

Tianyun Jiang

Fudan University, Shanghai | 23307110360@m.fudan.edu.cn | +86 182 5122 0858

Research interests: Robot Learning, World Models, Reinforcement Learning, and LLM Agents.

EDUCATION

Fudan University

Sep 2023 – Present

B.Sc. student, Data Science and Big Data Technology

GPA: 3.85/4.0

School of Data Science. Selected coursework: Neural Networks and Deep Learning, Introduction to Probability and Statistics, NLP and Large Language Models, Optimization Methods (all graded A).

University of California, Berkeley

Aug – Dec 2025

UCEAP exchange student, Statistics

GPA: 4.0/4.0

Coursework: Data Structures and Algorithms (CS 61B), Statistical Machine Learning, Introduction to Time Series, and Regression Analysis.

RESEARCH EXPERIENCE

DeltaWAM: Change-Centric Visual Foresight via Delta Tokens for Efficient World-Action Model

First author | Preprint: arXiv:2609.33177 (In submission of ICLR 2027) | [GitHub Link](#)

Jun – Sep 2026

EXcel-Lab, University of Central Florida | Prof. Yuzhang Shang

Existing World-Action Models demonstrate spatio-temporal redundancy and computational overhead. DeltaWAM shifts the predictive world unit to a single delta, which models the difference of DINO latent between consecutive frames. This paradigm proves to be more dynamic-aware and action-forcing compared with predicting full DINO latent at each timestep—the success rate improves **13.8%** on the Libero Benchmark under the same architecture and training recipe. Our work builds on pretrained DeltaWorld, which predicts delta tokens autoregressively without iterative denoising. Pretrained DeltaWorld only conditions on history contexts, we inject high-level semantics via Gated AdaLN to make its predictions more task-relevant. Pretrained DeltaWorld first gets finetuned on demonstrations, and then it gets frozen during the training stage for the Action DiT, which receives both current DINO latent and predicted future delta tokens. With only **120.3M** trainable parameters and **256** GPU hours on 2 H100 GPUs, DeltaWAM achieves **92.8%** average success rates on Libero and demonstrates procedural robustness on Libero-Pro with low inference latency (142.1ms per action chunk). In this project I conducted literature review, paper writing and most of the experiments (except the real-world experiments on Piper).

PD-IPO: Primal-Dual Identity Preference Optimization

Contribute to the construction of algorithm, theory proof and implementation of experiments | [GitHub Link](#) Apr – Aug 2026

Prof. Yihan Du's group, Singapore University of Technology and Design (SUTD)

Primal-Dual Identity Preference Optimization (PD-IPO) is a novel reinforcement learning algorithm for KL-regularized constrained policy optimization that integrates IPO-style squared log-ratio losses with projected Lagrange-multiplier updates. The goal is to maximize the policy's expected rewards without violating the cost constraints. The theoretical framework rigorously establishes non-asymptotic $\mathcal{O}(1/\sqrt{K})$ bounds over K primal-dual iterations for both suboptimality and constraint violations, accounting for preference estimation and importance-sampling errors. Furthermore, we extended the framework to an online setting (O-PD-IPO) utilizing an upper-confidence bound (UCB) for rewards and a lower-confidence bound (LCB) for costs, enabling safe and efficient exploration without requiring uniform policy coverage assumptions. In a simulated setting, O-PD-IPO converges to higher object values than PD-IPO through online exploration. In this project, I co-wrote algorithms and proofs, and constructed the simulated experimental framework for online PD-IPO, supporting empirical investigation of preference-based policy optimization under cost constraints.

Surgical WAM: A World-Action Model for Data-Efficient Surgical Robot Learning

Second author | Preprint: arXiv:2608.11204

Mar – Jul 2026

EXcel-Lab, University of Central Florida | Prof. Yuzhang Shang

The project introduces the Surgical World-Action Model (Surgical WAM), a unified generative model designed to overcome the scarcity of action-labeled data in surgical robot learning. It utilizes a two-stage training approach where the model first learns surgical visual dynamics from abundant, action-free endoscopic videos before being fine-tuned with a limited budget of action-labeled demonstrations. During deployment, Surgical WAM operates as a closed-loop controller that jointly predicts future observations and robot action chunks. When evaluated on a suite of surgical manipulation tasks on the SurRoL benchmark, action-free video pretraining improved the model's average success rate from 63.5% to 77.8%, demonstrating that leveraging unlabeled video is a highly data-efficient and practical strategy for scaling reliable surgical robot control. In this project I contributed to literature review and implementation of the experiments.

