# HAONAN GE

Southeast University, Nanjing, China

gehaonan82@gmail.com | Haonan-Ge.github.io | Haonan-Ge | Google Scholar

## Research Interests

Large Language Model, Multi-modal modeling, MLLM reasoning, Safety in MLLM.

## Education

**Southeast University** **Nanjing, China**
*B.Eng. in Electrical & Computer Engineering* Sep 2022 – Expected Jun 2026
**GPA**: 3.72/4.0

**University of California, Irvine** **Irvine, CA**
*Exchange in Computer Science* Jun 2025 – Sep 2025
**GPA**: 4.0/4.0

## Publications

*** denotes equal contribution**

[1] **Haonan Ge**, Yiwei Wang, Kai-Wei Chang, Hang Wu, Yujun Cai. FrameMind: Frame-Interleaved Video Reasoning via Reinforcement Learning, *ICLR 2026 (under review)*. arXiv:2509.24008.

[2] **Haonan Ge**, Yiwei Wang, Ming-Hsuan Yang, Yujun Cai. MRFD: Multi-Region Fusion Decoding with Self-Consistency for Mitigating Hallucinations in LVLMs, *EMNLP 2025 Findings*. arXiv:2508.10264.

[3] Haotian Xia*, **Haonan Ge***, Junbo Zou*, Hyun Woo Choi, Xuebin Zhang, Danny Suradja, Botao Rui, Ethan Tran, Wendy Jin, Zhen Ye, Xiyang Lin, Christopher Lai, Shengjie Zhang, Junwen Miao, Shichao Chen, Rhys Tracy, Vicente Ordonez, Weining Shen, Hanjie Chen. SportR: A Benchmark for Multimodal Large Language Model Reasoning in Sports, *ICLR 2026 (under review)*.PDF.

[4] Hang Wu, Yujun Cai, **Haonan Ge**, Hongkai Chen, Ming-Hsuan Yang, Yiwei Wang. RefineShot: Rethinking Cinematography Understanding with Foundational Skill Evaluation, *ICLR 2026 (under review)*.arXiv:2510.02423.

## Research Experience

**University of California, Merced & The University of Queensland** **Merced, CA**
*Research Intern (Remote); Supervised by Prof.* ***Yujun Cai*** *&* ***Yiwei Wang*** May 2025 – Sep 2025

**Project 1:** *FrameMind: Frame-Interleaved Video Reasoning via Reinforcement Learning* (**lead**)

- Proposed **Frame-Interleaved CoT (FiCOT)**: a multi-turn loop that alternates textual reasoning with active visual perception, enabling the model to *request* video evidence during reasoning (e.g., target frames or short clips).
- Introduced **Dynamic-Resolution Frame Sampling (DRFS)** to expose the policy to temporal–spatial trade-offs; developed **DRFS-GRPO** for group-relative optimization from outcome rewards (no frame-level labels).
- Implemented a compact **agent runtime**(prompt templates, tool schema, parsing/verification, 1–3 turn cap) and ran controlled ablations on short/long-video benchmarks (e.g., MVBench/MLVU/VideoMME).
- Demonstrated strong gains on long-video benchmarks (e.g., MLVU, VideoMME) with improved efficiency and interpretability.

**Project 2:** *RefineShot: Rethinking Cinematography Understanding with Foundational Skill Evaluation*

- Refined **ShotBench** with consistent option restructuring, releasing the expanded benchmark **RefineShot** for reliable evaluation.
- Performed the first critical analysis of **ShotVL** behavior, identifying issues in reasoning consistency and instruction adherence; proposed diagnostic checks.
- Designed an extended evaluation protocol to jointly assess task accuracy and foundational skills, improving fairness and reproducibility.

**University of California, Irvine & Rice University** **Irvine, CA & Houston, TX**
*Research Intern (On-site); Supervised by Prof.* ***Weining Shen*** *&* ***Hanjie Chen*** Jun 2025 – Sep 2025
**Project:** *SportR: A Benchmark for Multimodal LLM Reasoning in Sports* (**co-led**)

- Led **badminton** CoT annotations with progressive QA hierarchy; enforced quality control and consistency.
- Designed experiments for infraction/type/penalty/explanation/tactics and explicit **grounding** tasks.
- Implemented **SFT** pipeline and **GRPO**-based **RL** training (reward: LLM-as-Judge for QA; **GIoU** for grounding).
- Conducted RL ablations and runs, validating post-SFT → RL improvements and cross-modal generalization.
- Benchmarked open-source MLLMs (Qwen-VL, LLaVA-Next, GLM-4.5V, DeepSeek-VL) with unified prompts and judge ensemble.

**The University of Queensland** **Remote**
*Research Intern (Remote); Supervised by Prof.* ***Yujun Cai*** Feb 2025 – May 2025
**Project:** *MRFD: Multi-Region Fusion Decoding with Self-Consistency for Mitigating Hallucinations in LVLMs* (EMNLP 2025 Findings; **first author**)

- Proposed a **training-free** decoding method that selects salient regions via cross-attention, generates per-region responses, and fuses them using **Jensen–Shannon Divergence**–based reliability weights.
- Designed **region-aware prompts** (CoT-inspired) and a **consistency-weighted fusion** procedure to improve factual grounding and reduce hallucinations without modifying model parameters.
- Built evaluation scripts for POPE/CHAIR/MME-Hallucination; observed consistent gains over strong decoding baselines (precision/F1 and CHAIR scores).

## Services

**Reviewer**: CVPR 2026

## Skills

**Programming**: Python, C/C++; Matlab; HTML/CSS/JS
**ML/Systems**: PyTorch, Ray, FSDP; NumPy/Pandas; Git, Docker
**Writing**: LaTeX
**OS**: Linux (Ubuntu), Windows
**Languages**: Mandarin (Native), English (Fluent)

## References

**Prof. Yujun Cai**
*Assistant Professor, The University of Queensland*
E-mail: yujun.cai@uq.edu.au
Department: Electrical Engineering and Computer Science

**Prof. Yiwei Wang**
*Assistant Professor, University of California, Merced*
E-mail: yiweiwang2@ucmerced.edu
Department: Computer Science, UC Merced NLP Lab

**Prof. Weining Shen**
*Associate Professor, University of California, Irvine*
E-mail: weinings@uci.edu
Department: Statistics

**Prof. Hanjie Chen**
*Assistant Professor, Rice University*
E-mail: hanjie@rice.edu
Department: Computer Science, Chili Lab