

EDUCATION

Peking University

Expected 07/2026

B.S. in Physics

- **GPA:** 3.5/4.0
- **Relevant Fields:** Reinforcement Learning, Computer Vision, AI for Science.

ACADEMIC VISITS

Shenzhen International Quantum Academy

01/2025 – 02/2025

AWARDS

Alishan Scholarship

Top 3%, 09/2025

Outstanding Research Award

Top 10%, 09/2025

National Endeavor Scholarship

09/2025

PUBLICATIONS

Let's Verify and Reinforce Image Generation Step by Step

CVPR 2025

Renrui Zhang*, Chengzhuo Tong*, Zhizheng Zhao*, Ziyu Guo*, Huaidong Zhang, Manyuan Zhang, Peng Gao, Hongsheng Li

(* Co-first authorship)

RESEARCH EXPERIENCE

Reward and Policy Distribution Optimization in GRPO

03/2025 – 07/2025

University of Illinois Urbana-Champaign (UIUC) | Advisor: Prof. Minjia Zhang

Conducted a comprehensive analysis of Group-Based Reward Policy Optimization (GRPO), proposing systemic solutions for two fundamental limitations:

- **Refining Credit Assignment:** Addressed the coarse granularity of sample-level feedback. Designed a **Token-Level Critic** to estimate token vitality ($b_\phi(t)$) and developed a **Probability Compensation** mechanism to amplify gradients for "rare gems" (high-reward, low-probability samples), significantly enhancing learning efficiency.
- **Mitigating Exploration Sparsity:** Diagnosed zero-gradient issues in "All-Fail" groups caused by distribution mismatch. Proposed **Resource Re-allocation** to focus compute on hard instances and a **Negative Advantage** strategy to suppress failure modes. Also explored **Dense Reward Shaping** via Ground Truth Likelihood for mathematical reasoning tasks.

Chain-of-Thought Reasoning for Advanced Image Generation

09/2024 – 01/2025

CUHK & ByteDance | Advisor: Dr. Renrui Zhang

- Applied **Chain-of-Thought (CoT)** to autoregressive image generation, integrating Direct Preference Optimization (DPO) to enhance complex reasoning capabilities.
- Proposed the **Potential Assessment Reward Model (PARM)** to adaptively evaluate intermediate steps. The model selects high-potential reasoning paths, effectively balancing test-time compute efficiency with generation quality.

- Enhanced the *Show-o* model, achieving a **+24% improvement on GenEval** and surpassing Stable Diffusion 3 by **+15%** in benchmark performance.

Improving Crystal Structure Prediction via Niggli Reduction

07/2025 – 09/2025

The Chinese University of Hong Kong (CUHK) | Advisor: Prof. Shengchao Liu

- Identified a critical limitation in the state-of-the-art **DiffCSP** framework: reliance on $E(3)$ equivariance fails to capture **lattice basis transformation invariance**, leading to physically invalid predictions.
- Proposed a principled framework integrating **Niggli reduction** to align predictions with canonical lattice representations, effectively closing the theoretical gap.
- Designed a **differentiable Proxy Loss** to enable end-to-end training on these canonical representations, ensuring the physical validity of generated crystal structures.

Resistive Plate Counter Development

03/2024 – 07/2024

Peking University | Advisor: Prof. Qite Li

- Assembled and commissioned a Resistive Plate Counter (RPC) detector for high-energy physics experiments.
- Developed signal processing pipelines and machine learning-based feature extraction algorithms for precise Particle Identification (PID).

SELECTED COURSE PROJECTS

Neural Network Solver for Electric Fields

12/2024

Course: Physics and AI | Advisor: Prof. Yanqing Ma

- Implemented a **Physics-Informed Neural Network (PINN)** to solve complex electric field distributions.
- Designed custom loss functions integrating boundary conditions and PDE residuals to approximate physical solutions.

SKILLS

Languages	Python, LaTeX, CERN ROOT, C++
Tools	PyTorch, Git, Linux Environments