Guanyu Hou

☎ +86 19934322578 | @ hou.guanyu@student.zy.cdut.edu.cn | 🔗 Homepage

Education

Chengdu University of Technology **2021.9 - 2025.7**

Software Engineering Oxford Brookes College

First Class Degree

The University of Manchester **Approaching enrollment**

Master of Science in Artificial Intelligence

Publications

Watch Out for Your Guidance on Generation! Exploring Conditional Backdoor Attacks against Large Language Models **2024.11 - online**

The 39th Annual AAAI Conference on Artificial Intelligence (CCF-A) (Oral, top 5%)

🔗 https://ojs.aaai.org/index.php/AAAI/article/view/34819

Jiaming He, Wenbo Jiang, **Guanyu Hou**, Wenshu Fan, Rui Zhang, Hongwei Li

PRESS: Defending Privacy in Retrieval-Augmented Generation via Embedding Space Shifting **2024.9 - online**

2025 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2025) (CCF-B)

🔗 https://ieeexplore.ieee.org/document/10887843

Jiaming He*, Cheng Liu*, **Guanyu Hou***, Wenbo Jiang, Jiachen Li (* equal contribution)

Weaponizing Tokens: Backdooring Text-to-Image Generation via Token Remapping **2025.3- Accepted**

IEEE International Conference on Multimedia&Expo 2025 (ICME 2025) (CCF-B)

Jiaming He, Wenbo Jiang, **Guanyu Hou**, Qiyang Song, Guo Ji and Hongwei Li

Data Stealing Attacks against Large Language Models via Backdooring **2024.7- Online**

Electronics (JCR-Q2)

🔗 https://www.mdpi.com/2079-9292/13/14/2858

Jiaming He, **Guanyu Hou*** , Xinyue Jia, Yangyang Chen, Wenqi Liao, Yinhang Zhou, Rang Zhou (* Corresponding Author)

Embedding Based Sensitive Element Injection against Text-to-Image Generative Models **2024.4 - online**

2024 9th International Conference on Intelligent Computing and Signal Processing (ICSP 2024)

🔗 https://ieeexplore.ieee.org/document/10743442

Benrui Jiang, Kan Chen, **Guanyu Hou**, Xiying Chen, Jiaming He (Equal contribution)

Evaluating Robustness of Large Audio Language Models to Audio Injection: An Empirical Study **2025.5 - Under Review**

The 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP 2025) (CCF-B)

Guanyu Hou, Jiaming He, Yinhang Zhou, Ji Guo, Yitong Qiao, Rui Zhang, Wenbo Jiang

When Hallucinated Concepts Cross Modals: Unveiling Backdoor Vulnerability in Multi-modal In-context Learning **2025.2- Under Review**

The 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP 2025) (CCF-B)

Jiaming He, Yitong Qiao, **Guanyu Hou**, Wenbo Jiang, Zihan Wang, Qiyang Song, Hongwei Li

Language Skills

English

IELTS 7.0

Tool Skills

Programme Languages

Python, Java, C/C++

Machine Learning Frameworks

Pytorch, Sklearn

Databases

MySQL

Tools

Git, JIRA