

# 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%

- Supervisor: [Prof. Fangzhen Lin](#)

Hong Kong SAR

Sep. 2020 – Jul. 2021

### Shanghai University (SHU)

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

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

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: 15 papers, Citation: 300+)

[1] [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. **ICLR 2025**

[2] [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)**

[3] [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)**

[4] [Lion: Adversarial Distillation of Proprietary Large Language Models](#)

**Yuxin Jiang**, Chunkit Chan, Mingyang Chen, Wei Wang. **EMNLP 2023 (Main, Oral)**

[5] [Global and Local Hierarchy-aware Contrastive Framework for Implicit Discourse Relation Recognition](#)

**Yuxin Jiang**, Linhan Zhang, Wei Wang. **ACL 2023 (Findings)**

- [6] [Improved Universal Sentence Embeddings with Prompt-based Contrastive Learning and Energy-based Learning](#)  
Yuxin Jiang, Linhan Zhang, Wei Wang. **EMNLP 2022 (Findings)**
- [7] [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)**
- [8] [When Evolution Strategy Meets Language Models Tuning](#)  
Bo Huang, Yuxin Jiang, Mingyang Chen, Yi Wang, Hongyang Chen, Wei Wang. **COLING 2025**
- [9] [AMR-DA: Data Augmentation by Abstract Meaning Representation](#)  
Ziyi Shou, Yuxin Jiang, Fangzhen Lin. **ACL 2022 (Findings)**
- [10] [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. **EMNLP 2024 (Main)**
- [11] [RevisEval: Improving LLM-as-a-Judge via Response-Adapted References](#)  
Qiyuan Zhang, Yufei Wang, Tiezheng Yu, Yuxin Jiang, Chuhan Wu, Liangyou Li, Yasheng Wang, Xin Jiang, Lifeng Shang, Ruiming Tang, Fuyuan Lyu, Chen Ma. **ICLR 2025**
- [12] [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)**
- [13] [Audience Persona Knowledge-Aligned Prompt Tuning Method for Online Debate](#)  
Chunkit Chan, Jiayang Cheng, Xin Liu, Yauwai Yim, Yuxin Jiang, Zheyang Deng, Haoran Li, Yangqiu Song, Ginny Y. Wong, Simon See. **ECAI 2024**
- [14] [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)**
- [15] [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)**

## Honors and Awards

2024	ACL 2024 Outstanding Paper Award, TOP 1%	Association for Computational Linguistics
2023	ICASSP 2023 Top 3% Paper Award, TOP 1%	IEEE ICASSP
	Research Travel Grant Award, TOP 5%	HKUST
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, C++, SQL, Matlab, HTML, etc.

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