Yaofeng Su

yaofengsu@outlook.com | GitHub: InfiniteLoopCoder, 3.3k stars | Google Scholar, 170 citations

Research Interest

My research interest lies in multimodal generative models as a means to build faithful world representations. Through RL, distillation, and agent frameworks, I have contributed to audio-visual generation, autonomous driving, and video understanding. Going forward, I aim to advance multimodal foundation models with principled theoretical insights.

Education

Fudan University Shanghai, China
B.S. in Computer Science *Sep. 2023 – Jul. 2027*
- GPA 3.84/4.00, Rank 7/97. Advised by **Prof. Wenchao Ding** on autonomous driving risk assessment.

University of North Carolina at Chapel Hill Chapel Hill, NC
Exchange Student *Aug. 2025 – Dec. 2025*
- Dean's List. Researched VLA models and LLM agents under the supervision of **Prof. Huaxiu Yao**.

Representative Honors

Tencent Outstanding Student Award (Hunyuan Team) Sep. 2025
Fudan University Outstanding Student Award (Top 8%) Oct. 2024

Internship Experience

Joy Future Academy, JD Dec. 2025 – Present
Intern, Multimodal Foundation Model Research Department
- Developed audio-visual distillation algorithms and real-time streaming generation architectures. Core contributor to OmniForcing (ECCV 2026) and JoyAI-Echo. Currently exploring world models with memory mechanisms. Supervised by **Zeyue Xue**.

Invited Talks

Invited Talk, Lightricks (LTX-Video Team) May 2026
- Presented joint audio-visual streaming generation and self-forcing distillation methodology to Lightricks' CTO and Director of GenAI Research, covering LTX-2 foundation model adaptation and ongoing LTX-2.3 collaboration.

Publications

OmniForcing: Unleashing Real-time Joint Audio-Visual Generation ECCV 2026, **Oral**

Yaofeng Su*, Yuming Li*, Zeyue Xue, Jie Huang, Siming Fu, Haoran Li, Haoyang Huang, Nan Duan

- Designed asymmetric block-causal alignment and joint self-forcing distillation to convert offline bidirectional audio-visual diffusion models into real-time streaming generators at 25 FPS. Featured as a representative real-time audio-visual generation baseline in Alibaba Wan Streamer.

SimpleMem: Efficient Lifelong Memory for LLM Agents ICML 2026

Jiaqi Liu*, Yaofeng Su*, Peng Xia, Siwei Han, Zeyu Zheng, Cihang Xie, Mingyu Ding, Huaxiu Yao

- Designed semantic structured compression and intent-aware retrieval for LLM agent lifelong memory, improving F1 by 26.4% on LoCoMo while reducing token consumption by 30×.

Learning a Unified Risk Map for Autonomous Driving in Partially Observable Environments IEEE RA-L 2026

Jie Jia*, Yaofeng Su*, Zeyu Bao, Yun Hong, Bingzhao Gao, Zhongxue Gan, Wenchao Ding

- Addressed occlusion-induced collision risk in autonomous driving by modeling hidden vehicle motion distributions with adversarial-guided diffusion, enabling risk-aware planning without full observability.

ReAgent-V: A Reward-Driven Multi-Agent Framework for Video Understanding NeurIPS 2025

Yiyang Zhou*, Yangfan He*, Yaofeng Su, Siwei Han, Joel Jang, Gedas Bertasius, Mohit Bansal, Huaxiu Yao

- A multi-agent framework that leverages video understanding to generate reward signals for downstream tasks. Contributed reward-based VLA post-training, surpassing RL baselines by 9.8% on VLA alignment.

JoyAI-Echo: Pushing the Frontier of Long Audio-Visual Generation Tech Report, 2026

Echo Team @ Joy Future Academy, JD

- A long-form multi-shot framework with cross-modal memory and DMD distillation, achieving 7.5× inference speedup and minute-level coherent audio-video story generation.

Skills

Programming: Python, C/C++, PyTorch, Triton, Linux, Git, Docker | **Languages**: English (TOEFL 108)
AI: Reinforcement Learning, Diffusion Models and SDE, LLM/VLM Distillation, Distributed Training, AI Agent, Algorithm Analysis