

Aiwei Liu

☎ +86 18851830977 | @ liuaw20@mails.tsinghua.edu.cn | 🌐 exlaw.github.io | 📍 Tsinghua University, Beijing, China

I am a fourth-year Ph.D. student. My research interests lie in the fields of natural language processing and deep learning. In particular, I am interested in building more robust and secure natural language generation algorithms and focus on:

- The robustness of converting a database schema related question into a SQL query.
- Watermarking for Large Language Models.
- Safety alignment of Large Language Models.

EDUCATION

Tsinghua University

Ph.D. in Software Engineering; Advisor: Associate professor Lijie Wen

Beijing, China

Sep 2020 – present

Nanjing University

B.Eng. in Software Engineering; GPA: 4.6/5.0, ranking: 5/220

Nanjing, China

Sep 2016 – Jun 2020

PUBLICATIONS

Direct Large Language Model Alignment Through Self-Rewarding Contrastive Prompt Distillation

- **Aiwei Liu**, Haoping Bai, Zhiyun Lu, Xiang Kong, Simon Wang, Jiulong Shan, Meng Cao, Lijie Wen
- In Proceedings of Association for Computational Linguistics (ACL), 2024

Semantic Enhanced Text-to-SQL Parsing via Iteratively Learning Schema Linking Graph

- **Aiwei Liu**, Xuming Hu, Li Lin, Lijie Wen
- In Proceedings of the 28th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD), 2022

A Semantic invariant Robust Watermark for Large Language Models

- **Aiwei Liu**, Leyi Pan, Xuming Hu, Shiao Meng, Lijie Wen
- In Proceedings of the International Conference on Learning Representations (ICLR) 2024

An Unforgeable Publicly Verifiable Watermark for Large Language Models

- **Aiwei Liu**, Leyi Pan, Xuming Hu, Shu'ang Li, Lijie Wen, Irwin King, Philip S. Yu
- In Proceedings of the International Conference on Learning Representations (ICLR) 2024

Character-level White-Box Adversarial Attacks against Transformers via Attachable Subwords Substitution

- **Aiwei Liu**, Honghai Yu, Xuming Hu, Shu'ang Li, Li Lin, Fukun Ma, Yawen Yang, Lijie Wen
- In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2022

Preventing and Detecting Misinformation Generated by Large Language Models

- **Aiwei Liu**, Qiang Sheng, Xuming Hu
- In Proceedings of the 47th International ACM SIGIR Conference (SIGIR), 2024

Exploring the Compositional Generalization in Context Dependent Text-to-SQL Parsing

- **Aiwei Liu**, Wei Liu, Xuming Hu, Shuang Li, Fukun Ma, Yawen Yang, Lijie Wen
- In The Findings of Association for Computational Linguistics (Findings of ACL), 2023

GDA: Generative Data Augmentation Techniques for Relation Extraction Tasks

- Xuming Hu*, **Aiwei Liu***, Zeqi Tan, Xin Zhang, Chenwei Zhang, Irwin King, Philip S. Yu
- In The Findings of Association for Computational Linguistics (Findings of ACL), 2023

An Entropy-based Text Watermarking Detection Method

- Yijian Lu, **Aiwei Liu**, Dianzhi Yu, Jingjing Li, Irwin King
- In Proceedings of Association for Computational Linguistics (ACL), 2024

On the Cross-lingual Consistency of Text Watermark for Large Language Models

- Zhiwei He, Binglin Zhou, Hongkun Hao, **Aiwei Liu**, Xing Wang, Zhaopeng Tu, Zhuosheng Zhang, Rui Wang
- In Proceedings of Association for Computational Linguistics (ACL), 2024

On the Cross-lingual Consistency of Text Watermark for Large Language Models

- Zhiwei He, Binglin Zhou, Hongkun Hao, **Aiwei Liu**, Xing Wang, Zhaopeng Tu, Zhuosheng Zhang, Rui Wang
- In Proceedings of Association for Computational Linguistics (ACL), 2024

RAPL: A Relation-Aware Prototype Learning Approach for Few-Shot Document-Level Relation Extraction

- Shiao Meng, Xuming Hu, **Aiwei Liu**, Shu'ang Li, Fukun Ma, Yawen Yang, Lijie Wen
- In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2023

AMR-based network for aspect-based sentiment analysis

- Fukun Ma, Xuming Hu, **Aiwei Liu**, Yawen Yang, Philip S. Yu, Lijie Wen
- In Proceedings of Association for Computational Linguistics (ACL), 2023

CHEF: A Pilot Chinese Dataset for Evidence-Based Fact-Checking

- Xuming Hu, Zhijiang Guo, Guanyu Wu, **Aiwei Liu**, Lijie Wen, Philip S. Yu
- In The 2022 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2022

