Jiayue Pu

University of Chinese Academy of Sciences
Beijing, China

+86-13625869881
pujiayue22@mails.ucas.ac.cn
pujiayue0222@gmail.com
@pujiayue

Education

University of Chinese Academy of Sciences *Aug. 2022 - Present*
B.S. in Computer Science Beijing, China

– GPA: 3.96 / 4.0
– Rank: 2 / 87
– Specialized Courses: *AI Computing Systems*(98), *Fundamentals of Computer Graphics*(96), *Discrete Mathematics*(95), *Digital Circuits*(95), *Introduction to Machine Learning*(94), *Foundations of Theoretical Computer Science*(94), *UC Berkeley CS194: Advanced Large Language Model Agents*(A).

Research Experience

Key Laboratory of AI Safety, Institute of Computing Technology, CAS *Sep. 2023 - Present*
Research Assistant, advised by Prof. Xueqi Cheng and Prof. Fei Sun Beijing, China

∗ **Semantic-level LLM Unlearning Evaluation Metric (SEE)**: Addressed the limitations of existing unlearning evaluation methods that rely on exact string matching and fail to detect semantically equivalent expressions. Proposed a novel evaluation metric based on semantic entailment that enforces low prompt-target semantic similarity and directly measures whether model outputs semantically entail target knowledge, enabling reliable assessment of true forgetting. **First-author paper**: SEE: Semantic Extraction-Based Evaluation for Reliable LLM Unlearning, expected submission to ICLR 2025.
∗ **LLM Unlearning Mechanisms Survey**: Systematically investigated and summarized key technical challenges in the LLM unlearning field, with primary responsibility for literature review and methodological analysis of "Unlearning Evaluation" and "Inference-time Unlearning" sections. **Survey paper**: LLM Unlearning Survey, expected submission to TACL 2025.

Prof. Dawn Song's group at UC Berkeley *Feb. 2025 - Present*
Research collaborator, advised by Prof. Dawn Song and Yiyou Sun California, USA

∗ **Hallucination in LLM Agent:** Designed and developed a systematic benchmark platform to analyze the root causes of hallucinations in LLM agents across multi-environment interactions. Built comprehensive evaluation frameworks to identify key scenarios where agents generate hallucinations and proposed corresponding mitigation strategies, providing crucial technical support for enhancing agent reliability. **Second-author paper**: Clicking the Void: LLM Agent is Hallucinating and Where to Find Them, submitted to EMNLP 2025.

Course Projects in UCAS *Sep. 2024 - Jan. 2025*
Projects leader, advised by Prof. Lin Gao and Prof. Wei Chen Beijing, China

∗ **Aspect-based Sentiment Classification:** Fine-tuned multiple models including BERT, SVM, CNN and LLM PE etc. to achieve over 75% accuracy.
∗ **Ray Tracing Project:** Developed a ray tracing system with realistic light propagation and reflection, incorporating interactive features, stereoscopic textures, and adaptive sampling for optimized rendering.

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*
Outstanding Individual in Social Practice *University of Chinese Academy of Sciences* *2023,2024*

Technical Skills

Programming Languages: C, Python, Verilog, HTML/CSS, VB, Go
Tools and Frameworks: Git, Linux, NumPyPyTorch

Publc Affairs

Peer Mentor: Class 2412 *Aug. 2024 - Present*
Principal Cellist: UCAS Philharmonic Orchestra *Mar. 2023 - Present*
President: the Undergraduate Choir of UCAS *Aug. 2023 - Aug. 2024*

updated on July 17, 2025