Keyan Guo

917-374-6606 | keyanguo@buffalo.edu | keyanUB.github.io

Research Intersests

- Artificial Intelligence and Machine Learning (AI/ML) Security
- Trustworthy Generative AI
- Online Hate, Harassment and Abuse Detection, Explanation and Mitigation
- Security and Privacy in Online Social Networks
- Autonomous Vehicle Cybersecurity

Education

University at Buffalo, SUNY Buffalo, NY
Ph.D. in Computer Science and Engineering *January 2022 – present*

- Advisor: Prof. Hongxin Hu
- GPA: 4.0/4.0

University at Buffalo, SUNY Buffalo, NY
M.S. in Engineering Science *July 2019 – June 2021*

- Supervisor: Prof. Mingchen Gao
- GPA: 3.8/4.0

Qingdao University Qingdao, China
Bachelor of Engineering in Information Engineering *July 2014 – June 2018*

- GPA: 3.6/4.0

Professional Publications

- **Keyan Guo**, Ayush Utkarsh, Wenbo Ding, Isabelle Ondracek, Ziming Zhao, Guo Freeman, Nishant Vishwamitra, Hongxin Hu. “Moderating Illicit Online Image Promotion for Unsafe User Generated Content Games Using Large Vision-Language Models”. Accepted by ***33rd USENIX Security Symposium (USENIX Security)(Top conference in Computer Security. Known as a “Big 4” Security Conference)***, 2024

- Nishant Vishwamitra*, **Keyan Guo***, Farhan Tajwar Romit, Isabelle Ondracek, Long Cheng, Ziming Zhao, Hongxin Hu. “Moderating New Waves of Online Hate with Chain-of-Thought Reasoning in Large Language Models”. In *Proceedings of* ***45th IEEE Symposium on Security and Privacy (S&P) (Top conference in Computer Security. Known as a “Big 4” Security Conference, Acceptance rate:14.9%)***, 2024

- **Keyan Guo**, Alexander Hu, Jaden Mu, Ziheng Shi, Ziming Zhao, Nishant Vishwamitra, Hongxin Hu. “An Investigation of Large Language Models for Real-World Hate Speech Detection”. In *Proceedings of 22st IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2023

- Nishant Vishwamitra, **Keyan Guo**, Liao Song, Jaden Mu, Zheyuan Ma, Long Cheng, Ziming Zhao, Hongxin Hu. “Understanding and Analyzing COVID-19-related Online Hate Propagation Through Hateful Memes Shared on Twitter”. In *Proceedings of IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2023

- Ebuka Okpala, Nishant Vishwamitra, **Keyan Guo**, Liao Song, Long Cheng, Hongxin Hu, Yongkai Wu, Xiaohong Yuan, Jeannette Wade, Sajad Khorsandroo. “AI-Cybersecurity Education Through Designing AI-based Cyberharassment Detection Lab”. Accepted by *IEEE Frontiers in Education Conference (FIE)*, 2023

- Nishant Vishwamitra, **Keyan Guo**, Hongxin Hu, Ziming Zhao, Long Cheng, Feng Luo. "Understanding and Measuring Robustness of Vision and Language Multimodal Models". In *Proceedings of International Conference on Secure Knowledge Management (SKM)*, 2023

- Wenbo Ding, Liao Song, **Keyan Guo**, Ziming Zhao, Hongxin Hu. "Exploring Vulnerabilities in Voice Command Skills for Connected Vehicles". In *Proceedings of EAI International Conference on Security and Privacy in Cyber-Physical Systems and Smart Vehicles (EAI SmartSP)*, 2023

- **Keyan Guo**, Wentai Zhao, Jaden Mu, Nishant Vishwamitra, Ziming Zhao, Hongxin Hu. "Understanding the Generalizability of Hateful Memes Detection Models Against COVID-19-related Hateful Memes". In *Proceedings of 21st IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2022

- Shaik Sabiha, **Keyan Guo**, Foad Hajiaghajani, Chunming Qiao, Hongxin Hu, Ziming Zhao. "Demo: Understanding the Effects of Paint Colors on LiDAR Point Cloud Intensities". In *Workshop on Automotive and Autonomous Vehicle Security (AutoSec)*, 2022

