Aofan Liu

mailto:af.liu@stu.pku.edu.cn
Tel: 13418788132

Base Shenzhen · On Site

GitHub Page: github.com/Fab-Liu

Education Background

Xiamen University (985) Software Engineering B.Eng. 2020.09-2024.06

- Overall Ranking: Top 3%; Weighted GPA: 94.4/100; IELTS 7.0

Peking University (985) Computer Science M.Sc Until Now

- Directly admitted to Peking University for postgraduate study
- Research Interests: Large Models, Multimodal Systems, Large Model Security

Work Experience

GBA Artificial Intelligence Research Institute 2024.09-2024.11

Shenzhen | Multimodal Group | Algorithm Intern Algorithm Intern |

- Participated in research on semi-automatic annotation technology for large-scale multimodal datasets
- Studied the security defense mechanisms for bypassing visual adversarial examples in aligning LLMs (Large Language Models)
- Contributed to optimizing conditional generation probabilities on small harmful datasets and research on multimodal model security

Beijing Academy of Artificial Intelligence (BAAI) 2024.01-2024.06

Beijing | Multimodal Group | Algorithm Intern |

- Participated in research on semi-automatic annotation technology for large-scale multimodal datasets
- Studied the security defense mechanisms for bypassing visual adversarial examples in aligning LLMs
- Contributed to optimizing conditional generation probabilities on small harmful datasets and research on multimodal model security
- Contributed to model SFT (Supervised Fine-Tuning) and DPO (Direct Preference Optimization) training processes based on LLaMa Factory, helping improve model accuracy by 5% on specific tasks

Peking University V2X National Key Laboratory 2023.01-2023.10

Guangdong Province, Shenzhen | National Key Laboratory | Research Assistant |

- Independently built a fuzz database for over 60 CVEs (Common Vulnerabilities and Exposures) using AFLFuzz and LibFuzzer

Xiamen University, Computational and Data Science Laboratory 2021.09-2022.01

Fujian Province, Xiamen | Computational and Data Science Laboratory Research Assistant |

- Participated in experimental design, implementation, management, and monitoring for several research papers
- Assisted senior researchers in drafting sections of multiple papers using LaTeX
- Coordinated remote server scheduling for the laboratory, tracked research progress of junior researchers, and summarized findings

Academic Experience

Research on NLPL Probing Tasks Based on Context Augmentation 2024.09-2024.11

- Explored enhancing the ability of language models (LLMs) in natural language understanding tasks for code by augmenting the comment sections of code
- Evaluated the impact of context augmentation on NLPL (Natural Language Processing for Programming Languages) model comprehension, testing the performance of LLMs in multimodal tasks by introducing enriched contextual information
- Utilized large open-source code repositories containing rich comments, designed tasks, and assessed the effects of context augmentation
- Compared the performance differences between context augmentation and traditional methods (e.g., using only code structure or comment sections)

VisualDAN: Exposing Vulnerabilities in VLMs with Visual-Driven DAN Commands 2024.04-2024.06

- Evaluated the vulnerabilities in LLM security caused by visual input and explored the "jailbreaking" ability of visual adversarial examples
- Investigated the use of visual modality to output classic Jailbreak Prompt DAN series commands, achieving significant results
- Assessed the effects of attacks on various VLMs (e.g., MiniGPT-4, InstructBLIP, LLaVA) through experimental setups
- Conducted both manual and automated evaluations to determine the impact of adversarial examples on model outputs
- Compared the optimization loss and "jailbreaking" effects of visual versus text-based attacks, testing the effectiveness of existing defense technologies like DiffPure against visual adversarial examples

PiCo: Jailbreaking Multimodal Large Language Models via Pictorial Text and Code Instruction 2024.01-2024.04

- Investigated methods for jailbreaking aligned LLMs, including prompt injection, adversarial attacks, jailbreaks, and data poisoning
- Proposed the Toxicity and Helpfulness Evaluator, akin to F1-Score, for benchmarking and evaluating multimodal large models
- Focused on cross-modal attacks on MLLMs, particularly the security vulnerabilities of advanced models like Gemini-Pro and GPT-4
- PiCo successfully bypassed the security defenses of several advanced MLLMs, with an average attack success rate (ASR) of 56.27% on Gemini Pro Vision and 32.27% on GPT-4V

Research on Semi-Automatic Annotation Technology for Large-Scale Multi-Modal Datasets 2024.02-2024.04

