Caoyuan Ma

University of Tokyo
Graduate School of Information Science and Technology
Ph.D. Student — Multimodal LLM Safety

+81 70 8488 2313
mcy@g.ecc.u-tokyo.ac.jp
GitHub: Miles629
Google Scholar

Education

The University of Tokyo *Oct. 2025 – Present*
Ph.D. in Information Science and Technology — Research Focus: Multimodal LLM Safety

Wuhan University, School of Computer Science *Sep. 2022 – Jun. 2025*
M.S. in Computer Science and Technology — Human-centric AIGC and 3D Vision

Wuhan University, School of Computer Science *Sep. 2018 – Jul. 2022*
B.Eng. in Software Engineering — GPA: 3.87/4.00; top 1% in two academic years

Research Interests

Research Interests: Multimodal intelligence, with a focus on post-training, alignment, safety, and embodied spatial reasoning. Research portfolio

Research and Industry Experience

Multimodal Foundation Models

JD Explore Academy — Research Intern *Apr. 2026 – Present*

- Audited every sample across six benchmarks, categorized image and question patterns, and built targeted collection, cleaning, and multi-agent cross-validation pipelines that produced 7.74 million samples totaling 15.6 billion tokens.
- Improved the average score of Qwen3-VL-30B-A3B from 48.3 to 55.5; optimized the data mixture for the mainline model, enabling the specialist model to match top-tier 300B-model performance.
- Built data for ZeroBench-main, improving the specialist model by approximately 10 points during SFT; developed CoT data collection pipelines for Claude Opus and Fable to support agent training.

Shanghai AI Laboratory (Remote) — Research Intern *Sep. 2025 – Apr. 2026*

- Supported base-model experiments and evaluation for *SafeWork-R1*.
- Independently studied semi-self-evolving LLM safety; manuscript in revision.

Video Generation

Hedra (U.S., Remote) — Research Intern *2025*

- Selected and applied DMD and Self-Forcing to production video models, delaying noticeable visual degradation from approximately 8 to 30 seconds.

StepFun — Research Intern *2025*

- Led a four-person team to build a large-scale video-data curation pipeline for model training, processing more than 30 TB of video.

3D Human Reconstruction

Noitom & JD Group — Research Intern *2022 – 2024*

- Built a synchronized IMU motion-capture and multi-camera system to collect multi-view human motion data.
- Developed a NeRF-based human-centric 6D reconstruction method spanning 3D space and 3D skeleton-driven motion, resulting in the first-authored CVPR 2024 paper *HumanNeRF-SE*.

Awards and Honors

- **Scholarships:** National Scholarship (twice; 2/208), First-Class Scholarship (twice; top 5%), and SPRING GX.
- **Honor:** Outstanding Graduate of Wuhan University (top 10%).
- **Competitions:** Meritorious Winner, Mathematical Contest in Modeling; First Prize, National Undergraduate IoT Contest.

Publications

*Equal contribution. My name is shown in **bold**. Conference/journal versions supersede duplicate preprints.

1. **HumanNeRF-SE: A Simple yet Effective Approach to Animate HumanNeRF with Diverse Poses**
   Caoyuan Ma, Yu-Lun Liu, Zhixiang Wang, Wu Liu, Xinchen Liu, and Zheng Wang. *CVPR 2024.*

2. **Learning a Dynamic Neural Human via Poses Guided Dual Spaces Feature**
   Caoyuan Ma, Runqi Wang, Wu Liu, Zhiming Zhou, and Zheng Wang. *IEEE AVSS 2024.*

3. **SafeCap: Improving LVLM Safety with Image Captioning Reinforcement Learning**
   Caoyuan Ma, Liuwenpu, Weichu Xie, Tian Gu, Shilei Zhao, Lingxi Min, Shuai Dong, Yuqi Xu, Ji Zhao, Ziyue Wang, et al.. *arXiv:2608.10513, 2026.*

4. **SafeRI: Recognition and Intervention for Token-Level Safety Intervention in Large Vision Language Models**
   Caoyuan Ma*, Tian Gu*, Liuwenpu, Weichu Xie, Shuai Dong, Yuqi Xu, Ji Zhao, Ziyue Wang, et al.. *arXiv:2609.03544, 2026.*

5. **You Think, You Act: The New Task of Arbitrary Text to Motion Generation**
   Runqi Wang*, **Caoyuan Ma***, Guopeng Li*, Hanrui Xu, Yizhou Li, and Zheng Wang. *ICCV 2025.*

6. **Subjective Camera 1.0: Bridging Human Cognition and Visual Reconstruction through Sequence-Aware Sketch-Guided Diffusion**
   Haoyang Chen*, Dongfang Sun*, **Caoyuan Ma***, Shiqin Wang, Kewei Zhang, Zheng Wang, and Zhixiang Wang. *ICCV 2025.*

7. **Leader is Guided: Interactive Motion Generation via Lead-Follow Paradigm and Trajectory Guidance**
   Runqi Wang*, **Caoyuan Ma***, Jian Zhao, Hanrui Xu, Dongfang Sun, Haoyang Chen, Lin Xiong, Zheng Wang, and Xuelong Li. *ACM MM 2025.*

8. **Behave Your Motion: Habit-Preserved Cross-Category Animal Motion Transfer**
   Zhimin Zhang*, Bi'an Du*, **Caoyuan Ma**, Zheng Wang, and Wei Hu. *ACM MM 2025.*

9. **Causal Deciphering and Inpainting in Spatio-Temporal Dynamics via Diffusion Model**
   Yifan Duan, Jian Zhao, Junyuan Mao, Hao Wu, Jingyu Xu, Shiqin Wang, **Caoyuan Ma**, Kai Wang, Kun Wang, and Xuelong Li. *NeurIPS 2024.*

10. **Improving Stack Overflow Question Title Generation with Copying Enhanced CodeBERT Model and Bi-Modal Information**
    Fengji Zhang, Xiao Yu, Jacky Keung, Fuyang Li, Zhiwen Xie, Zhen Yang, **Caoyuan Ma**, and Zhimin Zhang. *Information and Software Technology, 148:106922, 2022.*

11. **Right Makes Might: Aligning Verified Hidden States Empowers RL Reasoning**
    Ziyue Wang, Aolin Yuan, Yongfu Zhu, Shuai Dong, Wu Liu, et al., including **Caoyuan Ma**. *arXiv:2606.03234, 2026.*

12. **SafeWork-R1: Coevolving Safety and Intelligence under the AI-45 Law**
    Shanghai AI Laboratory et al., including **Caoyuan Ma**. *arXiv:2507.18576, 2025.*