# Guanyan Ou

📞 (+86) 134-1635-1655   @ guanyanou@outlook.com

Guanyan Ou is a fourth-year undergraduate student at Sun Yat-sen University. His research interest lies in trustworthy AI and privacy, with the overarching goal of making AI safe and fair.

## Competences & Languages

| | |
|---|---|
| **AI & Data Analysis** | Python, Pytorch; SQL; R language |
| **App Development** | Back-end: Java, Springboot; Front-end: React, JavaScript, Typescript; C# |
| **3D Modeling** | Cinema 4D |
| **Languages** | Chinese — Native |
| | English — TOEFL: 105 (R-29, L-28, S-23, W-25) (Sep/09/2023) |

## Education

| | |
|---|---|
| Expected Jul. 2024 | School of Software Engineering, **Sun Yat-sen University** |
| Sep. 2020 | Bachelor in Software Engineering, GPA: 3.8/4.0 |

## Research & Publications

**Paper on Adversarial Attacks on CV Classifiers**                    **October 2022 — October 2023**

Han Wu, **Guanyan Ou (Co-First Author)**, Weibin Wu, Zibin Zheng. "Improving Transferable Targeted Adversarial Attacks with Model Self-Enhancement." *Under Review of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024.*

- Proposed a model-enhancement-based method, outperforming state-of-the-art baselines by 12.2%.
- Improved the success rates of targeted black-box adversarial attacks toward computer vision classifiers.
- Employed unstructured network pruning and sharpness-aware knowledge distillation to improve adversarial perturbation generation.

**Patent for AI-assisted Tool**                    **July 2022 — February 2023**

Yuhong Nan, Zhaoxin Cai, **Guanyan Ou**, Zibin Zheng. "An Adversarial Method and Apparatus for Recognizing Misleading UI." *Patent CN116149533A.* [Link]

- Developed an Android app to help users identify fake UI patterns, such as fake closing buttons.
- Protected users from being misled or triggering unexpected activities by blocking part of inputs.
- Identified real button by the accessibility of Android system comparing with results from computer vision model.

**Ongoing Research on Red-teaming**                    **August 2023 — Present**

- Inducing black-box generative language models to produce harmful outputs.
- Using reinforcement learning framework and language models to generate human-comprehensible and natural prompts.

## Projects & Experiences

**RPCpico**                    **June 2022 — July 2023**

*The course project of "Computer Network"*

- Implemented a remote procedure call (RPC) program in Java, with HTTP protocol and Protobuf encoding.
- Provided load balance by a registration center, transferring request to the corresponding server.

**Blog Project**                    **November 2022 — January 2023**

*The course project of "Intermediate Training for Software Engineering"*

- Developed a multi-user blog website comprising both the back-end and front-end for users, along with a separate front-end for the management interface.

## Awards & Honors

**SYSU Collegiate Programming Contest**                    **2021 & 2022**

*Second Prize*

- A university-wide programming contest.

**SYSU Novice Collegiate Programming Contest**                    **2021**

*First Prize*

- A university-wide programming contest for students without experience in programming contests.