

Yuqing Yang

E-mail: yyang063@usc.edu * *Phone:* +1-213-696-4881

Website: <https://ayyyq.github.io/>

Address: 1246 W 30th St, Los Angeles, California, 90007, USA

Education

University of Southern California

PhD student in Computer Science

Los Angeles, USA

August 2024 - Expected May 2028

Advisor: Prof. Robin Jia

Research Focus: Large Language Models, Model Diagnostics.

Fudan University

Master of Science in Computer Science

Shanghai, China

Advisor: Prof. Xipeng Qiu

September 2021 - June 2024

Research Focus: Natural Language Processing, including Information Extraction and LLM Alignment.

University of Chinese Academy of Sciences

Bachelor of Engineering in Computer Science

Beijing, China

GPA: 3.94/4.00 Rank: 3/112

September 2017 - June 2021

Thesis Title: *Research on Key Methods of Event Factuality Prediction for Unstructured Texts.*

Experience

Amazon Bedrock Core Science

Mentor: Qi Zhu

Santa Clara, USA

May 2025 - Aug 2025

- Worked on remaining issues of LLMs in table reasoning.

Generative Artificial Intelligence Research Lab

Advisor: Pengfei Liu

Shanghai, China

June 2023 - June 2024

- Worked on LLM alignment, with a particular focus on knowledge and reasoning of LLMs.

Amazon AWS AI

Mentor: Qipeng Guo

Shanghai, China

July 2021 - June 2023

- Worked on information extraction, especially focusing on relation extraction and event extraction.

Publications

* indicates equal contribution.

- **Yuqing Yang**, Robin Jia. When Do LLMs Admit Their Mistakes? Understanding the Role of Model Belief in Retraction. *arXiv preprint 2025*.
- **Yuqing Yang**, Yan Ma, Pengfei Liu. Weak-to-Strong Reasoning. *EMNLP Findings 2024*.
- Steffi Chern*, Zhulin Hu*, **Yuqing Yang***, Ethan Chern, Yuan Guo, Jiahe Jin, Binjie Wang, Pengfei Liu. BeHonest: Benchmarking Honesty of Large Language Models. *arXiv preprint 2024*.
- Zhen Huang, Zengzhi Wang, Shijie Xia, Xuefeng Li, Haoyang Zou, Ruijie Xu, Run-Ze Fan, Lyumanshan Ye, Ethan Chern, Yixin Ye, Yikai Zhang, **Yuqing Yang**, Ting Wu, Binjie Wang, Shichao Sun, Yang Xiao, Yiyuan Li, Fan Zhou, Steffi Chern, Yiwei Qin, Yan Ma, Jiadi Su, Yixiu Liu, Yuxiang Zheng, Shaoting Zhang, Dahua Lin, Yu Qiao, Pengfei Liu. OlympicArena: Benchmarking Multi-discipline Cognitive Reasoning for Superintelligent AI. *NeurIPS Datasets and Benchmarks 2024*.

- **Yuqing Yang**, Ethan Chern, Xipeng Qiu, Graham Neubig, Pengfei Liu. Alignment for Honesty. *NeurIPS 2024*.
- Kai Lv, **Yuqing Yang**, Tengxiao Liu, Qinghui Gao, Qipeng Guo, Xipeng Qiu. Full Parameter Fine-tuning for Large Language Models with Limited Resources. *ACL 2024 (Oral)*.
- Tengxiao Liu, Qipeng Guo, **Yuqing Yang**, Xiangkun Hu, Yue Zhang, Xipeng Qiu, Zheng Zhang. Plan, Verify and Switch: Integrated Reasoning with Diverse X-of-Thoughts. *EMNLP 2023*.
- Kai Lv*, Shuo Zhang*, Tianle Gu, Shuhao Xing, Jiawei Hong, Keyu Chen, Xiaoran Liu, **Yuqing Yang**, Honglin Guo, Tengxiao Liu, Yu Sun, Qipeng Guo, Hang Yan, Xipeng Qiu. CoLLiE: Collaborative Training of Large Language Models in an Efficient Way. *EMNLP Demo 2023*.
- **Yuqing Yang***, Qipeng Guo*, Xiangkun Hu, Yue Zhang, Xipeng Qiu, Zheng Zhang. An AMR-based Link Prediction Approach for Document-level Event Argument Extraction. *ACL 2023*.
- Qipeng Guo*, **Yuqing Yang***, Hang Yan, Xipeng Qiu, Zheng Zhang. DORE: Document ordered relation extraction based on generative framework. *EMNLP 2022 Findings*.
- Pengfei Cao, Yubo Chen, **Yuqing Yang**, Kang Liu, Jun Zhao. Uncertain local-to-global networks for document-level event factuality identification. *EMNLP 2021*.

Skills

English Proficiency	TOEFL - 108
Programming Languages	Python, C/C++, verilog
Professional Services	Reviewer for ARR, NeurIPS 2025, ICLR 2026
Teaching Assistant	Probability Theory and Mathematical Statistics, Fudan Undergrad Course, Fall 2022