Chenhang Cui

(+86)15209187803 | chenhangcui@gmail.com | Google scholar | GitHub page | Chenhang's homepage

Education

National University of Singapore, School of Computing (NExT++ Lab) — Singapore
Ph.D. in Computer Science, Advisor: Prof. Chua Tat-Seng; with Dr. Fei Shen — *Jan. 2026 - Present*

University of Electronic Science and Technology of China — Chengdu, China
Bachelor of Yingcai Honors College — *Sept. 2021 - Jun. 2025*

The Middle School Attached To Northwestern Polytechnical University — Xi'an, China
High School Diploma — *Sept. 2018 - Jun. 2021*

- GPA: 3.98/4.00 Average Score: 90.51 Rank: 2/70 TOEFL: 97
- About me: I am very interested in large models and aspire to become an artificial intelligence scientist in the future. I enjoy making friends and in my daily life, I love running and traveling.
- Research interests: Large Vision-language Models (LVLMs): Trustworthiness in large vision-language models, (hallucination, attack, etc.), Alignment of LVLMs, RL in LVLMs; Multi-view Representation Learning: Exploring better representations for multimodal data in the real world

Publications

Conference paper (*Equal contribution)

- 1. **Chenhang Cui**, Gelei Deng, An Zhang, Jingnan Zheng, Yicong Li, Lianli Gao, Tianwei Zhang, Tat-Seng Chua. *Safe + Safe = Unsafe? Exploring How Safe Images Can Be Exploited to Jailbreak Large Vision-Language Models.* NeurIPS, 2025. [pdf]
- 2. **Chenhang Cui**, An Zhang, Yiyang Zhou, Zhaorun Chen, Gelei Deng, Huaxiu Yao, Tat-Seng Chua. *Fine-Grained Verifiers: Preference Modeling as Next-Token Prediction in Vision-Language Alignment.* ICLR, 2025. [pdf]
- 3. Peng Xia, Siwei Han, Shi Qiu, Yiyang Zhou, Zhaoyang Wang, Wenhao Zheng, Zhaorun Chen, **Chenhang Cui**, Mingyu Ding, Linjie Li, Lijuan Wang, Huaxiu Yao. *Mmie: Massive multimodal interleaved comprehension benchmark for large vision-language models.* ICLR, 2025. [pdf]
- 4. Haoqin Tu*, **Chenhang Cui***, Zijun Wang, Yiyang Zhou, Bingchen Zhao, Junlin Han, Wangchunshu Zhou, Huaxiu Yao, Cihang Xie. *How Many Unicorns Are in This Image? A Safety Evaluation Benchmark for Vision LLMs.* ECCV, 2024. [pdf]
- 5. Yiyang Zhou*, **Chenhang Cui***, Jaehong Yoon, Linjun Zhang, Zhun Deng, Chelsea Finn, Mohit Bansal, Huaxiu Yao. *Analyzing and Mitigating Object Hallucination in Large Vision-Language Models.* ICLR, 2024. [pdf]
- 6. Yiyang Zhou*, **Chenhang Cui***, Rafael Rafailov, Chelsea Finn, Huaxiu Yao. *Aligning Modalities in Vision Large Language Models via Preference Fine-tuning.* ICLR Workshop, 2024. [pdf]
- 7. Zhaorun Chen, Yichao Du, Zichen Wen, Yiyang Zhou, **Chenhang Cui**, Zhenzhen Weng, Haoqin Tu, Chaoqi Wang, Zhengwei Tong, Qinglan Huang, Canyu Chen, Qinghao Ye, Zhihong Zhu, Yuqing Zhang, Jiawei Zhou, Zhuokai Zhao, Rafael Rafailov, Chelsea Finn, Huaxiu Yao. *MJ-Bench: Is Your Multimodal Reward Model Really a Good Judge?* ICML Workshop, 2024. [pdf]
- 8. **Chenhang Cui**, Yazhou Ren, Jingyu Pu, Jiawei Li, Xiaorong Pu, Tianyi Wu, Yutao Shi, Lifang He. *A Novel Approach for Effective Multi-View Clustering with Information-Theoretic Perspective.* NeurIPS, 2023. [pdf]
- 9. **Chenhang Cui**, Yazhou Ren, Jingyu Pu, Xiaorong Pu, Lifang He. *Deep Multi-view Subspace Clustering with Anchor Graph.* IJCAI, 2023. [pdf]
- 10. Jingyu Pu, **Chenhang Cui**, Xinyue Chen, Yazhou Ren, Xiaorong Pu, Zhifeng Hao, S Yu Philip, Lifang He. *Adaptive Feature Imputation with Latent Graph for Deep Incomplete Multi-View Clustering.* AAAI, 2023. [pdf]

