

Shuo Han

Email: shuo.han.25@ucl.ac.uk |Address: London, UK

Education Background

University College London	2025.03 – Present
• Doctor of Philosophy in Computer Science	
Northwestern University	2022.09 – 2024.08
• Master of Science in Statistics and Data Science	
Boston University	2019.09 – 2022.08
• Bachelor of Arts in Computer Science and Statistics	

Research Interests

Trustworthiness, security, and domain-specific applications of Large Language Models (LLMs), with a focus on addressing reliability and security concerns.

Publication

- **Shuo Han**, Tao Tan, Yuantian Miao, Xiao Chen, Nan Sun., “Prompting Instability: An Empirical Study of LLM Robustness in Code Vulnerability Detection”, AJCAI, 2025
- Zelei Cheng, Xian Wu, Jihao Yu, **Shuo Han**, Xin-Qiang Cai, Xinyu Xing., “Soft-Label Integration for Robust Toxicity Classification”, NeurIPS, 2024
- Chenli Wang, Juyang Wu, Xing Yang, Junfei Wang, Jian Shu, Jiazhong Lu, Yuanyuan Huang, **Shuo Han**., “MC-GAN: an Adversarial Sample Defense Algorithm”, ICCWAMTIP, 2024
- Jian Shu, Bo Xian, Chenli Wang, Jiazhong Lu, Yuanyuan Huang, **Shuo Han**., “A Botnet Data Collection Method for Industrial Internet”, ICCWAMTIP, 2024

Research Experiences

Research in Trustworthy LLMs

University of New South Wales	2024.05 – 2025.09
• Design experiments to test the robustness and uncertainty of LLM responses for cybersecurity tasks.	

Research in AI for Security

Northwestern University	2023.12 – 2024.05
Project Background: Toxicity detection in human-LLM interactions often relies on single-annotator labels that can be biased, so we aim to use crowdsourced labels for more balanced and accurate assessments.	
• Crafted toxic prompts using prompt engineering techniques and annotated them through third-party companies and LLMs. Integrate these crowdsourced annotations using a soft-labeling technique.	

Research in AI Security

Advanced Cryptography and System Security Key Laboratory	2023.05-2023.08
• Explored applying model compression techniques to enhance the model structure, achieving lower computational costs and improved accuracy for Generative Adversarial Networks.	

Academic Services

Graduate Teaching Assistant

Northwestern University

Primary responsibilities: Host office hours, grade assignments, and lead project presentation sessions.

- STAT 332-0/IBIS 432, Spring 2023, Class size:30
- STAT 303-2, Winter 2023, Class size: 100

Volunteer

- The Seventeenth International Conference on Web Search and Data Mining, 2024

Skills

- Languages: Mandarin (native), English, Korean Beginner, Japanese Beginner
- Software: Adobe Illustrator, MS OFFICE
- Programming language: Python, Java, C, R, SQL, CSS, HTML, Java Script, OCaml
- Framework/Technology: Pytorch, Tensorflow, Linux, Git, HuggingFace, Pandas, Numpy, Matplotlib