Mian Wu

tuboushuu@gmail.com Mian Wu mianwu01 Mian Wu mianwu01

Education

Shanghai Jiao Tong University Sep. 2021 – Aug. 2026
B.E in Electrical and Computer Engineering

University of California, Berkeley Sep. 2023 – Sep. 2025
Visiting Student

Publications

"Can Video Models Reason Step-by-Step? The Autoregressive Paradigm for Chain-of-Frames"
Mian Wu*, Bingda Tang*, Disen Lan, Harold Chen *Submitted to CVPR 2027*

– Develop an autoregressive video reasoning framework with on-policy correction, training models to recover from errors in their own generated trajectories rather than solely imitating expert demonstrations. The approach improves long-horizon visual reasoning, reduces error accumulation, and enables more reliable recovery from intermediate mistakes.

"Causal Memory for LLM Agents"
Mian Wu, Yujia Zheng *Submitted to CLeaR 2027*

– Formalize an LLM agent's memory as a temporal causal structure over its past records, recovered from execution logs with causal discovery, to select the records a decision needs and trace the sources it relied on.

"Full Visibility Makes Agent Swarms Converge Early and Spread Overfitted Code"
Xiaoyang Cao*, Yanlong Liu*, **Mian Wu***, Xuan Jiang, Yanchen Liu, Ao Qu, Sraavya Sambara, Yuhan Tang, Awad Abdelhalim, Jinhua Zhao, Paul Pu Liang, Michiel Bakker *Submitted to AAAI 2027*
WMAC Workshop

– Show that a shared leaderboard makes swarms of 32 LLM coding agents converge on one program lineage about twice as fast as a three-peer follow graph and spreads test-case-overfitted code, ranking first of six conditions on the agents' own grader and fourth on the official ALE-Bench judge.

"Beyond the Sampling Frontier: From Evolutionary Discoveries to Generalizable Capability"
Mian Wu, Harman Singh, Boxue Yang, Kurt Keutzer, Aviral Kumar, Sewon Min **NeurIPS 2026**
EvoRobust Workshop

– Propose Evolutionary Self-Improvement (ESI), which runs verifier-guided evolutionary search on training problems the model cannot solve by sampling and fine-tunes the same model on the discovered programs, improving direct generation on unseen hard competitive-programming problems beyond sampling-based self-training.

"RLAC: Reinforcement Learning with Adversarial Critic for Free-Form Generation Tasks"
Mian Wu, Gavin Zhang, Sewon Min, Sergey Levine, Aviral Kumar **ICLR 2026**

– Propose a method for verifying free-form generation outputs using a dynamically trained critic that co-evolves with the generator (RL policy), without relying on static reward models or exhaustive enumerations of possible errors.

"Goal-Driven Reward by Video Diffusion Models for Reinforcement Learning"
Qi Wang*, **Mian Wu***, Yuyang Zhang*, Mingqi Yuan, Wenyao Zhang, Haoxiang You, Yunbo Wang, Xin Jin, Xiaokang Yang, Wenjun Zeng **CVPR 2026**

– Developed a demo-free reward framework that leverages pretrained video diffusion models to turn goal-conditioned videos into reward signals for RL agents, removing the need for hand-crafted reward design.

Projects

VLM-Guided Waypoint Reward Shaping for Sparse-Reward RL May 2024 – Sep. 2024

– **Technologies**: deep reinforcement learning, reward shaping, intrinsic motivation, vision-language models

– Used a VLM to propose and verify semantic waypoints from game frames, serving as a dense reward signal that enabled PPO to outperform RND-based baselines.

Curriculum Learning for Large-Scale Model Training Jan. 2024 – Aug. 2024

– **Technologies**: ConvNeXt, curriculum learning, composite dataset design, data augmentation

– Designed a composite dataset framework to control data presentation order across multiple large-scale datasets. Investigated the impact of curriculum learning on training efficiency under limited-epoch scenarios and varying training set sizes.

RL-based Inverse Design of Photonic Crystals Aug. 2023 – May 2024

– **Technologies**: generative models, deep reinforcement learning, curriculum learning, physics-informed design optimization

– Transitioned from conditional GAN-based generation to a PPO inverse-design framework, achieving higher bandgaps while enforcing physical feasibility.

Honors & Awards

Sep. 2023	Berkeley Global Access Scholarship	*University of California, Berkeley*
Oct. 2022	Undergraduate Excellent Scholarship	*Shanghai Jiao Tong University*
Nov. 2022	Silver Medal	*The University Physics Contest*

Skills

`Toolchain`	Linux, Docker, Huggingface, DeepSpeed, wandb
`Frameworks & Libraries`	PyTorch, Transformers, TRL, PEFT, vLLM, Diffusers, Gym
`Foundation Model`	LLM/MLLM fine-tuning, prompting; LoRA, RLHF
`Reinforcement Learning`	World Model, Actor-Critic, DDPG; MuJoCo, Hydra