Xun Huang

✉ : robertoxunhuang@gmail.com | ⊙ : https://github.com/bastoni95 |

Education

- **Nanjing University of Science and Technology (NJUST) [🌐]** *B.Eng in Data Science and Big Data Technology*
  Sept. 2022 - Current Wuxi, China

- GPA: 89.32/100, rank top 5%
- IELTS: 6.5

- **Nanyang Technological University (NTU) [🌐]** *Jun. 2025 - Current*
  Remote Research Intern at NTU supervised by ***Dr. Xiaojun Jia****.*

Projects

Partial List

- **Securing LLM Agents against IPI via Dynamic Auditing and Observation Sanitization** *Oct. 2025 - Jan. 2025*
  - Designed a dynamic oversight framework to mitigate Indirect Prompt Injection (IPI) by integrating a tri-tier tool risk stratification mechanism that balances security overhead with operational autonomy;
  - Developed a dual-verification auditing module consisting of a Trajectory Validator and a Parametric Verifier to enforce semantic alignment between agent actions and user instructions, effectively neutralizing runtime manipulations;
  - Engineered an LLM-based Observation Curator using semantic re-synthesis to strip malicious instructions from untrusted tool outputs while maintaining high-fidelity information for task completion;
  - Validated the framework on the AgentDojo benchmark, achieving a near 0% ASR while maintaining superior Benign Utility (BU) across multiple domains; paper submitted to ICML 2026 (**first author**).

- **CC-BOS: Jailbreak Optimization Framework for LLMs under Classical Chinese Contexts***Jun. 2025 - Sep. 2025*
  - Proposed the first adversarial prompt generation method leveraging Classical Chinese contexts, revealing security vulnerabilities of LLMs in cross-lingual settings;
  - Constructed an eight-dimensional strategy space and designed a bio-inspired optimization algorithm (Fruit Fly Search) to enable automated jailbreak prompt generation in black-box scenarios;
  - Developed a two-stage translation module to ensure consistency and reliability of cross-lingual evaluation;
  - Validated on multiple mainstream LLMs, achieving higher success rates and efficiency than existing methods; paper accepted to ICLR 2026 (**first author**).

- **ECIDS: A Lightweight Semantic Packet-Level Intrusion Detection System** *Apr. 2025 - Sept. 2025*
  - Leverages Word2Vec to convert packet byte sequences into static semantic feature vectors, combined with a lightweight classifier for real-time network traffic analysis, meeting the low-latency requirements of UAV D2G (Drone-to-Ground) communication scenarios;
  - Proposes EC-SJT (Supervised Joint Training) and EPC (Contrastive Projection via Contrastive Loss) to enhance semantic-classification alignment;
  - ECIDS is deployed as a proxy on the PX4 flight control system, and MAVLink protocol traffic is monitored in real time through QGroundControl (QGC) to verify the feasibility of edge deployment;
  - A manuscript has been completed and is currently being prepared for submission (**first author**).

- **BERTector: LLM-based Intrusion Detection System** *Mar. 2025 - Sept. 2025*
  - Conducted fine-tuning of a BERT-based model on network traffic data to enable accurate detection of potential malicious activities.
  - Developed the NSS-Tokenizer, which preserves semantic structure in continuous traffic data, enhancing tokenization efficiency and model performance;
  - Incorporated Low-Rank Adaptation (LoRA) technology to reduce model parameter size and improve fine-tuning efficiency, making the model lightweight and easy to deploy;
  - Completed the writing of a related academic paper (**co-first author**,paper).

- **Vulnerability Reproduction Platform** *Sep. 2024 - Jun. 2025*

- Designed and developed a vulnerability testing platform with Python, Django, and Docker, integrating detection, analysis, and exploit verification capabilities;
- Implemented containerization for dynamic resource scheduling and environment isolation, improving system security and resource utilization;
- Built user-facing features with Vue and Element UI, completing platform, user, and tool management modules with full container lifecycle support;
- Integrated security toolkit and EXP/POC resources, providing centralized management and contributing to efficient vulnerability verification; obtained a software copyright.

Work Experiences

- **Nanyang Technological University [🌐]** *Jun. 2025 - Current*

  Remote Research Intern supervised by ***Dr. Xiaojun Jia****.*

  - Actively involved in research and discussions on AI safety, with a particular focus on LLMs security and agent security.

Honours and Awards

Selected Awards

- **Competition Awards**
  - Excellence Award, *Works Contest of the National College Student Information Security Contest.* *Aug. 2025*
  - Second Prize, *National University Mathematics Competition, Jiangsu Province* *Jan. 2024*
  - Third Prize, *"TIPDM CUP" Data Mining Challenge*, Online *Jun. 2023*
  - Second Prize, *"TIPDM CUP" Data Mining Challenge, Jiangsu Province*, Online *Jun. 2023*
- **School Honours**
  - Shuangli Scholarship, NJUST (Top 3%) *Sep. 2024*
  - Merit Student, NJUST *Nov. 2024*
  - Second-Class Excellent Student Scholarship, NJUST(Top 10%) *Mar. 2024*
  - Third Prize, Outstanding Student Scholarship, NJUST (Top 15%) *Mar. 2023, Sep. 2023, Sep. 2024, Mar. 2025*

Panorama

Partial List

- **Students' Activities**
  - Academic Guide, Academic Guidance Center, NJUST *Sept. 2024 - Feb. 2025*
    - ∗ Academic Peer Mentor
  - Football Team of the School of Cyberspace Security, NJUST *Mar. 2023 - Present*
    - ∗ Captain