL algorithm

Honors and Awards

AAAI 2024 Student Scholarship Dec 2023

Top Ten Excellent Students Nomination Award of South China University of Technology Nov 2023

China National Scholarship Sept 2023

Meritorious Winner of The Mathematical Contest in Modeling (MCM) May 2023

Kaggle **Silver** medal (Top 5%) July 2021

- CommonLit Readability Prize：Rate the complexity of literary passages for grades 3-12 classroom use Kaggle

Kaggle Bronze medal (Top 6%) Mar 2022

- Evaluating Student Writing: Analyze argumentative writing elements from students grades 6-12

SCUT First Prize Scholarship Oct 2022

Patent

Self-supervised pre-training method, system and medium for Chinese Pinyin spelling correction. Sept 2022

IP No: 202211156374.3

Academic Services

Program Committee: ICLR (2025), AAAI (2024), ACL ARR (Feb. 2024 - June 2024)

SKILLS

Programming Languages: Python, C++, C, Matlab, Java, Latex
Framework & Tools: PyTorch, TensorFlow, Django
Languages: English, Chinese (Native)