Zeyi Liao

☑ lzy37ld@gmail.com / liao.629@osu.edu | ☑ X | ♠ Personal Website | ☎ Google Scholar | ⋒ Me

(U) Research Interests

I am broadly interested in general AI, with a focus on NLP and multimodal (esp. vision-language) AI. My ultimate goal is to build helpful yet robust and reliable AI systems that enhance and enrich human life.

A Education

 The Ohio State University Ph.D. in Computer Science and Engineering 	Aug. 2023 – May. 2028 (expected)
• University of Southern California M.S in Electronic and Computer Engineering	Aug. 2021 – May. 2023
• Beijing Jiaotong University B.S in Electronic Engineering	Aug. 2016 – May. 2020

Research Experience	
• Research Assistant, The Ohio State University Advisor: Prof. Huan Sun	Aug. 2023 – Present
• Research Assistant, University of Southern California Advisor: Prof. Xiang Ren	Aug. 2022 – May. 2023
• Indepedent Research, Beijing Jiaotong University	Aug. 2020 – May. 2021

Publications (selected, see full list on)

Joint Demonstration and Preference Learning Improves Policy Alignment with Human Feedback

Chenliang Li, Siliang Zeng*, Zeyi Liao*, Jiaxiang Li, Dongyeop Kang, Alfredo Garcia, Mingyi Hong Preprint on Arxiv, 2024. [PDF]

[6] Amplegcg: Learning a universal and transferable generative model of adversarial suffixes for jailbreaking both open and closed llms

Zeyi Liao, Huan Sun

Conference on Language Modeling (COLM), 2024. [PDF], **У** X on AmpleGCG-plus

[5] AttributionBench: How Hard is Automatic Attribution Evaluation?

Yifei Li, Xiang Yue, Zeyi Liao, Huan Sun

Annual Conference of the Association for Computational Linguistics (Findings of ACL), 2024. [PDF]

[4] A Trembling House of Cards? Mapping Adversarial Attacks against Language Agents

Lingbo Mo, Zeyi Liao, Boyuan Zheng, Yu Su, Chaowei Xiao, Huan Sun Preprint on Arxiv, 2024. [PDF]

[3] In Search of the Long-Tail: Systematic Generation of Long-Tail Knowledge via Logical Rule **Guided Search**

Huihan Li, Zeyi Liao*, Yuting Ning*, Siyuan Wang, Xiang Lorraine Li, Ximing Lu, Faeze Brahman, Wenting Zhao, Yejin Choi, Xiang Ren

Preprint on Arxiv, 2023. [PDF]

^{*} means equal contribution

[2] Chatcounselor: A large language models for mental health support

June M Liu, Donghao Li, He Cao, Tianhe Ren, Zeyi Liao, Jiamin Wu

ACM International Conference on Information and Knowledge Management (CIKM), 2023. [PDF]

[1] RobustLR: A diagnostic benchmark for evaluating logical robustness of deductive reasoners Soumya Sanyal, Zeyi Liao, Xiang Ren

Conference on Empirical Methods in Natural Language Processing (EMNLP), 2022. [PDF]

♣ Teaching Experience

• CSE 3521: Survey of Artificial Intelligence I: Basic Techniques @ OSU

Aug. 2023 – Jan. 2024

Q Awards

• Master Honor Student, University of Southern California

• First Class Scholarship, Beijing Jiaotong University

• Merit Student, Beijing Jiaotong University

Aug. 2022 Aug. 2020/2019/2018

Aug. 2018/2017