

JINGYUAN HUANG

Phone Number: (762) 205-4534 • Email: jh19696@uga.edu • Website: harveyyellow.github.io

EDUCATION

- **Ph.D. in Computer Science**, University of Georgia **Aug 2025 – Present**
Research Focus: Multimodal LLMs, Trustworthy LLM
Supervisor: Prof. Ninghao Liu
- **B.S. in Artificial Intelligence**, The Chinese University of Hong Kong **Sep 2021 – Jun 2025**

RESEARCH INTERESTS

Multimodal LLMs, Computational Social Science, Model Bias & Fairness

PUBLICATIONS

AI Sees Your Location—But With A Bias Toward The Wealthy World.

Jingyuan Huang[†], Jen-tse Huang[†], Ziyi Liu, Xiaoyuan Liu, Wenxuan Wang, Jieyu Zhao.

In Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP). (2025).

Not All Countries Celebrate Thanksgiving: On the Cultural Dominance in Large Language Models.

Wenxuan Wang, Wenxiang Jiao, **Jingyuan Huang**, Ruyi Dai, Jen-tse Huang, Zhaopeng Tu, Michael R. Lyu.

In Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (ACL). (2024).

A Picture is Worth a Thousand Toxic Words: A Metamorphic Testing Framework for Content Moderation Software.

Wenxuan Wang, **Jingyuan Huang**, Chang Chen, Pinjia He, Jiazhen Gu, Michael R. Lyu.

In Proceedings of the 38th IEEE/ACM International Conference on Automated Software Engineering (ASE). (2023).

Validating Multimedia Content Moderation Software via Semantic Fusion.

Wenxuan Wang, **Jingyuan Huang**, Chang Chen, Jiazhen Gu, Jianping Zhang, Weibin Wu, Pinjia He, Michael R. Lyu.

In Proceedings of the 32nd ACM SIGSOFT International Symposium on Software Testing and Analysis (ISSTA). (2023).

A Spectrum Evaluation Benchmark for Medical Multi-Modal Large Language Models.

Jie Liu, Wenxuan Wang, Yihang Su, **Jingyuan Huang**, Wenting Chen, Yudi Zhang, Cheng-Yi Li, Kao-Jung Chang, Xiaohan Xin, Linlin Shen, Michael R. Lyu.

In Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (ACL). (2025).

([†] denotes equal contribution)

RESEARCH EXPERIENCE

- **Research Collaborator**, University of Southern California (USC) **May 2024 – May 2025**
(Remote, with Prof. Jieyu Zhao)
 - Co-led research quantifying significant geopolitical and economic biases in the location-recognition capabilities of state-of-the-art VLMs.
 - Spearheaded the core research direction, designed the end-to-end experimental pipeline, and executed all analyses, securing a co-first author EMNLP 2025 publication.
- **Research Assistant**, ARISE Lab, CUHK **May 2022 – May 2025**
(with Prof. Michael R. Lyu)
 - Engineered a novel multimodal attack vector that bypassed content moderation systems by distributing toxic cues, achieving a 90% evasion rate.
 - Led the complete experimental lifecycle, from dataset generation to technical writing, for two top-tier conference papers (ISSTA 2023, ASE 2023) on AI safety.
 - Co-authored "Asclepius," a novel benchmark for evaluating the medical reasoning capabilities of Vision-Language Models.
 - Supervised and mentored two undergraduate students on their final year projects in LLM reliability.
- **Research Collaborator**, Tencent AI Lab **Oct 2022 – Sep 2023**

- Initiated and directed a research project investigating cultural dominance and bias in multilingual Large Language Models.
- Architected the complete experimental framework, from novel dataset curation to evaluation metric design, culminating in an ACL 2024 publication.
- Managed all primary experiments and authored the corresponding results and analysis sections of the paper.

TECHNICAL SKILLS

- **Programming Languages:** Python, C, SQL
- **AI/ML Frameworks:** PyTorch, TensorFlow, Hugging Face, scikit-learn, Pandas, NumPy
- **Developer Tools:** Git, LaTeX