



Jiayue Pu

University of Chinese Academy of Sciences
Beijing, China

+86-13625869881

✉ pujiayue22@mailsucas.ac.cn

✉ pujiayue0222@gmail.com

@pujiayue

EDUCATION

University of Chinese Academy of Sciences

B.S. in Computer Science

– GPA: 3.96 / 4.0

– Rank: 2 / 79

– Specialized Courses: *AI Computing Systems(98)*, *Introduction to Computer Science(97)*, *Calculus(97)*, *Fundamentals of Computer Graphics(96)*, *Discrete Mathematics(95)*, *Digital Circuits(95)*, *Introduction to Machine Learning(94)*, *Foundations of Theoretical Computer Science(94)*, *Fundamentals of Programming and Experiments(94)*, *Stanford CS224n: Natural Language Processing with Deep Learning(self-learning)*.

Aug. 2022 - Present

Beijing, China

RESEARCH EXPERIENCE

Key Laboratory of AI Safety, Institute of Computing Technology, CAS

Research Assistant, advised by Prof. Xueqi Cheng and Prof. Fei Sun

Sep. 2023 - Present

Beijing, China

– **Research Focus:** Trustworthy Machine Learning & Large Language Model Security

* **Unlearning Evaluation Frameworks for Trustworthy AI**

- Leading the development of evaluation metrics for LLM unlearning effectiveness and privacy preservation
- Introducing new definitions of memory under unlearning scenarios and utilizing their inverse as normalized metrics to evaluate unlearning algorithms
- Authoring the evaluation methodology section for systematic review papers on machine unlearning in LLMs

* **Memorization Analysis of LLM**

- Analyzing existing definitions of LLM memorization and their limitations in practical, domain-specific applications, particularly in scenarios such as unlearning
- Utilizing jailbreak algorithms to investigate LLM memorization, analyzing both fact-level and string-level retention.

Prof. Dawn Songs group at UC Berkeley

Research collaborator, advised by Prof. Dawn Song and Yiyou Sun

Feb. 2025

California, USA

– **Research Focus:** Hallucination & LLM Agent

- * Developed a systematic evaluation benchmark to detect, analyze, and further mitigate hallucinations in LLM-based agents, improving trustworthiness and adaptability in complex real-world applications

The Association for Computational Linguistics

Volunteer, advised by Prof. Fei Sun

Aug. 2024

Bangkok, Thailand

– Communicated with international scholars and attended academic events, including Keynotes, Orals, Posters etc.

HONORS AND AWARDS

National Scholarship (Top 2%) *Ministry of Education of the People's Republic of China* 2024

First Class Academic Scholarship (Top 5%) *University of Chinese Academy of Sciences* 2023, 2024

Model Student (totally 20 school-wide) *University of Chinese Academy of Sciences* 2024

Triple-A Student *University of Chinese Academy of Sciences* 2023, 2024

Outstanding Communist Youth League Member *University of Chinese Academy of Sciences* 2023, 2024

TECHNICAL SKILLS AND INTERESTS

Programming Languages: C, Python, Verilog, HTML/CSS, VB, Go

Tools and Technologies: Git, Linux

Frameworks: PyTorch

Areas of Interest: Music, especially singing and playing the cello

PUBLIC AFFAIRS

Peer Mentor: Class 2412

University of Chinese Academy of Sciences

Aug. 2024 - Present

Principal Cellist: UCAS Philharmonic Orchestra

University of Chinese Academy of Sciences

Mar. 2023 - Present

President: the Undergraduate Choir of UCAS

University of Chinese Academy of Sciences

Aug. 2023 - Aug. 2024