

YUANPENG LI

☎ + (1) 949-618-8992 ✉ yuanpl2@uci.edu 🌐 yuanpeng-li.github.io 📄 github.com/Yuanpeng-Li 🔗 linkedin.com/in/yuanpengli

EDUCATION

University of California, Irvine

Doctor of Philosophy (PhD) in Statistics

Irvine, CA

Sep. 2024 – Present

Jilin University – School of Mathematics

Bachelor of Science Degree, Major in Mathematics and Applied Mathematics

Changchun, China

Aug. 2020 – Jul. 2024

RESEARCH EXPERIENCE

Probe-FairGRPO: Gradient-Conflict-Aware Multi-Objective RL Fine-Tuning of LLMs

Irvine, CA

Graduate Researcher, UC Irvine

May 2026 – Present

- Reframed multi-reward GRPO fine-tuning (correctness / format / length rewards) as *multi-task RL* to bring fair-gradient combination (FairGrad) into LLM post-training; to make this tractable on an LLM, a lightweight *probe* builds a per-reward $K \times K$ gradient Gram matrix over a small parameter subset (lm-head / last layers / LoRA) to cheaply estimate inter-objective gradient conflict and alignment.
- A FairGrad solver maps the probe Gram to adaptive per-reward weights α (down-weighting conflicting, up-weighting aligned objectives); clipping each reward's GRPO loss independently before weighting lets one scalar backward preserve the exact weighted gradient under PPO/GRPO clipping — unit-test-verified against the biased mix-then-clip alternative.
- Built a reproducible multi-objective GRPO training-and-eval stack on verl 0.8 / vLLM / FSDP for multi-GPU training (standalone, unit-tested core) with W&B diagnostics; configured GSM8K / DAPO-Math training and AIME / MATH-500 / OlympiadBench / AMC23 / GPQA evaluation on Qwen2.5-0.5B/7B and Llama-3.1-8B-Instruct.

TOPPO: Rethinking PPO for Multi-Task RL with Critic Balancing

Irvine, CA

First author; under review (arXiv:2605.11473); Advisors: Prof. Annie Qu, Prof. Rui Miao

Jul. 2023 – Apr. 2026

- Designed *Critic Balancing* for multi-task PPO — per-task PopArt value normalization, critic LayerNorm, and per-side gradient combiners (PCGrad / CAGrad / FairGrad) — to fix critic-side gradient ill-conditioning where tail tasks stall.
- On Meta-World MT50, beat published SAC- and ARS-family baselines with 90.88% mean and 56.50% worst-10 tail-task success at 1/22.7 the parameters (a +27.2% worst-10 gain over ARS-LoRA).

PROJECTS

gradescope-mcp | Python, Model Context Protocol, uv

- Built an MCP server exposing Gradescope to AI agents — 34 tools across course management, grading, rubric CRUD, and statistics (all writes gated behind explicit confirmation), plus a reusable human-approved batch-grading skill covered by 30 automated tests.

StudyGround | JavaScript, Node, Claude Code Plugin, Pyodide

- Built a Claude Code plugin for AI-assisted self-study: a web reader with inline *Ask* buttons, in-browser Python via Pyodide, and NotebookLM-style RAG over course materials (BM25 plus embedding search with page-anchored citations).

TEACHING

- Graduate Teaching Assistant**, UC Irvine Statistics (Sep. 2024 – Present): STATS 7, 67, 110, 120C, and 205P (Bayesian Data Analysis, MDS).

INTERNSHIP EXPERIENCE

Synkrotron

Xi'an, China

Intern, AI Research Department

Dec. 2022 – Jan. 2023

- Configured remote Linux devices (via FRP) for an automated road-patrol pipeline and studied computer-generated simulation datasets to improve real-world autonomous-driving test performance.

AWARDS

- Scholarship in the academic year 2022-2023 Oct. 2023
- The Chinese Mathematics Competitions: **National 3rd Prize** Dec. 2021
- Jilin Province Mathematics Competitions: **Provincial 2nd Prize** Dec. 2021

SKILLS

- LLM Post-Training & RL:** PyTorch, Hugging Face, vLLM, verl, FSDP, PPO/GRPO, RLHF/RLAIF, LoRA/PEFT
- Agent & Retrieval Tooling:** MCP, RAG, BM25, vector search, Claude Code Plugin, Pyodide
- Infra & Engineering:** Python, C, Git, Docker, Linux, SLURM, Ray, W&B, uv, pytest