

Tuen-Yue Tsui

📍 Philadelphia, PA ✉️ tuenyue.tsui@gmail.com 🌐 TSUITUENYUE 🌐 Personal Website

Education

University of Pennsylvania

Sept 2024 – May 2026

MSE in Robotics

- GPA: 4.0/4.0

Wuhan University

Sept 2020 – Jun 2024

BE in Artificial Intelligence

- GPA: 3.75/4.0

Selected Research Experience

Minimal Witness Reinforcement Learning

July 2026

Shanghai AI Laboratory, Collaborators: Yuqiang Li, Zhehong Ai

- Formalized minimal-witness identification: recovering the antichain of subset-minimal sufficient sets from a black-box sufficiency verifier alone. Proved that every RL objective that scores rollouts separately or only through the success probability (PPO, GRPO, MaxRL, soft RL, GFlowNets) has an optimum that misses this antichain.
- Derived the credit assignment from the minimality objective itself: each success certifies its up-set in the subset lattice, a group is valued by the measure of the certified region, and each rollout is credited by what the region loses without it. Positive credit holds exactly for the group-minimal successes.
- Obtained an exact planner and an unbiased group policy gradient. MWRL recovered 67% of the antichain in prime-implicant enumeration, where size-penalized baselines recovered at most 24%; it also amortized minimal reaction-condition discovery to unseen Suzuki–Miyaura substrates and recovered families of minimal circuits in Qwen3 language models.

The Lattice of Transition Laws

Aug 2026 – Sept 2026

University of Pennsylvania, Advisors: Jiatao Gu, Lingjie Liu

- Placed diffusion, autoregression and the hybrids between them on one corruption lattice, and defined the cost of a decoding schedule as the dependence its parallel steps discard.
- Proved that the fewest steps of a zero-cost schedule are set by the geometry of the data; for data Markov on a graph they equal the graph's treedepth, logarithmic in sequence length and linear in the side length of a grid.
- Predicted the ranking of schedules before decoding with a pairwise-dependence kernel estimated from pre-trained weights, and verified most predictions across text, image and video generation.

Noise-Level Adjacency in Diffusion Training

Aug 2026 – Sept 2026

University of Pennsylvania, Advisors: Jiatao Gu, Lingjie Liu

- Proved a lower bound on the fitting error of a network shared across noise levels, set by the distances between neighbouring posterior means minus the change the network can make between neighbouring embeddings.
- Showed that the benefit of weight sharing rests on adjacency in the noise-level embedding, while its direction is a design choice: reversing the embeddings leaves the error unchanged and shuffling them raises it, on a Gaussian mixture and on CIFAR-10.
- Showed that separate networks on contiguous ranges match the shared network, and that dividing the levels at equal shares of these distances lowers the error of all but the smallest networks.

Mori–Zwanzig Formulation of Bayesian Filtering

Sept 2025 – May 2026

University of Pennsylvania, Advisors: Konrad Kording, Jiatao Gu, Lingjie Liu

- Showed that the information a projection filter discards bounds its cost only on average: on a single stream, a projection that discards little can cost arbitrarily more in the future.
- Applied the Mori–Zwanzig formalism to the filter's own belief: under the Fisher-metric projection, every projection filter is a Markovian closure up to a defect, and its error is the sum of its defects and of the returns of what it discards.
- Derived the memory in closed form for Gaussian references, decaying $k + 1$ times as fast as the filter forgets

for an order- k closure and as a power law for a static parameter, and from these rates a rule for what a filter should resolve.

Projects

Vectorized Environment and Multi-Agent RL for Teamfight Tactics

Ongoing

- **Multi-Agent RL + Simulation.** Building a GPU-resident vectorized simulator of Teamfight Tactics in PyTorch, in the style of Isaac Lab: many complete eight-player games step in one process with all state on the device, covering the shared card pool, shop, economy, items, legal-action masks and batched combat, behind a Gymnasium interface, for multi-agent RL in which eight policies compete in each game.

Apodexis: Mathematical Proof-Chain Visualization Tool

[github link](#)

- **Research Tool + Formal Methods.** Built a native SwiftUI macOS app for organizing long mathematical and scientific proof chains as typed dependency graphs, with semantic proof relations, branch/fork workflows, LaTeX rendering, formal-hole detection, local JSON persistence, and Markdown export.

Scalable Quadruped Imitation from Monocular Video

[github link](#)

- **CV/CG + Robotics.** Retargeted quadruped motions from monocular videos via a learnable skeleton; removed MoCap/manual keypoints.

Minimum-Snap Trajectory Generation for Quadrotors

[github link](#)

- **SLAM + Planning + Control.** Integrated VIO, SE(3) controller, and ray-casting-powered A* (**30×** faster: **22 s** $\rightarrow$ **0.7 s**; **1st** on leaderboard) for path planning with a time-optimal snap solver.

Selected Engineering Experience

Large Language Model Pretraining Pipeline on Multiple GPUs

- Built and ran an end-to-end pretraining pipeline for a 3B-parameter language model, covering data scraping, cleaning, deduplication, tokenization, distributed training, and experiment monitoring.
- Trained the model on 8 NVIDIA B200 GPUs in parallel, gaining hands-on experience with large-scale GPU training workflows, data throughput bottlenecks, and multi-GPU debugging.

Language Model Post-Training with Verifiable Rewards

- Post-trained Qwen3-4B-Base with RLVR in verl on a multi-GPU Slurm cluster, comparing GRPO, MaxRL and MWRL under the same model, data and budget (256 problems with 16 rollouts per step, 500 steps).
- Built a data-sufficiency reasoning task whose minimal answers form an exactly known antichain, with its verifier and reward; registered custom group advantage estimators in verl's Ray workers, evaluated held-out recall@k with vLLM, and streamed the training curves live to a web page.

Reinforcement Learning Policies for Embodied Agents

- Trained RL policies on large scaled simulations for humanoid locomotion and quadruped object-centric interaction tasks.
- Designed observation spaces, command interfaces, reward terms, and curriculum/randomization settings for policy learning in high-throughput simulation environments.

Skills

Foundation Models: LLM pre-training, post-training (SFT, DPO, RLHF/RLAIF), evaluation, inference

Generative Modeling: diffusion and autoregressive models and the hybrids between them, masked diffusion, flow matching, decoding schedules, few-step generation, noise-level design

RL / Agents: policy optimization, reward modeling, group-relative policy gradients, curriculum learning, embodied RL, RL for scientific discovery

Theory: partial orders and lattices, antichains and minimal sufficient sets, credit assignment, dependency structure of data, lower bounds

Robotics / Simulation: motion planning, trajectory optimization, VIO/SLAM, whole-body control, imitation learning, ROS, Drake, Isaac Lab, Newton, Genesis

Systems / Tools: Python, C++, C, PyTorch, JAX, Linux, Git, Slurm, distributed training, multi-GPU training workflows

Languages: Chinese (Native), English (Full Professional)