- Contributed to building a promptable vision-based model capable of segmenting, recognizing, and describing any target within an image
- Developed a human-in-the-loop collaborative annotation framework based on a hybrid supervised large model, inspired by the SAM architecture
- Built a semi-automatic interactive annotation engine based on datasets like MSCoCo, CityScape, and Mapillary
- Improved annotation efficiency by 1-2 orders of magnitude and constructed a high-quality multimodal dataset of 500,000 images

AccuracyFuzz: Targeted Fuzz Testing Tool Based on FineTuned Large Language Models 2023.08-2024.01

- Developed a Transformer-based method to predict vulnerabilities at a finer granularity of the line level
- Used pre-trained CodeBERT models and self-attention mechanisms to achieve higher accuracy and efficiency
- Applied large models to conduct pattern testing of vulnerable software function locations
- This method significantly outperforms existing approaches in function-level prediction and line-level vulnerability detection, offering more precise and cost-effective vulnerability identification

Bert Sentiment Analysis: Prompting sentiment analysis based on Bert 2023.02-2023.04

- Trained and evaluated a model using the ChnSentiCorp dataset, which contains nearly 10,000 online reviews
- Solved the sentiment analysis task through a prompting method, converting the task into an MLM task using templates
- Fine-tuned the MLM head and evaluated model performance by predicting sentiment labels for reviews ("0" for negative, "1" for positive) on validation and test datasets

Competition& Project

18th "Huaqi Cup" Financial Innovation Application Competition | National First Prize 2022.06–2023.04

- Developed a catalog storage program using Solidity
- Created a custom star image generation program using HTML/CSS and JavaScript
- Contributed to the development and debugging of a deep learning program for artistic image style transfer

8th China International "Internet+" Innovation and Entrepreneurship Competition | National Third Prize 2022.04-2022.10

- Participated in the development of the business plan framework and organized business students to complete the writing of the business plan and the creation of the PPT
- Used regression analysis and weighted averages to determine the initial launch cities and national store expansion intentions for smart knee protectors
- Applied PEST and Ansoff Matrix models to analyze the potential and risks in the smart healthcare industry

Club and Organizational Experience

NASA Programming Challenge | North America | Team Leader 2022.02-2022.04

- Coordinated a team of 4 members from China, Pakistan, the UK, and India
- Wrote a 7,000+ word project description and app introduction documentation
- Developed a mobile app using Kotlin in 72 hours together with the team

厦门大学区块链协会 | 活动部 |副部长 2021.09-2023.09

- Participated in and organized blockchain-related lecture series jointly hosted by Xiamen University and the Blockchain Association
- Gained an initial understanding of the operational mechanisms of mainstream tokens
- Attended the "Blockchain + Finance" seminar on campus, discussing the application of non-fungible tokens (NFTs)

AIESEC **国际志愿者&诺丁汉大学 | 马来西亚 | 国际志愿者** 2021.08-2021.10

- Provided 20 general education English lessons for refugee children from surrounding countries
- Assisted 80+ students from around the world with homework guidance and grading
- Coordinated and scheduled the timetables for 100+ volunteers from all over the world during the event

专业技能

Programming Languages

- Proficient in Python (version 3.x)
- Experienced in developing and maintaining web applications with frameworks such as Django or Flask
- Familiar with Python standard libraries and third-party libraries/frameworks such as NumPy, Pandas, Django, Flask, etc.

Development Environment

- Familiar with Linux/Unix operating systems, including basic command-line operations and system management
- Experienced with Git for version control and team collaboration, familiar with platforms such as GitHub or GitLab, and proficient in Docker containerization of applications

Data Mining and Web Scraping

- Proficient in using the Requests library for HTTP requests
- Experienced with HTML/XML parsing using BeautifulSoup or LXML
- Familiar with JavaScript-rendered pages and using Selenium for data scraping
- Able to store scraped data in databases such as SQLite, MySQL, MongoDB, etc.

获奖经历

- Citi Bank Cup Financial Application Innovation Competition | National First Prize 2023.02-2023.06
- Mathematical Contest in Modeling (MCM/ICM) | National First Prize 2023.02-2023.02
- Higher Education Press Cup National Mathematical Modeling Competition | National Second Prize 2022.11-2022.11
- 8th China International "Internet+" Innovation and Entrepreneurship Competition | National Third Prize 2022.04-2022.10
- 7th China International "Internet+" Innovation and Entrepreneurship Competition | National Silver Prize 2021.07-2021.10

技能与特长

Language Proficiency：Chinese (native); English (IELTS 7.0);

Hobbies and Interests：Rock climbing, scuba diving, writing, video editing (PR, CapCut)