PROJECT EXPERIENCE

RIGOROUS: Real-Time Interruptible LLM Agents for Social Deduction Games

Three-person project | GitHub Link

Mar – Jun 2026

In RIGOROUS, we introduce an end-to-end Large Language Model (LLM) agent designed for the complex social deduction game Werewolf, aiming to overcome the limitations of rigid turn-taking and the reliance on expensive human preference data. Built on the Qwen3-4B-Instruct model, the agent is trained through a two-stage pipeline: a Supervised Fine-Tuning (SFT) phase that distills high-quality gameplay trajectories from a stronger teacher model, followed by Kahneman-Tversky Optimization (KTO) that automatically generates preference labels using a recursive "speech-versus-silence" counterfactual branch simulation. Furthermore, the system incorporates a streaming interruption mechanism that allows agents to dynamically decide whether to interrupt an ongoing utterance in real time, demonstrating the viability of fluid conversational dynamics in AI interactions. Evaluations in the 8-player Werewolf Arena demonstrate that the SFT phase significantly elevates the agent's performance, and the win rate against the base model is 77%. The subsequent KTO optimization provides a further 10% gain over its SFT baseline. Additionally, separate testing validates the functional success of the real-time turn-switching mechanism, recording 1,173 interruptions while the SFT model decisively won 81 out of 99 completed interruptible games against the base model. I contributed to writing the report and conducting experiments (I mainly worked on training and evaluating the models).

Beyond Static Vectors: Is It Possible to De-homogenize and Debiase Preliminary Social Matching via Agent-Mediated Dialogues?

Independent project

Apr – Jun 2026

In this project, I explore using LLM agents for preliminary social matching. It focuses on scenarios like dating and recruitment to address algorithmic homogenization and historically encoded biases shown in traditional algorithms, such as computing static-vector similarity among users. I deployed agents to engage in counterpart-specific, low-information dialogues. After interacting, they generate an auditable follow-up score and a concise rationale for human review. Evaluations demonstrated the effectiveness of this dynamic interaction. In a speed dating simulation, it successfully reduced recommendation overlap by 22.9% ($p = 0.0047$) among comparable users. In a recruitment benchmark, it brought the adjusted gender odds ratio much closer to parity (from 0.695 to 0.881). These results highlight how AI agents can create individualized differentiation and weaken demographic associations in social gatekeeping platforms.

TECHNICAL SKILLS

Programming and Mathematics Proficient in Python and PyTorch; C, Java, R, and MATLAB. Have a solid mathematic foundation for understanding algorithms and theoretical proofs.

Deep learning, Robot Learning and RL Familiar with the basics in deep learning and robot learning, as well as common algorithms in reinforcement learning. Gain hand-on experiences by training and

deploying a pick-and-place policy on SO-101.

MY CURRENT RESEARCH INTERESTS

Here I list some research problems that I frequently think about:

- GPT-Astra has demonstrated strong 3D understanding and agentic planning for robotics. I think this is significant improvement in system 2, which provides the support of high-level task planning. How can we further work on system 1, which relates more to reactive low-level control, to make it coordinate better with the system 2? This can be formulated as a problem of **instruction following**: after accumulating lots of experience in atomic behaviors, the policy is steered towards finishing various tasks under various environments with the help of agentic planning. Language is very suitable for reasoning, but is limited for bridging from high-level semantics to low-level control. Certain behaviors can't be precisely described by language, especially in the dexterous manipulation. **Sometimes we might still need a world model in the middle as the interface.** Another question I have is **how many computational budgets show we allocate for the planning agent in different tasks?** This matters because there is trade-off between task success rates and deployment latency, and naturally we want to find a balanced point. The properties of different tasks also differ. **Can we utilize GPT-Astra mainly for curriculum training of the policy and discard it at inference?**
- I found out that the **training loss of the policy**(for continuous actions it's usually a flow-matching loss) **correlates very weakly with the actual success rates.** The evaluation loss during the training stage is not so informative. The metric we rely on more is the number of episodes or training steps, which depends on empirical experiences. The data efficiency of different models is different, so we can't solely rely on that. How can we get a more reliable metric (like a new training loss) that correlates more with the models' performances? This problem is hard because **there is stochasticity both in the policy's predictions and the world dynamics**, and we might have to take this into account. For example, the recovery ability has little to do with the loss in imitation learning, currently it can only get evaluated through actual rollouts. The stochasticity of the policy and the environment will affect the rollouts.
- How can we improve the introspection ability of the robot so that it is **more aware of task uncertainties?** For tasks that are more like extrapolation and not so out of domain, reasoning and planning definitely help a lot for the execution. However, in failure modes or novel tasks the robot can still need humans' interventions(i.e. collect more demonstrations). **Can we make the policy actively and intelligently ask for help when it's necessary?** The failure-driven active learning might be meaningful for zero-shot or few-shot adaptations.