

KEYAN ZHANG

zhangky03@gmail.com ◇ Shanghai, China ◇ github.com/zhangky11

EDUCATION

Shanghai Jiao Tong University (SJTU)

Bachelor of Engineering in Information Security

Sept. 2020 - June 2024 (expected)

Shanghai, China

- **Major GPA:** 3.91/4.30, **Rank:** 6/86
- **Scholarship:** Shao Qiu Innovation Scholarship (Top 1%), Undergraduate Excellence Scholarship (Top 5%)
- **A/A+ Courses:** Data Structure, Principle of Databases, Computer Architecture, Deep Learning and 40 others.

University of Texas at Austin (UT Austin)

Exchange Program in Electrical and Computer Engineering

Aug. 2023 - Dec. 2023 (expected)

Austin, Texas, USA

- **Courses:** ADV Computer Vision (Grad-level), Software Engineering and Design, Information & Cryptography.

PUBLICATIONS

- Lu Chen, Siyu Lou, **Keyan Zhang**, et al. “*HarsanyiNet: Computing Accurate Shapley Values in a Single Forward Propagation*”. The 40th International Conference on Machine Learning (ICML), 2023
- Haodong Zhao, **Keyan Zhang**, Wenjing Fang, et al. “*Safety of Split Learning: A Survey*”. CHINESE JOURNAL OF COMPUTERS. (*Under Review*)

RESEARCH EXPERIENCE

Explainable AI: Computing Shapley Value in a Single Forward Propagation

Jan. 2022 - May 2023

Advised by Prof. Quanshi Zhang, John Hopcroft Center for Computer Science, SJTU

- Collaborated to propose an innovative neural network to calculate *Shapley* values in a single forward propagation by redistributing the *Harsanyi Interactions*, facilitating the attribution of inputs in order to explain the black box.
- Reduced the error to just **10%** compared to the state-of-the-art (SOTA) *Shapely* value methods and improved time complexity from $O(2^n)$ to $O(1)$, while maintaining its equivalence to the exact *Shapley* value.
- Co-authored a research paper accepted by ICML 2023 as the third author.

Diffusion Watermark: Identifying Images Generated by Stable Diffusion

July 2023 - Present


Advised by Prof. Qiang Liu, Statistical Learning & AI Group, UT Austin

- Designed a semantic-level watermark implementation method to differentiate generated images, modifying *Stable Diffusion* under the guidance of *BLIP* captioning method and *CLIP* score function.
- Achieved **90%** accuracy with strong robustness against adversarial inputs, protecting *Intellectual Property* (IP).

PROFESSIONAL EXPERIENCE

Alibaba Ant Group | Machine Learning Engineer Intern | Python, Pytorch, Tensorflow

Oct. 2022 - Sept. 2023

- Conducted a survey for submission, summarizing existing Split Learning attacks and defense techniques.
- Contributed to *SecretFlow* , an open source framework for Privacy-Preserving Machine Learning, implementing two different kinds of Label Leakage Attacks for Split Learning, achieving an accuracy rate of over **90%**.

Dbappsecurity | Cybersecurity Engineer Intern | SQL, HTML, JavaScript

July 2022 - Aug. 2022

- Deployed *Nessus*-like vulnerability scanning tools and performed attack tests on *Cyber Ranges* using *Metasploit* platform, and standard network attack methods like *SQL injection*, *XSS*, and *DoS*.
- Utilized tools such as *iptables* to build firewalls, employed the *Snort* intrusion detection system to explore external attack patterns and set up *bastion hosts* to monitor possible internal security threats.

ENTREPRENEURSHIP EXPERIENCE

Bright Eyes: AI Multimodal Diagnosis System for Orbital Disease

Mar. 2022 - Feb. 2023

Advised by Prof. Huifang Zhou, School of Medicine, SJTU

- Developed solutions for early diagnosis of orbital diseases with *Multimodal Machine Learning*, attaining an unmatched accuracy of **93%**, which exceeded human experts using only 2D images for diagnosis.
- Built a web application for orbital disease detection based on our algorithm using *Vue.js* and *Django*, which provided services to over **1.5k** people (36 patients diagnosed and treated on time).
- Won the **Golden Award** (Top 0.005% nation-wide) of the *China College Students “Internet+” Competition*, the highest award bestowed on the best university students in entrepreneurship and innovation competitions.

TEACHING EXPERIENCE

• **Teaching Assistant:** Computer Graphics, Harvard Summit for Young Leaders in China

June 2022

• **Lecturer:** Split Learning Attacks and Defenses, VolS-Young Forum, SJTU

Nov. 2022