Xu Minmin

✉ xuminmin729@gmail.com 🔗 https://xmm-world.github.io/

Education

University of Science and Technology of China (USTC) , Hefei, China 09/2021 – 07/2025

B.Eng. in Computer Science and Technology, GPA: 3.75/4.30

Publications

(* stands for equal contribution)

[1] **(EMNLP24 Main)** Direct Multi-Turn Preference Optimization for Language Agents.
Li Haoran*, **Xu Minmin***, Zhou Yichen, Chen Zixuan, Gao Wenhao
[PDF] [Code]
[2] **(ICML25)** AdvAgent: Controllable Blackbox Red-teaming on Web Agents.
Sun Jiacheng, Liu Yutong, He Ziming, Peng Zeyu, Tang Lingbo, **Xu Minmin**, Qiu Han, Luo Bo
[Page] [PDF] [Code]

Research Experience

XLANG Lab, HKU Hong Kong

Research Assistant to Prof. Yu Su 04/2025 – present

Topic: **Computer Use Agent**

Secure Learning Lab, UIUC & UChicago Chicago, IL

Research Assistant to Prof. Luo Bo 07/2024 – 12/2024

Topic: **LLM Agent Security**

Lab for Data Science, USTC Hefei, China

Research Assistant to Prof. Gao Wenhao and Prof. He Xiangyu 07/2023 – 06/2024

Topic: **Reinforcement Learning & Language Agents**

Projects

DMPO: Direct Multi-Turn Preference Optimization for Language Agents 07/2023 – 06/2024

- Eliminated the partition function in the BT model and derived the DMPO loss function for language agents in multi-turn scenarios.
- Provided a theoretical explanation for the necessity of adding length normalization to the DPO loss function.
- Conducted extensive experiments on three agent datasets, demonstrating the effectiveness of the DMPO loss in reducing compounding errors and improving robustness to trajectory length disparity.

AdvAgent: Controllable Blackbox Red-teaming on Web Agents 07/2024 – 12/2024

- Developed a black-box attack framework, AdvAgent, which exploited vulnerabilities in VLM-powered web agents by automatically generating and injecting adversarial prompts into web pages.
- Achieved high attack success rates with AdvAgent while maintaining stealthiness and controllability.

Multi-Agent Collaborative Defense for Multi-Lingual Scenarios 09/2024 – 12/2024

- Designed a multi-agent collaborative pipeline that improved defense effectiveness in multi-lingual scenarios.

Selected Awards and Honors

Scholarship

- USTC Fellowship Level-A 2025
- Outstanding Student Scholarship in USTC 2024
- Outstanding Student Scholarship in USTC 2023
- Outstanding Student Scholarship in USTC 2022

Awards

- School Outstanding Psychological Committee Member 2023
- School Outstanding Psychological Committee Member 2022

Additional Information

Skills

- **Programming Languages:** Python, C/C++, HTML, Bash, SQL, Verilog
- **Tools and Frameworks:** Git, LaTeX, PyTorch, Markdown

Selected Courses

- Mathematical Logic (93), Graph Theory (91), Data Structures (92), Advanced Programming and Practice (91), Web Information Processing and Application (95), Principles and Techniques of Compiler (98)