Arxiv

- 1. **Chenhang Cui***, Yiyang Zhou*, Xiangyu Yang, Shirley Wu, Linjun Zhang, James Zou, Huaxiu Yao. Holistic Analysis of Hallucination in Large Vision-Language Models: Bias and Interference Challenges [pdf]

Research Experience

PhD: Trustworthy and Aligned Vision-Language Models Jan. 2026 – Present

Advisor: Prof. Chua Tat-Seng; working with Dr. Fei Shen (NUS, NExT++ Lab)

- **Topic**: Hallucination, safety, and RL-based alignment for large vision-language models.

Research Intern: Safety and Multimodal RL Alignment for VLLMs Aug. 2024 – Aug. 2025

Host: Prof. An Zhang (NUS, NExT++ Lab)

- **Topic**: Safety and Multimodal RL Alignment for VLLMs
- **Description**: Focused on enhancing the safety and alignment of Vision-Language Large Models (VLLMs) by addressing issues of unsafe and misaligned multimodal content generation. The project involves evaluating existing VLLMs, developing new alignment techniques, and assessing their safety.
- **Achievements**: One paper has been accepted by ICLR, 2025. One paper has been accepted by NeurIPS, 2025.

Research Intern: Reducing Hallucination in Large Vision-Language Models May 2023 – Jan. 2024

Host: Prof. Huaxiu Yao (UNC MURGe-Lab)

- **Topic**: Reducing object hallucination problems in Large Vision-Language models
- **Description**: Aiming to reduce hallucination in existing Large Vision-Language models (e.g., LLaVa, MiniGPT-4, mPlug-owl).
- **Achievements**: Reduced hallucination by 23% without compromising text diversity in MiniGPT-4. Human and GPT assessments rank our methods as number one. Short version accepted by NeurIPS Workshop, 2023; long version accepted by ICLR. (Conference - 1)

Undergraduate Research: Representation Learning from Multi-view Perspective May 2022 – Sep. 2023

Advisor: Prof. Yazhou Ren (UESTC)

- **Topic**: Consistency and Complementary of Multimodal Representations
- **Description**: Enhancing representation learning through consistency from multimodal data, applied to downstream ML tasks.
- **Achievements**: Three papers completed (Conference - 2, 4, 5)

Services

Conference Reviewer

- ARR 2024, KDD 2024, ICLR 2025, CVPR 2025, ICLR 2026

Journal Reviewer

- Neural Networks

Honors & Awards

UESTC Undergraduate Academic Scholarship 2022, 2023, 2024

HUAWEI Scholarship 2023

The Mathematical Contest in Modeling: Honorable Mention (15%), Student Advisor 2024

SenseTime Scholarship: This award is presented to undergraduate students with outstanding contributions in the field of computer science. Only 25 students received this award in 2024. 2024

Technical Skills

Programming: Python, Latex, Git, Matlab, Shell, MarkDown, Html
Communication: Chinese (native), English (TOEFL: 97 - Reading: 24, Listening: 27, Speaking: 21, Writing: 25)