A Multi-level Supervised Contrastive Learning Framework for Low-Resource Natural Language Inference

- Shu'ang Li, Xuming Hu, Li Lin, **Aiwei Liu**, Lijie Wen, Philip S. Yu
- IEEE/ACM Transactions on Audio, Speech, and Language Processing

Improving Open Relation Extraction With Search Documents Under Self-Supervisions

- Xuming Hu, Zhaochen Hong, Chenwei Zhang, **Aiwei Liu**, Shiao Meng, Lijie Wen, Irwin King, Philip S. Yu
- IEEE/ACM Transactions on Audio, Speech, and Language Processing

Entity-to-Text based Data Augmentation for various Named Entity Recognition Tasks

- Xuming Hu, Yong Jiang, **Aiwei Liu**, Zhongqiang Huang, Pengjun Xie, Fei Huang, Lijie Wen, Philip S. Yu
- In The Findings of Association for Computational Linguistics (Findings of ACL), 2023

Enhancing Cross-lingual Natural Language Inference by Soft Prompting with Multilingual Verbalizer

- Shuang Li, Xuming Hu, **Aiwei Liu**, Yawen Yang, Fukun Ma, Philip S. Yu, Lijie Wen
- In The Findings of Association for Computational Linguistics (Findings of ACL), 2023

Gaussian prior reinforcement learning for nested named entity recognition

- Yawen Yang, Xuming Hu, Fukun Ma, Shuang Li, **Aiwei Liu**, Lijie Wen, Philip S. Yu
- IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)

PREPRINTS

A comprehensive evaluation of ChatGPT's zero-shot Text-to-SQL capability

- **Aiwei Liu**, Xuming Hu, Lijie Wen, Philip S. Yu

A survey of text watermarking in the era of large language models

- **Aiwei Liu***, Leyi Pan*, Yijian Lu, Jingjing Li, Xuming Hu, Lijie Wen, Irwin King, Philip S. Yu

MarkLLM: An Open-Source Toolkit for LLM Watermarking

- Leyi Pan, **Aiwei Liu**, Zhiwei He, Zitian Gao, Xuandong Zhao, Yijian Lu, Binglin Zhou, Shuliang Liu, Xuming Hu, Lijie Wen, Irwin King

Refiner: Restructure Retrieval Content Efficiently to Advance Question-Answering Capabilities

- Zhonghao Li, Xuming Hu, **Aiwei Liu**, Kening Zheng, Sirui Huang, Hui Xiong

Entropy-Based Decoding for Retrieval-Augmented Large Language Models

- Zexuan Qiu, Zijing Ou, Bin Wu, Jingjing Li, **Aiwei Liu**, Irwin King

ChatCite: LLM Agent with Human Workflow Guidance for Comparative Literature Summary

- Yutong Li, Lu Chen, **Aiwei Liu**, Kai Yu, Lijie Wen

On the Robustness of Document-Level Relation Extraction Models to Entity Name Variations

- Shiao Meng, Xuming Hu, **Aiwei Liu**, Fukun Ma, Yawen Yang, Shuang Li, Lijie Wen

WORK EXPERIENCE

University of Illinois Chicago

Visiting Scholar Advisor: Prof Philip S. Yu (ACM Fellow)

- Project: Trustworthy of Large Language Models

University of Illinois Chicago

July. 2024 – Present

Apple AIML Group

Research Intern

Mar 2023 – Present

- Project: Prompt Difficulty Evaluation, Safety alignment for LLM
- Accomplishments: 1) Developed a LLM-based automatic attributes identification methods for prompt difficulty evaluation. 2) A safety alignment method for LLM that does not require manual annotation of preference data.

Chinese University of Hong Kong

Visiting Scholar Advisor: Prof Irwin King (IEEE Fellow)

Chinese University of Hong Kong

Nov. 2023 – May 2024

- Project: Watermark for Large Language Models
- Accomplishments: 1) Developed an unforgeable publicly verifiable watermark for Large Language Models 2) Write a comprehensive survey about the text watermarking in the era of LLMs.

SELECTED AWARDS

Tsinghua Excellent Student Award (SecondClass)

2022

Outstanding Graduates Nanjing University

2020

China Electronics Technology Group Scholarship

2019

National Scholarship

2018

Hainan Airlines Scholarship

2017

PROFESSIONAL SERVICES

Program Committee/ Reviewer

- The Annual Meeting of the Association for Computational Linguistics (ACL)
- The Annual Conference on Empirical Methods in Natural Language Processing (EMNLP)
- The Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)
- The Annual Conference of the European Chapter of the Association for Computational Linguistics (EACL)
- The ACM WWW International World Wide Web Conference (WWW)