SHUO XU

Email: xushuo23@gmail.com   Phone: +86 15067777600

EDUCATION

Sorbonne University   **Paris, France**

MSc in Mathematics and Applied Mathematics (Statistical Learning & Algorithms)   *2024.09 – Present*

- GPA：14.52/20 - Mention B (Second Distinction, $\approx 3.6/4.0$)
- Core Courses: Reinforcement Learning, Deep Learning, Generative Models; Stochastic Optimization, Statistical Learning, High-Dimensional Statistics; Advanced Probability, Concentration Inequalities, Extreme Value Theory

Renmin University of China & Sorbonne University   **Beijing, China**

BSc in Mathematics and Applied Mathematics (Sino-French Dual Degree Program)   *2021.09 – 2025.07*

- Core GPA (Mathematics courses): $\approx 3.5/4.0$
- Core Courses: Measure Theory & Probability, Linear Algebra, Optimization, Real Analysis, Functional Analysis, Mathematical Statistics, Numerical Analysis, Ordinary Differential Equations, Mathematical Modeling

RESEARCH EXPERIENCE

Self-Rewarded Diffusion Models for Image Reasoning Generation   **Beijing, China**

Assistant Researcher under Zhiyuan Ma (Tsinghua University)   *2025.06 – Present*

- Designed a fully label-free, rule-free self-reward mechanism that extracts structural stability signals from the generation distribution, and integrates the self-reward as a relative advantage signal into policy updates to guide diffusion models in reasoning over relative position, counting, and spatial consistency.
- Leveraging relative preference learning and MaxEnt IRL – based reward modeling ideas from LLMs, constructed relative rewards via clustering consistency and majority voting, and built a reinforcement learning fine-tuning pipeline within the DDPO framework supporting multi-sample sampling, cluster ranking, and reward extraction.
- Developed a learnable Inference Consistency Module (ICM) using VQ-GAN discrete latent representations to model reasoning-related structural relations, whose outputs serve as supplementary reward signals during training.

Outcome: Experiments ongoing; evaluation on public benchmarks against DDPM, Stable Diffusion, and DDPO, targeting submission to NeurIPS / ICML / ICLR / CVPR.

ACADEMIC PROJECTS & THESIS

Autonomous Kart Racing Control with Reinforcement Learning (SuperTuxKart)   **Paris, France**

Reinforcement Learning Course Project   *2025.12*

- In the SuperTuxKart reinforcement learning environment, designed a raw observation wrapper to structure high-dimensional heterogeneous observations into box and sequence features, encoded via MLP and Transformer respectively.
- Implemented PPO under an Actor – Critic framework to learn racing control policies with hybrid continuous – discrete action spaces.
- Addressed training instability and gradient explosion through learning-rate tuning, gradient clipping, and checkpointing, and completed deployment and evaluation scripts for participation in the unified racing benchmark.

Probability Measure Formulation of MDPs and Model-free Monte Carlo Trap Recognition   **Beijing, China**

Undergraduate Thesis (Rank 1/32, University Excellent Thesis Award)   *2024.10-2025.03*

- Formulated a formal MDP model from a probability and measure-theoretic perspective, facilitating a principled understanding of the relationships among value functions, policies, and stochastic processes.
- After analyzing the inefficiency of traditional Monte Carlo methods in environments with absorbing states, designed and implemented a sampling-statistics – based extension, Monte Carlo Trap Recognition (MCTR), to identify and avoid potential absorbing regions.
- Conducted algorithmic experiments in custom environments, demonstrating that MCTR improves exploration efficiency over standard MC methods in the presence of absorbing states.

ADDITIONAL INFORMATION

- Computing Skills: Python, Pytorch, reinforcement learning frameworks (Gymnasium), Latex
- Languages: Chinese (native), English (CET-6), French (TCF B1)
- Teaching Assistant: Led tutorials and graded assignments for sophomore courses in Probability Theory and Mathematical Statistics