

Junyi Wu

Ph.D. Candidate in Mathematics, Purdue University

Applied Science | Post-Training | Reinforcement Learning

West Lafayette, IN | +1 (765) 767-3390 | wu2393@purdue.edu

junyi-w.github.io | github.com/Junyi-W | Google Scholar | LinkedIn

Education

Purdue University

Ph.D. in Mathematics; GPA: 3.97/4.00; Advisor: Prof. Guang Lin

Fall 2024 – Expected May 2028

West Lafayette, IN

University of Chinese Academy of Sciences

B.S. in Mathematics and Applied Mathematics; GPA: 3.91/4.00

2017 – 2022

Beijing, China

Exchange: National University of Singapore, Spring 2020 (GPA: 4.0/4.0)

Research Experience

Purdue University

Graduate Research Assistant, Advisor: Prof. Guang Lin

Spring 2025 – Present

West Lafayette, IN

Post-Training & Preference Alignment

DIDR: KL-Regularized Reward Fine-Tuning of Generative Models

Junyi Wu, Weijian Luo, Haoyang Zheng, Ruizhe Zhang, Guang Lin | *NeurIPS 2026* | Code

- Derived a **post-training objective** for one-step text-to-image generators that shares the distribution-level optimum of **KL-regularized RLHF**, aligning reward maximization with reference-model regularization.
- Developed a **Monte Carlo reward-gradient estimator** using differentiable denoising rollouts through a frozen reference model; enabled preference alignment from text prompts without image training data.
- Built the **DDP/FSDP, BF16** training stack for **2.6B and 6.15B** models on **8 H100 GPUs**, jointly training a generator and an auxiliary score network with frozen reward and reference models.
- Evaluated alignment beyond the training reward: improved Z-Image ImageReward from 0.39 to **1.08** and FID from 36.7 to **22.1** versus its one-step initialization, collaborating with hi-Lab, Xiaohongshu Inc.

COCOA: Supervised Fine-Tuning & Preference Optimization for VLA Policies

Yi Wu, Manav Kulshrestha, Junyi Wu, Kai Cheng, Nan Jiang, Aniket Bera, Lin Tan | *Under review, ICLR 2027*

- Derived a two-stage **SFT** $\rightarrow$ **offline preference optimization** recipe for action correction, adapting an SRPO-style objective to rectified-flow policies with a **Flow-DPO**-style likelihood-ratio approximation.
- Co-developed an **action-conditioned correction policy** that learns to repair proposed actions while keeping the base VLA frozen, separating policy improvement from base action generation.
- Improved **CALVIN** five-instruction success from 77.8% to **81.9%** on FLOWER VLA and average completed sequence length from 3.92 to **4.03** on OpenPi $\pi_{0.5}$.

Quantum Safe Stochastic Linear Bandits

Ruizhe Zhang, Junyi Wu, Guang Lin | *NeurIPS 2026*

- Co-developed **constrained Thompson sampling** for learning unknown rewards and safety constraints, with high-probability action feasibility and **polylogarithmic regret** under coherent quantum feedback.
- Designed BCO-QLinTS to select safe actions via quantum convex optimization with a safe-set oracle, avoiding explicit estimation of the constraint matrix.

Selected Project

LLM Fine-Tuning, Inference & Representation Analysis

Aug. – Oct. 2026

Independent experiments with Qwen language models

- Implemented **full-parameter fine-tuning** experiments for **Qwen2.5-0.5B** on FineWeb-Edu with BF16 and gradient accumulation; compared representation-fusion and prefix-rereading variants against controlled baselines.
- Evaluated recurrent inference on **Qwen3-4B-Instruct** over **14,042 MMLU** and **1,000 SciQ** questions; compared Jacobian Lens and Logit Lens readouts on **Qwen3-4B/8B** across six probing tasks.

Technical Skills

Post-Training & RL: Reward fine-tuning, KL regularization, SFT, preference optimization, Thompson sampling

Programming & ML: Python, C++, PyTorch, JAX, HuggingFace Transformers/Diffusers, NumPy

Distributed Training: DDP, FSDP, NCCL, CUDA, BF16, gradient accumulation

Tools: Git, Linux, Docker, Weights & Biases, LaTeX

Service: Reviewer, NeurIPS 2026 | Co-Mentor, Purdue SURF (Summer 2026)