JEN-TSE HUANG

 https://github.com/penguinnnnn   https://penguinnnnn.github.io/
 Sha Tin, Hong Kong   (852) 65757071   jthuang@cse.cuhk.edu.hk

EDUCATION

The Chinese University of Hong Kong   *Aug. 2020 - Present*
Ph.D. in Computer Science
– A member of Automated Reliable Intelligent Software Engineering (ARISE) Lab
– Supervised by Prof. Michael R. Lyu

Peking University   *Sep. 2015 - Jun. 2019*
B.Sc. in Computer Science
– A member of Yuanpei College

EXPERIENCE

Tencent AI Lab, Shenzhen   *Mar. 2022 - Oct. 2023*
Research Intern
– Supervisor: Dr. Wenxiang Jiao and Dr. Zhaopeng Tu
– Research Topic: Evaluation of Large Language Models

CSE Dept., The Chinese University of Hong Kong, Hong Kong   *Feb. 2020 - Jul. 2020*
Research Assistant
– Supervisor: Prof. Michael R. Lyu
– Research Topic: Interpretability, Reliability and Robustness in Deep Learning Models

SenseTime Research, Beijing   *Feb. 2018 - Jun. 2019*
Research Intern
– Supervisors: Dr. Chen Qian and Prof. Bolei Zhou
– Research Topics: Pose Estimation; Interpretability in Generative algorithms

PREPRINTS

1. **Jen-tse Huang**, Wenxuan Wang, Man Ho Lam, Eric John Li, Wenxiang Jiao, Michael R. Lyu. Revisiting the Reliability of Psychological Scales on Large Language Models. *arXiv Preprint: 2305.19926*. 2023.

2. **Jen-tse Huang**, Man Ho Lam, Eric John Li, Shujie Ren, Wenxuan Wang, Wenxiang Jiao, Zhaopeng Tu, Michael R. Lyu. Emotionally Numb or Empathetic? Evaluating How LLMs Feel Using EmotionBench. *arXiv Preprint: 2308.03656*. 2023.

3. Wenxuan Wang, Wenxiang Jiao, Jingyuan Huang, Ruyi Dai, **Jen-tse Huang**, Zhaopeng Tu, Michael R. Lyu. Not All Countries Celebrate Thanksgiving: On the Cultural Dominance in Large Language Models. *arXiv Preprint: 2310.12481*. 2023.

4. Wenxuan Wang, Zhaopeng Tu, Chang Chen, Youliang Yuan, **Jen-tse Huang**, Wenxiang Jiao, Michael R. Lyu. All Languages Matter: On the Multilingual Safety of Large Language Models. *arXiv Preprint: 2310.00905*. 2023.

5. Tian Liang, Zhiwei He, **Jen-tse Huang**, Wenxuan Wang, Wenxiang Jiao, Rui Wang, Yujiu Yang, Zhaopeng Tu, Shuming Shi, Xing Wang. Leveraging Word Guessing Games to Assess the Intelligence of Large Language Models. *arXiv Preprint: 2310.20499*. 2023.

6. Yuxuan Wan, Wenxuan Wang, Wenxiang Jiao, Yiliu Yang, Youliang Yuan, **Jen-tse Huang**, Pinjia He, Michael R. Lyu. A $\wedge$ B $\Leftrightarrow$ B $\wedge$ A: Triggering Logical Reasoning Failures in Large Language Models. *arXiv Preprint: 2401.00757*. 2024.

7. Wenxuan Wang, Juluan Shi, Zhaopeng Tu, Youliang Yuan, **Jen-tse Huang**, Wenxiang Jiao, Michael R. Lyu. The Earth is Flat? Unveiling Factual Errors in Large Language Models. *arXiv Preprint: 2401.00761*. 2024.

8. Wenxuan Wang, Haonan Bai, Yuxuan Wan, **Jen-tse Huang**, Youliang Yuan, Haoyi Qiu, Nanyun Peng, Michael R. Lyu. New Job, New Gender? Measuring the Social Bias in Image Generation Models. *arXiv Preprint: 2401.00763*. 2024.

