Yichi Zhang

Zhejiang University • yichizhang@zju.edu.cn • (86) 13884308849

Education

Zhejiang University Hangzhou, China
B.S. in Information Engineering (Major). Aug. 2021 - Now

- GPA: 3.93/4.00. Minor in Computer Science and Technology.
- Performed well in the following courses: Calculus, Probability Theory, Complex Analysis, Computer Systems, Data Structures, Object-Oriented Programming, Machine Learning
- Third-class Scholarship, Zhejiang University

Research Experience

Zhejiang Lab Hangzhou, China
Research Intern May. 2023 - Mar.2024

- Collaborated with Prof. Xiaogang Xu to design a novel, high-accuracy, and straightforward feature for identifying outputs from generative AI, with a particular emphasis on images produced by the Diffusion Model.
- Developed a thorough grasp of AI Generated Content (AIGC) and diverse generative AI models, including but not limited to Diffusion Models and Generative Adversarial Networks, mastering their underlying principles and practical applications.
- Authored a scholarly paper titled *Diffusion Noise Feature: Accurate and Fast Generated Image Detection*.

NESA Lab, Zhejiang University Hangzhou, China
Research Intern May. 2023 - Now

- Under the mentorship of Prof. Shouling Ji, dedicated to exploring privacy and security challenges within the realm of machine learning.
- Gained proficiency in the application of backdoor attack and defense strategies across supervised and unsupervised learning environments. Additionally, delved into backdoor techniques specific to federated learning systems, striving to develop an efficient yet uncomplicated backdoor mechanism.
- Acquired knowledge on the robustness of foundational models in Large Language Models (LLMs) and Large Vision Models (LVMs), as well as the privacy and security considerations for these models.

Skills

Programming Languages: Proficient in Python, C++, and Matlab, with a strong foundation in algorithm development and data manipulation.
Scientific Writing: Skilled in using LaTeX for producing high-quality technical documents and research papers, ensuring clear and professional presentation of complex information.

Research Interests

- Trustworthy AI
- Generative AI
- Machine Learning