Jingyuan **Huang**

1155173905@link.cuhk.edu.hk | harveyyellow.github.io | Huang Jingyuan

Education

The Chinese University of Hong Kong — *HONG KONG*

Undergraduate Student in Artificial Intelligence — *Sep '21 - Present.*

GPA: 3.6/4

Research Experience

Summer Research with Assistant Professor Jieyu Zhao (USC) — *REMOTE*

Research Collaborator — *May '24 - Present.*

- Researched bias in Vision-Language Models (VLMs), currently writing for submission to ACL in February.
- Conducted the majority of the literature review and all experimental work, contributing ideas that were incorporated into the research.
- Designated as one of two co-first authors on the paper.

ARISE Lab (CUHK) — *HONG KONG*

Research Assistant — *May '22 - Present.*

- Evaluating the Safety and Reliability of Multi-modal models and Large language models
- Developed the core idea of the first and second papers: distributing information across different modalities to circumvent AI moderation systems, revealing that only 10% of processed toxic content was flagged by AI in 2021.
- Led the majority of the experimental design and execution, dataset collection and generation, and authored the experimental section of the paper.
- Collaborated on creating a benchmark to evaluate the medical capabilities of Vision-Language Models (VLMs).
- Currently leading two students in their final year projects within the lab.

Tencent AI Lab — *HONG KONG*

Research Collaborator — *Oct '22 - Sep '23.*

- Collaborated closely with researchers to produce a top-tier conference paper.
- Proposed the core idea: investigating the cultural biases of large language models across different languages.
- Designed the experiment, collecting all relevant datasets and establishing evaluation metrics.
- Completed the majority of the experimental work and authored the experimental and result section of the paper.

Publication

Not All Countries Celebrate Thanksgiving: On the Cultural Dominance in Large Language Models — *ACL 2024*

Wenxuan Wang, Wenxiang Jiao, **Jingyuan Huang**, Ruyi Dai, Jen-tse Huang, Zhaopeng Tu, Michael R. Lyu — *(CCF A)*

The Annual Meeting of the Association for Computational Linguistics

A Picture is Worth a Thousand Toxic Words: A Metamorphic Testing Framework for Content Moderation Software — *ASE 2023*

Wenxuan Wang, **Jingyuan Huang**, Chang Chen, Pinjia He, Jiazhen Gu, Michael R. Lyu — *(CCF A)*

The IEEE/ACM International Conference on Automated Software Engineering

Validating Multimedia Content Moderation Software via Semantic Fusion — *ISSTA 2023*

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

The ACM SIGSOFT International Symposium on Software Testing and Analysis

Asclepius: A Spectrum Evaluation Benchmark for Medical Multi-Modal Large Language Models — *Under Review*

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

LLMs as GeoGuessr Masters: Exceptional Performance, Hidden Biases, and Privacy Risks — *Ongoing*

Jingyuan Huang*, Jen-tse Huang*, Ziyi Liu, Xiaoyuan Liu, Wenxuan Wang, Jieyu Zhao

Skills

Computer Skills, Machine Learning, Deep Learning, Python, C, SQL

Research Skills, Self-motivated; skilled at summarizing papers and understanding key concepts.