Bocheng Chen

Contact Information

Computer and Information Science University of Mississippi
Email: bchen5@olemiss.edu

Research Interests

My research focuses on threat modeling for cyber risks in AI-based systems. Identifying and mitigating realistic security vulnerabilities in LLM-based chatbot systems to ensure their secure operation.

Working Experience

University of Mississippi, Department of Computer and Information Science, MS, USA

Tenure-Track Assistant Professor **Aug, 2025 -**

Amazon, WA, USA

Applied Scientist Intern **May, 2024 - August, 2024**

Designed and implemented a RAG-based chatbot to streamline onboarding and automate ticket routing, created an LLM-based evaluation framework to benchmark pipeline performance, and developed a fine-tuning workflow that leveraged user feedback to enhance model accuracy and adaptability.

Education

Michigan State University, East Lansing, Michigan, USA

PhD, Computer Science, May, 2025

Shanghai Jiaotong University, Shanghai, China

BS, Electronic Engineering, May, 2020

Selected Publication

1. Guangliang Liu*, **Bocheng Chen***, Xitong Zhang, Kristen Johnson
   "Diagnosing the Performance Trade-off in Gender Stereotype Mitigation", submitted, 2025
2. Zhiyu Xue*, Guangliang Liu*, **Bocheng Chen**, Kristen Marie Johnson, Ramtin Pedarsani
   "No Free Lunch for Defending Against Prefilling Jailbreak Attack by In-Context Learning", submitted, 2025

Full Publication

1. **Bocheng Chen**, Nikolay Ivanov, Guangjing Wang, Qiben Yan
   "Multi-Turn Hidden Backdoor in Large Language Model-Powered Chatbot Models"
   The 19th ACM Asia Conference on Computer and Communications Security (ASIACCS), 2024
2. **Bocheng Chen**, Hanqing Guo, Qiben Yan
   "The Dark Side of Human Feedback: Poisoning Large Language Models via User Inputs"
   arXiv preprint, 2024
3. **Bocheng Chen**, Guangjing Wang, Hanqing Guo, Yuanda Wang, Qiben Yan
   "Understanding Multi-Turn Toxic Behaviors in Open-Domain Chatbots"
   The 26th International Symposium on Research in Attacks, Intrusions and Defenses (RAID), 2023
4. **Bocheng Chen**, Advait Paliwal, Qiben Yan
   "Jailbreaker in Jail: Moving Target Defense for Large Language Models"
   The 10th ACM CCS Workshop on Moving Target Defense (CCS MTD), 2023
5. **Bocheng Chen**, Nikolay Ivanov, Guangjing Wang, Qiben Yan
   "DynamicFL: Balancing Communication Dynamics and Client Manipulation for Federated Learning"
   The 20th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON), 2023

6. **Bocheng Chen**, Hanqing Guo, Yuanda Wang, Qiben Yan
"FlexLLM: Exploring LLM Customization for Moving Target Defense on Black-Box LLMs Against Jailbreak Attacks"
arXiv preprint, 2024
7. Guangjing Wang, Ce Zhou, Yuanda Wang, **Bocheng Chen**, Hanqing Guo, Qiben Yan
"Beyond Boundaries: A Comprehensive Survey of Transferable Attacks on AI Systems"
ACM Computing Surveys, 2025
8. Hanqing Guo, Junfeng Guo, **Bocheng Chen**, Yuanda Wang, Xun Chen, Heng Huang, Qiben Yan, Xiao Li
"AUDIO WATERMARK: Dynamic and Harmless Watermark for Black-box Voice Dataset Copyright Protection"
The 34th USENIX Security Symposium (USENIX), 2025
9. Yuanda Wang, **Bocheng Chen**, Hanqing Guo, Guangjing Wang, Weikang Ding, Qiben Yan
"ClearMask: Noise-Free and Naturalness-Preserving Protection Against Voice Deepfake Attacks"
Major revision for *The 20th ACM ASIA Conference on Computer and Communications Security (ASIACCS)*, 2025
10. Hanqing Guo, Guangjing Wang, **Bocheng Chen**, Yuanda Wang, Xiao Zhang, Xun Chen, Qiben Yan, Xiao Li
"WavePurifier: Purifying Audio Adversarial Examples via Hierarchical Diffusion Models"
The 30th Annual International Conference On Mobile Computing And Networking (MobiCom), 2024
11. Guangjing Wang, Hanqing Guo, Yuanda Wang, **Bocheng Chen**, Ce Zhou, Qiben Yan
"Protecting Activity Sensing Data Privacy Using Hierarchical Information Dissociation"
2024 IEEE Conference on Communications and Network Security (CNS), 2024
12. Yuanda Wang, Hanqing Guo, Guangjing Wang, **Bocheng Chen**, Qiben Yan
"VSMask: Defending Against Voice Synthesis Attack via Real-Time Predictive Perturbation"
The 16th ACM Conference on Security and Privacy in Wireless and Mobile Networks (WiSec), 2023
13. Guangjing Wang, Nikolay Ivanov, **Bocheng Chen**, Qi Wang, ThanhVu Nguyen, Qiben Yan
"Graph Learning for Interactive Threat Detection in Heterogeneous Smart Home Rule Data"
Proceedings of the ACM on Management of Data (SIGMOD), 2023
14. Hanqing Guo, Guangjing Wang, Yuanda Wang, **Bocheng Chen**, Qiben Yan, Li Xiao
"PhantomSound: Black-Box, Query-Efficient Audio Adversarial Attack via Split-Second Phoneme Injection"
The 26th International Symposium on Research in Attacks, Intrusions and Defenses (RAID), 2023
15. Mohannad Alhanahnah, Clay Stevens, **Bocheng Chen**, Qiben Yan, Hamid Bagheri
"IoTCom: Dissecting Interaction Threats in IoT Systems"
IEEE Transactions on Software Engineering (TSE), 2022

Professional Activities

Journal Reviewer of

- *Transactions on Information Forensics & Security (TIFS)* **2023-2024**
- *Transactions on Knowledge Discovery from Data (TKDD)* **2023-2024**