Shuo Han

Email: shuohan2024.1@u.northwestern.edu |Address: Evanston, IL

Education Background

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
- Awards: Dean's List

Research Interests

Large Language Models (LLMs) and AI security, focusing on enhancing LLM coding capabilities (e.g., code generation, vulnerability detection) and evaluating uncertainty in LLMs.

Publication

- Zelei Cheng, Xian Wu, Jihao Yu, **Shuo Han**, Xin-Qiang Cai, Xinyu Xing., "Soft-Label Integration for Robust Toxicity Classification", NeurIPS, 2024
- **Shuo Han**., "Hydro-GRNNI: Hydrological Graph Recurrent Neural Network for Imputation", Northwestern University, 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 Assistant in LLM for Security

University of New South Wales **2024.05 – Present**

- Design experiments to test the robustness and uncertainty of LLM responses for cybersecurity tasks.
- Develop an automated evaluation framework to assess LLMs' reliability in identifying and reasoning about security-related bugs.

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.
- Incorporated a bi-level optimization algorithm and GroupDRO loss based on topics to compute out-of-distribution loss, addressing distribution shifts caused by variations in annotators and topic difficulties.

Research Assistant in Data Mining

Northwestern University **2023.07-2024.08**

Project Background: The sparsity of U.S. hydrological data impedes effective assessment. To overcome this challenge, we explored using graph neural networks to impute missing data through spatiotemporal dependencies.

- Designed the model to capture upstream-downstream information among all monitoring stations, rather than relying solely on general latitude and longitude data, to better suit hydrological data.
- Use Graph Recurrent Neural Networks to capture spatial and temporal information, leveraging the recurrent structure for temporal data and the graph structure for spatial dependencies.

Research Assistant 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.

- Proposed a data collection method for the industrial internet that includes network traffic and industrial control features, enhancing the accuracy of botnet detection.
- Applied a Logistic Regression approach for botnet detection on the collected dataset, addressing both binary classification and multiclassification tasks.

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