Qiusi Zhan

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

Summary

Research Interests: Safety in Large Language Models (LLMs) and LLM agents.

Objective: Seeking research internship opportunities for Summer 2025.

Education

University of Illinois at Urbana-Champaign 08/2023 – Present
Ph.D. Computer Science, advised by Daniel Kang

University of Illinois at Urbana-Champaign 08/2021 – 12/2022
M.Eng. Electrical & Computer Engineering, advised by Heng Ji

Peking University 09/2017 – 07/2021
B.S. Computer Science, advised by Sujian Li

Publications

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

LLM Safety

"Adaptive Attacks Break Defenses Against Indirect Prompt Injection Attacks on LLM Agents"
Qiusi Zhan, Richard Fang, Henil Shalin Panchal, Daniel Kang
In submission

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

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

"LLM Agents can Autonomously Hack Websites"
Richard Fang, Rohan Bindu, Akul Gupta, Qiusi Zhan, Daniel Kang
In submission

"Teams of LLM Agents can Exploit Zero-Day Vulnerabilities"
Richard Fang, Rohan Bindu, Akul Gupta, Qiusi Zhan, Daniel Kang
In submission

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"
Qiusi Zhan*, Qi Zeng*, Heng Ji
NAACL 2022 Findings

*Indicates equal contribution

Internships

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 System
Manager: Lingfei Wu

Applied Scientist Intern ByteDance 04/2021 – 07/2021
Topic: Template-Based K-12 Math Problem Solving
Manager: Bo Zhao