9. Wenxiang Jiao, Wenxuan Wang, **Jen-tse Huang**, Xing Wang, Zhaopeng Tu. Is ChatGPT A Good Translator? Yes With GPT-4 As The Engine. *arXiv Preprints: 2301.08745*. 2023.

PUBLICATIONS

1. **Jen-tse Huang**, Wenxuan Wang, Eric John Li, Man Ho Lam, Shujie Ren, Youliang Yuan, Wenxiang Jiao, Zhaopeng Tu, Michael R. Lyu. Who is ChatGPT? Benchmarking LLMs' Psychological Portrayal Using PsychoBench. *In Proceedings of the Twelfth International Conference on Learning Representations (ICLR)*. 2024.

2. Youliang Yuan, Wenxiang Jiao, Wenxuan Wang, **Jen-tse Huang**, Pinjia He, Shuming Shi, Zhaopeng Tu. GPT-4 Is Too Smart To Be Safe: Stealthy Chat with LLMs via Cipher. *In Proceedings of the Twelfth International Conference on Learning Representations (ICLR)*. 2024.

3. Wenxiang Jiao, **Jen-tse Huang**, Wenxuan Wang, Zhiwei He, Tian Liang, Xing Wang, Shuming Shi, Zhaopeng Tu. ParroT: Translating During Chat Using Large Language Models. *In Findings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2023.

4. Wenxuan Wang, Jingyuan Huang, **Jen-tse Huang**, Chang Chen, Jiazhen Gu, Pinjia He, Michael R. Lyu. An Image is Worth a Thousand Toxic Words: A Metamorphic Testing Framework for Content Moderation Software. *In Proceedings of the 38th IEEE/ACM International Conference on Automated Software Engineering (ASE)*. 2023

5. Jianping Zhang, **Jen-tse Huang**, Wenxuan Wang, Yichen Li, Weibin Wu, Xiaosen Wang, Yuxin Su, Michael Lyu. Improving the Transferability of Adversarial Samples by Path-Augmented Method. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2023.

6. Wenxuan Wang, **Jen-tse Huang**, Weibin Wu, Jianping Zhang, Yizhan Huang, Shuqing Li, Pinjia He, Michael R. Lyu. MTTM: Metamorphic Testing for Textual Content Moderation Software. *In Proceedings of the 45th IEEE/ACM International Conference on Software Engineering (ICSE)*. 2023.

7. Wenxiang Jiao, Zhaopeng Tu, Jiarui Li, Wenxuan Wang, **Jen-tse Huang**, Shuming Shi. Tencent's Multilingual Machine Translation System for WMT22 Large-Scale African Languages. *In Proceedings of the Seventh Conference on Machine Translation (WMT)*. 2022.

8. **Jen-tse Huang**, Jianping Zhang, Wenxuan Wang, Pinjia He, Yuxin Su, Michael R. Lyu. AEON: A Method for Automatic Evaluation of NLP Test Cases. *In Proceedings of the 31st ACM SIGSOFT International Symposium on Software Testing and Analysis (ISSTA)*. 2022.

9. Jianping Zhang, Weibin Wu, **Jen-tse Huang**, Yizhan Huang, Wenxuan Wang, Yuxin Su, Michael R. Lyu. Improving Adversarial Transferability via Neuron Attribution-Based Attacks. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2022.

AWARD

1. **Postgraduate Studentship**, The Chinese University of Hong Kong — *2020 - 2024*

2. **Scholarship for Hong Kong, Macao, Taiwan and Overseas Chinese**, Peking University — *2018*

SERVICE

1. **Conference Reviewer**: CVPR'24, ACL'23, EMNLP'23

2. **Teaching Assistant**: Software Engineering (CSCI3100) in CUHK — *2021, 2022*

3. **Teaching Assistant**: Discrete Math for Engineers (ENGG2440A) in CUHK — *2020*