

Yuxin Jiang

☎ +86 132 4231 6959 | ✉ yjiangcm@connect.ust.hk | 🏠 WeChat: Yuxin_Jiang_ | 💻 yjiangcm.github.io

Research Interest

- **Instruction Tuning of Large Language Models**, especially on enhancing and evaluating the capability of language models to comprehend and execute complex instructions accurately.
- **Alignment Tuning of Large Language Models**, concentrating on averting models' unintended behaviors and ensuring LLMs align with human expectations.
- **Contrastive Learning in NLP**, focusing on leveraging contrastive learning to enhance the quality of embeddings and to enable more nuanced and context-aware language model performances.

Education

The Hong Kong University of Science and Technology (HKUST)

Ph.D. in Data Science and Analytics GPA: 3.9/4.3, Top 5%

- Supervisor: [Prof. Wei Wang](#)

Hong Kong SAR

Sep. 2021 – Jul. 2025 (Expected)

The Hong Kong University of Science and Technology (HKUST)

M.S. in Big Data and Technology GPA: 4.0/4.3, Top 5%

- Excellent Student Scholarship

Hong Kong SAR

Sep. 2020 – Jul. 2021

Shanghai University (SHU)

B.S. in Mathematics and Applied Mathematics GPA: 3.7/4.0, Top 10%

- Outstanding Graduates of Shanghai

Shanghai, China

Sep. 2016 – Jul. 2020

Internship

Speech and Semantic Group, Huawei Noah's Ark Lab

LLM Research Intern; Mentor: [Dr. Yufei Wang](#)

Hong Kong SAR

Sep. 2023 – Feb. 2024

- Conducted in-depth evaluations of large language models, including complex instruction following and long-context understanding. Successfully integrated the evaluation system into Pangu's development pipeline.
- Developed and proposed a novel "Learning to Edit" (LTE) framework, enabling effective and efficient knowledge editing within large language models.

Publications

(First Author: 7 papers, Accepted: 11 papers, Total: 14 papers, Citation: 200+)

The following are 11 accepted publications:

- [1] [Learning to Edit: Aligning LLMs with Knowledge Editing](#)
Yuxin Jiang, Yufei Wang, Chuhan Wu, Wanjun Zhong, Xingshan Zeng, Jiahui Gao, Liangyou Li, Xin Jiang, Lifeng Shang, Ruiming Tang, Qun Liu, Wei Wang. **ACL 2024 (Main)**
- [2] [FollowBench: A Multi-level Fine-grained Constraints Following Benchmark for Large Language Models](#)
Yuxin Jiang, Yufei Wang, Xingshan Zeng, Wanjun Zhong, Liangyou Li, Fei Mi, Lifeng Shang, Xin Jiang, Qun Liu, Wei Wang. **ACL 2024 (Main)**
- [3] [Lion: Adversarial Distillation of Proprietary Large Language Models](#)
Yuxin Jiang, Chunkit Chan, Mingyang Chen, Wei Wang. **EMNLP 2023 (Main, Oral)**
- [4] [Global and Local Hierarchy-aware Contrastive Framework for Implicit Discourse Relation Recognition](#)
Yuxin Jiang, Linhan Zhang, Wei Wang. **ACL 2023 (Findings)**
- [5] [Improved Universal Sentence Embeddings with Prompt-based Contrastive Learning and Energy-based Learning](#)
Yuxin Jiang, Linhan Zhang, Wei Wang. **EMNLP 2022 (Findings)**

- [6] [Dual Multi-head Co-Attention for Reading Comprehension of Abstract Meaning](#)
Yuxin Jiang, Ziyi Shou, Qijun Wang, Hao Wu, Fangzhen Lin. **SemEval 2021 (ACL Workshop)**
- [7] [AMR-DA: Data Augmentation by Abstract Meaning Representation](#)
Ziyi Shou, **Yuxin Jiang**, Fangzhen Lin. **ACL 2022 (Findings)**
- [8] [Exploring the Potential of ChatGPT on Sentence Level Relations: A Focus on Temporal, Causal, and Discourse Relations](#)
Chunkit Chan, Jiayang Cheng, Weiqi Wang, **Yuxin Jiang**, Tianqing Fang, Xin Liu, Yangqiu Song. **EACL 2024 (Findings)**
- [9] [Audience Persona Knowledge-Aligned Prompt Tuning Method for Online Debate](#)
Chunkit Chan, Jiayang Cheng, Xin Liu, Yauwai Yim, **Yuxin Jiang**, Zheyue Deng, Haoran Li, Yangqiu Song, Ginny Y. Wong, Simon See. **ECAI 2024**
- [10] [M4LE: A Multi-Ability Multi-Range Multi-Task Multi-Domain Long-Context Evaluation Benchmark for Large Language Models](#)
Wai-Chung Kwan, Xingshan Zeng, Yufei Wang, Yusen Sun, Liangyou Li, **Yuxin Jiang**, Lifeng Shang, Xin Jiang, Qun Liu, Kam-Fai Wong. **ACL 2024 (Main, Outstanding Paper Award)**
- [11] [Weighted Sampling for Masked Language Modeling](#)
Linhan Zhang, Qian Chen, Wen Wang, Chong Deng, Xin Cao, Kongzhang Hao, **Yuxin Jiang**, Wei Wang. **ICASSP 2023 (Top 3% Paper Recognition)**

The following are 3 under-review manuscripts:

- [12] [Bridging and Modeling Correlations in Pairwise Data for Direct Preference Optimization](#)
Yuxin Jiang, Bo Huang, Yufei Wang, Xingshan Zeng, Liangyou Li, Yasheng Wang, Xin Jiang, Lifeng Shang, Ruiming Tang, Wei Wang
- [13] [MT-Eval: A Multi-Turn Capabilities Evaluation Benchmark for Large Language Models](#)
Wai-Chung Kwan, Xingshan Zeng, **Yuxin Jiang**, Yufei Wang, Liangyou Li, Lifeng Shang, Xin Jiang, Qun Liu, Kam-Fai Wong
- [14] [Exploiting Correlations Between Contexts and Definitions with Multiple Definition Modeling](#)
Linhan Zhang, Qian Chen, Wen Wang, **Yuxin Jiang**, Bing Li, Wei Wang, Xin Cao

Honors and Awards

- 2024 **Outstanding Paper Award (Top 1%), ACL 2024**
- 2023 **Research Travel Grant Award, HKUST**
 Top 3% Paper Recognition, ICASSP 2023
- 2021 **Postgraduate Studentship (Top 5%), HKUST**
 School of Engineering Excellent Student Scholarship (Top 5%), HKUST
 School of Engineering Entrance Scholarship (Top 5%), HKUST
- 2020 **Outstanding Graduates of Shanghai (Top 1%), Shanghai Municipal Education Commission**
- 2016-19 **Grand Prize Scholarship (Top 3%), Shanghai University**
 Leadership Scholarship (Top 3%), Shanghai University
 Excellent Student (Top 3%), Shanghai University

Skills

Programming Skills: Python, Matlab, C++, HTML, etc.

Languages: English (IELTS 7.0, GRE 324), Mandarin Chinese (Native), Cantonese (Elementary).