Qiang Hu

Tel: (086) 13874700355; E-mail: huq2021@mail.sustech.edu.cn

EDUCATION

Southern University of Science and Technology, ShenZhen, China — Sep. 2021-Present
BS in Data Science and Big Data Technology (*Department of STAT-DS*) — Expected in June 2025
GPA: 3.93/4.00(1/53) [Transcript]
Research Interests: Trustworthy AI, LLMs, Efficient AI [Google Scholar]

KNOWLEDGE BACKGROUND

***Math*:** Mathematical Analysis (H), Advanced Linear Algebra (H), Discrete Mathematics, ODE
***Statistics*:** Probability Theory, Mathematical Statistics, Operational Research and Optimization, Statistical Learning, Statistical Linear Model, Time Series, Multivariate Statistical Analysis
***Computer Science*:** Introduction to Computer Programing (Java), Data Structure and Algorithm Analysis, Principles of Databases Systems, Artificial Intelligence, Computer Vision, Advanced NLP
Data Science: Distributed Storage and Parallel Computing (Hadoop), Big Data Analysis Software and Application (Spark), Data Science Practice, Statistical Data Analysis (SAS)

WORKING PAPERS

1. **Hu, Q.***, Zhang, H.*, & Wei, H. (2024). Defending Membership Inference Attacks via Privacy-aware Sparsity Tuning. *arXiv preprint (Submitted to ICML2025)*
2. Zhang, H.*, Gao, H.*, **Hu, Q**.*, Chen, G., Yang, L., Jing, B., ... & Yang, L. (2024). ChineseSafe: A Chinese Benchmark for Evaluating Safety in Large Language Models. *arXiv preprint*

RESEARCH EXPERIENCE

Intern in Machine Learning Group of Hongxin Wei, SUSTech — Jul 2023-Present

Vulnerability Disparity Against MIA, — *Jul 2023-Mar 2024*

- Investigated the vulnerability disparities of different data instances to MIA within the same dataset.
- Identified efficient approach to evaluate instance-level data privacy vulnerability in MIA scenarios.
- Explored several training-time mitigation strategies to enhance existing MI defenses.

ICL for LLM Content Risk Control (Collaborative project with Deepexi**)**, — *Mar 2024-Jul 2024*

Co-first Author

- Established a benchmark for LLM safety in domestic scenarios, including constructing datasets through web scraping techniques.
- Evaluated existing LLMs on our benchmark, revealing various safety risks in Chinese scenarios.
- Trained a judge model via WarmupICL for Content Risk Control based on the established dataset;

Defending Membership Inference Attacks via Privacy-aware Sparsity Tuning, *Jul 2024-Oct 2024*

First Co-first Author (Submitted to ICML 2025)[paper]

- Investigated membership inference in ML from a parameter perspective and discovered that privacy risk is only related to a small fraction of model parameters.
- Proposed a privacy-aware tuning method that improves the utility-privacy trade-off of existing defenses, achieving state-of-the-art performance.

ACTIVITIES & AWARDS

Provincial(HuNan) first prize in CPhO 2020(Chinese Physics Olympiad) — Oct. 2020
Excellent Student Scholarship, SUSTech — 2022, 2023, 2024
Provincial(GuangDong) third award in CUMCM 2022
(China Undergraduate Mathematical Contest in Modeling) — Sep. 2022
Provincial(GuangDong) second award in CMC 2023(Chinese Mathematics Competitions) — Oct. 2023
Outstanding Volunteer of the School of Science — 2023

LANGUAGES & SKILLS

Languages: Mandarin (Native), English（TOEFL iBT: 85）
Programming Languages: Python, Bash, SQL, R, SAS, Spark(including kafka&Ray), Java, Matlab;