Qiusi Zhan

Email: qiusiz2@illinois.edu | Homepage: zqs1943.github.io | Github: github.com/ZQS1943

Summary

Research Interests: Exploring safety in (multimodal) Large Language Models (LLMs) and LLM agents, with emphasis on risk identification and safety improvement.

Objective: Seeking research internship opportunities for Summer 2026.

Education

University of Illinois at Urbana-Champaign

Ph.D. Computer Science, advised by Prof. Daniel Kang 08/2023 – 08/2027 (Expected)

M.Eng. Electrical & Computer Engineering, advised by Prof. Heng Ji 08/2021 – 12/2022

Peking University 09/2017 – 07/2021

B.S. Computer Science, advised by Prof. Sujian Li

Publications

Full list at Google Scholar: https://scholar.google.com/citations?user=XaYJrgoAAAAJ&hl=en

LLM Safety

"SafeSearch: Do Not Trade Safety for Utility in LLM Search Agents"
Qiusi Zhan*, Angeline Yasodhara, Abdelrahman Zayed, Xingzhi Guo, Daniel Kang, Joo-Kyung Kim
In Submission

"Visual Backdoor Attacks on MLLM Embodied Decision Making via Contrastive Trigger Learning"
Qiusi Zhan*, Hyeonjeong Ha*, Rui Yang, Sirui Xu, Hanyang Chen, Liang-Yan Gui, Yu-Xiong Wang, Huan Zhang, Heng Ji, Daniel Kang
In Submission

"MM-PoisonRAG: Disrupting Multimodal RAG with Local and Global Poisoning Attacks"
Hyeonjeong Ha*, Qiusi Zhan*, Jeonghwan Kim, Dimitrios Bralios, Saikrishna Sanniboina, Nanyun Peng, Kai-wei Chang, Daniel Kang, Heng Ji
In Submission

"Adaptive Attacks Break Defenses Against Indirect Prompt Injection Attacks on LLM Agents"
Qiusi Zhan, Richard Fang, Henil Shalin Panchal, Daniel Kang
NAACL 2025 Findings, TrustNLP Workshop Spotlight

"InjecAgent: Benchmarking Indirect Prompt Injections in Tool-Integrated Large Language Model Agents"
Qiusi Zhan, Zhixiang Liang, Zifan Ying, Daniel Kang
ACL 2024 Findings

"Removing RLHF Protections in GPT-4 via Fine-Tuning"
Qiusi Zhan, Richard Fang, Rohan Bindu, Akul Gupta, Tatsunori Hashimoto, Daniel Kang
NAACL 2024

Information Extraction

"GLEN: General-Purpose Event Detection for Thousands of Types"
Qiusi Zhan*, Sha Li*, Kathryn Conger, Martha Palmer, Heng Ji, Jiawei Han
EMNLP 2023

"EA^{2}E: Improving Consistency with Event Awareness for Document-Level Argument Extraction"
Qi Zeng*, Qiusi Zhan*, Heng Ji
NAACL 2022 Findings

*Indicates equal contribution

Internships

Applied Scientist Intern Amazon 05/2025 – 08/2025
Topic: Safety alignment of LLM-based search agents
Manager: Joo-Kyung Kim

Applied Scientist Intern Microsoft 05/2024 – 08/2024
Topic: Multi-source information-augmented conversational LLM agents
Manager: Yu Hu

Research Scientist Intern JD.com Silicon Valley Labs 01/2022 – 05/2022
Topic: User-simulator assisted open-ended conversational recommendation
Manager: Lingfei Wu

Applied Scientist Intern ByteDance 04/2021 – 07/2021
Topic: Template-based K-12 math problem solving
Manager: Bo Zhao