

导师简介：刘勇，男，博士，讲师，硕士研究生导师。主要从事人工智能安全、生成模型水印、大模型内容安全等方面的研究工作。近年来，主持湖北省自然科学基金青年项目1项、十堰市重点实验室开放基金项目1项，参与国家重点研发计划项目1项、国家自然科学基金面上项目1项、腾讯犀牛鸟基金项目1项，并在国内外重要期刊和会议上共发表学术论文15篇。



研究方向：人工智能安全、生成模型水印、大模型内容安全等。

教授课程：人工智能语言及算法、自然语言处理

邮 箱：liuyong@huat.edu.cn

工作经历：

1. 湖北汽车工业学院 讲 师 2024.10-至今

近年来主要科研项目：

[1] 湖北省科技厅，基于频域可解释的神经网络模型水印研究，湖北省自然科学基金青年项目，2025-03-01至2027-02-28，**主持**

[2] 湖北汽车工业学院，黑盒神经网络鲁棒水印研究，电磁感应与节能技术十堰市重点实验室开放基金项目，2024-06 至 2026-06，**主持**

[3] 湖北汽车工业学院，面向黑盒场景下神经网络隐蔽水印研究，博士科研启动项目，2024-12至2028-12，**主持**

[4] 国家自然科学基金委员会，数字文创资产发行与流通应用示范，国家重点研发计划项目，2023-11至2026-10，**参与**

[5] 国家自然科学基金委员会，以神经网络模型为载体的安全隐写研究，国家自然科学基金面上项目，2024-01至2027-12，**参与**

[6] 腾讯公司和中国计算机学会，多媒体水印及取证新方法研究，CCF-腾讯犀牛鸟基金项目，**参与**

近年来代表性论文：

- [1] **Liu Yong**, Wu Hanzhou, Zhang Xinpeng. Robust and Imperceptible Black-box DNN Watermarking Based on Fourier Perturbation Analysis and Frequency Sensitivity Clustering[J]. IEEE Transactions on Dependable and Secure Computing, 2024, 21(6): 5766 - 5780. (CCF A, 网络与信息安全领域顶刊)
- [2] **Liu Yong**, Li Chen, Wang Zichi, Wu Hanzhou, Zhang Xinpeng. Transferable Adversarial Attack Based on Sensitive Perturbation Analysis in Frequency Domain[J]. Information Sciences, 2024, 678: 120971. (CCF B, 中科院一区Top)
- [3] **Liu Yong**, Zhang Li, Wu Hanzhou, Wang Zichi, Zhang Xinpeng. Reducing High-Frequency Artifacts for Generative Model Watermarking via Wavelet Transform[J]. IEEE Internet of Things Journal, 2024, 11(10): 18503-18515. (CCF C, 中科院一区Top)
- [4] Zhang Li, **Liu Yong**, Zhang Xinpeng, Wu Hanzhou. Suppressing High-Frequency Artifacts for Generative Model Watermarking by Anti-Aliasing[C]//Proceedings of the 2024 ACM Workshop on Information Hiding and Multimedia Security. 2024: 223-234. (CCF C, EI)
- [5] Zhang Li, **Liu Yong**, Liu Shaoteng, Yang Tianshu, Wang Yexin, Zhang Xinpeng, Wu Hanzhou. Generative Model Watermarking Based on Human Visual System[C]//International Forum on Digital TV and Wireless Multimedia Communications. Singapore: Springer Nature Singapore, 2022: 136-149. (EI)