Teaching Experience

Graduate Teaching Assistant(TA) Fall 2020 – Spring 2024
University at Buffalo, SUNY *Buffalo, NY*

- Teaching assistant for the following courses
  - CSE 574: Introduction to Machine Learning. Spring 2024
  - CSE 565: Computer Security. Fall 2023
  - CSE 460/560: Data Models and Query Languages. Spring 2023
  - CSE 460/560: Data Models and Query Languages. Fall 2022
  - CSE 460/560: Data Models and Query Languages. Spring 2022
  - CSE 460/560: Data Models and Query Languages. Fall 2021
  - CSE 460/560: Data Models and Query Languages. Spring 2021
  - CSE 368: Introduction to Artificial Intelligence. Fall 2020
- My TA responsibilities include
  - Designing and grading AI/ML course projects
  - Instructor in charge of the Artificial Intelligence (AI) Security
  - Designing and grading hands-on AI security labs
  - Participating in the development of course-related learning materials
  - Answering student questions about course knowledge, assignments and projects
  - Grading assignments, quizzes, exams, and other submissions
  - Filling-in for the instructor for face-to-face course lectures

Guest Speaker
The 18th International AAAI Conference on Web and Social Media *Buffalo, NY*

- ICWSM 2024 Tutorial.
  I and my team members presented our tutorial at the 18th ICWSM conference. The tutorial introduces the topic of machine learning-based online abuse defense, including our designed platform, current research, and self-developed hands-on labs.

University at Buffalo, SUNY *Buffalo, NY*

- SEAS 2023 Lightning Talk.
  I was invited to give a lightning talk about AI-related safety issues and our ongoing research projects in the School of Engineering and Applied Sciences at University of Buffalo. 12 Ph.D. students were invited to this event.

- CSE 702 Seminar: Machine Learning and Cybersecurity.
  I was invited to talk about AI-related safety issues and knowledge of adversarial machine learning (AML) in a seminar course at the University of Buffalo. Over 20 graduate students attended the seminar.

North Carolina A&T State University *Greensboro, NC*

- GenCyber 2022/2023.
  I was invited to speak on AI-related cybersecurity challenges and cyberbullying defense. Additionally, I presented our designed cybersecurity lab at two sessions at North Carolina State A&T University in 2022 and 2023.

Acedemic Experience

Program Committee

- Conference Committee. IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2024
- Session Chair. IEEE International Conference on Machine Learning and Applications (ICMLA), 2023
- Conference Committee. IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2023
- Artifacts Evaluation Committee. Annual Computer Security Applications Conference (ACSAC), 2023

Conference/Journal Paper Reviewer

- IEEE Transactions on Dependable and Secure Computing (TDSC), 2024
- IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2023
- International Conference on Mobility, Sensing and Networking (MSN), 2023
- IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCon), 2023
- IEEE International Conference on Machine Learning and Applications (ICMLA), 2022
- IEEE International Conference on Trust, Privacy and Security in Intelligent Systems, and Applications (TPS), 2022
- Information Systems Frontiers (ISFI), 2022, 2023

Invited Talk

- Great Lakes Security Day 2023.
  I was invited to present our research work about mitigating online hate in the evolving cyber environment at Great Lakes Security Dat (GLSD) 2023 on April 21st.GLSD brings together premier practitioners, researchers, students, and funding partners in security to share the latest advances, debate roadmaps, and future directions, create new collaborations, and seek new opportunities in cybersecurity in and around Western and Upstate New York.

Post Presentation

- IEEE Symposuim on Security and Privacy (S&P), 2024

Honors and Awards

- Celebration of Student Academic Excellence Showcase. University at Buffalo, SUNY, 2024
- CSE Best AI Poster Award. University at Buffalo, SUNY, 2023
- CSE Best Graduate Teaching Award. University at Buffalo, SUNY, 2022

Technique Skills

Languages: C/C++, Java, Python, JavaScript/TypeScript, HTML/CSS, MySQL, NoSQL, LaTeX
Frameworks: PyTorch, Tensorflow
Tools: Git/GitHub, HuggingFace, Vim, Jupyter, Node.js, VS Code, IntelliJ IDEA, MongoDB, Eclipse, Amazon Mechanical Turk

Libraries: Pandas, NumPy, Matplotlib, spaCy, scikit-learn, NetworkX, BeautifulSoup, Foolbox