

EDUCATION

- **Bachelor of Engineering in AI: Systems & Technology** *Sept 2021 - May 2025 (Expected)*
The Chinese University of Hong Kong CGPA: 3.644
 - Dean's List 2021-2022
 - Dean's List 2023-2024
 - ELITE Stream Scholarship 2022-2023
 - Relevant Coursework: Fundamentals of Machine Learning, Introduction to Computer Systems, Distributed and Parallel Computing, Foundations of Optimization, Foundation of Applied Deep Learning.

PERSONAL PROJECTS

- **OASIS** *Nov 2022 - Feb 2023*
A metamorphic testing framework to validate content moderation software, focusing on toxic images.
 - Conducted a preliminary study on over 5,000 real-world image messages and summarised over 20 image transform rules to hide toxic content.
 - Designed a comprehensive testing framework for textual toxic content spread via images, whose results can effortlessly evade software moderation.
- **Semantic Fusion: DUO** *Jul - Nov 2022*
A practical content moderation software testing tool that helps to improve software moderation ability.
 - Developed a general, effective methodology and tool for testing multimedia content moderation software.
 - Successfully attacked 7 popular softwares through 1,200 pieces of data, measurably indicating their weaknesses in multimedia content moderation.

EXPERIENCE

- **UG Summer Research Internship** *Jun - Aug 2023*
The Chinese University of Hong Kong, Hong Kong Offline
 - Project Name: Multilingual Safety of Large Language Model
 - Built a multilingual safety benchmark containing more than 26,000 pieces of data and 10 different languages for testing large language models' multilingual safety ability.
- **UG Summer Research Internship** *Jun - Aug 2024*
The Chinese University of Hong Kong, Hong Kong Offline
 - Project Title: Automatically Detecting Audio-Visual Inconsistencies in 3D Software Systems
 - Propose an automatic tool to detect audio and video inconsistencies in software, which utilizes LMMs and LLMs to extract information and analysis.

TECHNICAL SKILLS

Programming Languages: C, Python, Java, Javascript, HTML, CSS, PHP
Libraries: Python Libraries(Pandas, PyTorch, NumPy, etc.), OpenMP, cuda

PUBLICATIONS

- **All Languages Matter: On the Multilingual Safety of LLMs**
Wang, Wenxuan, Zhaopeng Tu, Chang Chen, Youliang Yuan, Jen-tse Hung, Wenxiang Jiao, and Michael Lyu
The 62nd Annual Meeting of the Association for Computational Linguistics, 2024
- **Validating Multimedia Content Moderation Software via Semantic Fusion**
Wang, Wenxuan, Jingyuan Huang, Chang Chen, Jiazhen Gu, Jianping Zhang, Weibin Wu, Pinjia He, and Michael Lyu
Proceedings of the 32nd ACM SIGSOFT International Symposium on Software Testing and Analysis, 2023
- **An Image is Worth a Thousand Toxic Words: A Metamorphic Testing Framework for Content Moderation Software**
Wang, Wenxuan, Jingyuan Huang, Jen-tse Huang, Chang Chen, Jiazhen Gu, Pinjia He, and Michael R Lyu
2023 38th IEEE/ACM International Conference on Automated Software Engineering (ASE), 2023