Yifu Luo

(+49) 1759402517 yifu599@gmail.com Homepage

EDUCATION

Ph.D., Technical University of Munich, Germany *Sep 2026 - Aug 2029*
Computer Science.
Advised by Prof. Dr. Gjergji Kasneci.

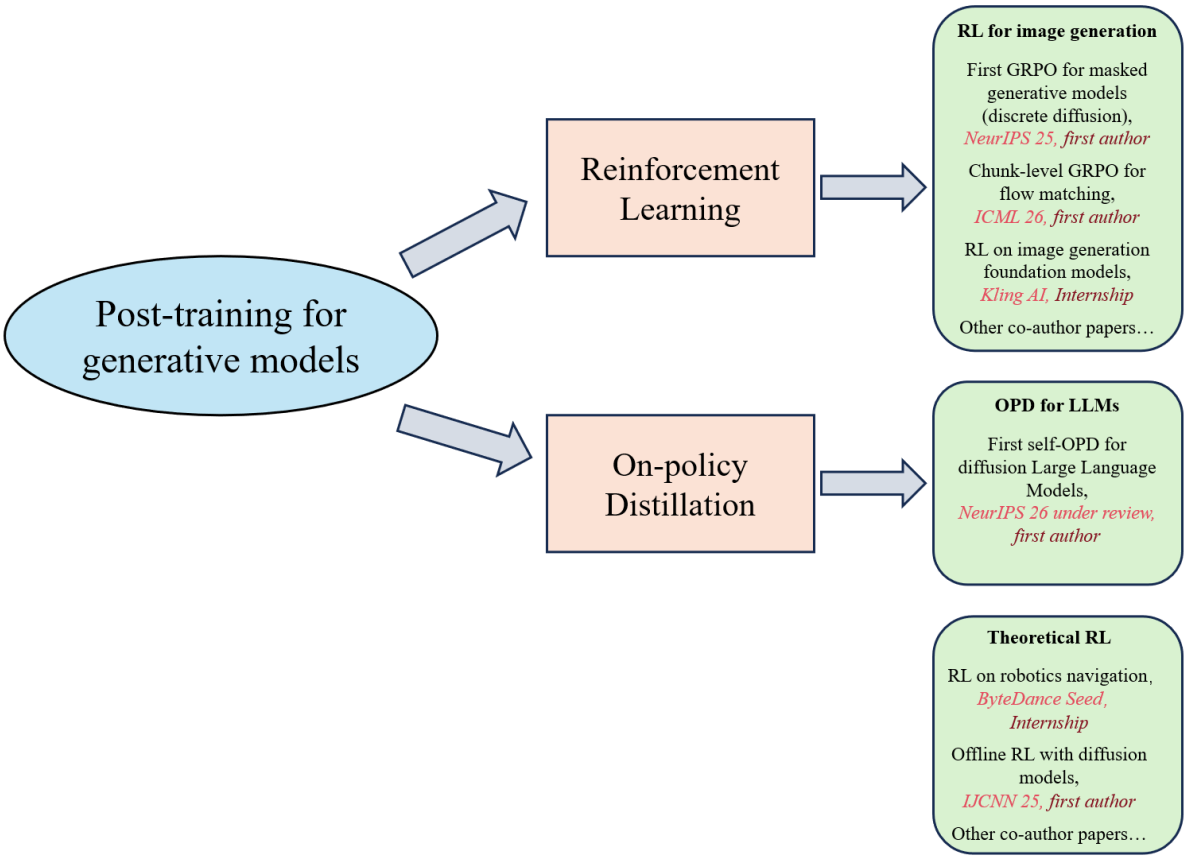
M.E., Tsinghua University, China *Sep 2023 - Jun 2026*
Artificial Intelligence Program, Electronic and Information Engineering.
Advised by Prof. Dr. Xueqian Wang.

B.E., Tsinghua University, China *Aug 2018 - Jun 2022*
Electronics Information Engineering

Exchange Program (M.Sc.), Technical University of Munich, Germany *Apr 2024 - Oct 2024*
Computer Science

RESEARCH INTEREST

I am interested in advancing the post-training paradigms of generative models, with a particular emphasis on reinforcement learning (RL) and on-policy distillation (OPD). My past works range from RL post-training to enhance diffusion models (image generation) to self-OPD on LLMs. I have also explored theoretical RL and its applications on robotics.



INDUSTRY RESEARCH EXPERIENCE

Research Intern, Kling AI *Jul 2025 - Oct 2025*
Mentored by Dr. Changqian Yu
– I enhanced flow-matching text-to-image models (FLUX) through RL post-training. I shifted GRPO from the step-level to the chunk-level yielding superior image generation performance. ***(ICML 26, first author)***
– I also developed reward models using Qwen-VL and InternVL for image generation **foundation models**, including the data construction and model training.

Research Intern, ByteDance Seed *Mar 2025 - Jul 2025*
Mentored by Dr. Yansheng Wang
– I developed RL and imitation learning methods for robotic decision-making from video generation.
– I also developed reward models using Qwen-VL for robotics spatial reasoning and decision-making.

ACADEMIC RESEARCH EXPERIENCE

Research Intern, Max Planck Institute for Intelligent Systems *Jan 2026 - May 2026*

Advised by PI. Shiwei Liu

– I proposed the first On-policy Self-distillation (OPSD) approach for diffusion Large Language Models (dLLMs). We shifted the OPSD formulation from prefix conditioning to self-suffix generation conditioning. Our approach yields superior performance in **math reasoning tasks**. ***(NeurIPS 26 under review, first author)***

Research Assistant, Tsinghua University *Sep 2023 - Jun 2026*

Advised by Prof. Xueqian Wang

– I focused on RL and its post-training paradigms to enhance generative models.

– I proposed the first GRPO approach for masked generative models in text-to-image generation, with a novel transition probability tailored for them. ***(NeurIPS 25, first author)***

– I also proposed a frequency-aware decision diffuser for offline RL. ***(IJCNN 25, first author)***

PUBLICATIONS

★ Highlight

Accepted

★ [1] **Yifu Luo**, Xinhao Hu, et al. "Reinforcement Learning Meets Masked Generative Models: Mask-GRPO for Text-to-Image Generation". In *Neural Information Processing Systems* ***(NeurIPS)***, 2025. [arxiv][code]

– Propose the first GRPO approach for masked generative models in text-to-image generation. We redefine the transition probability tailored specifically for masked generative models, and explore several useful strategies to further enhance our method.

★ [2] **Yifu Luo**, Haoyuan Sun, et al. "Principled RL for Flow Matching Emerges from the Chunk-level Policy Optimization". In *International Conference on Machine Learning* ***(ICML)***, 2026. [arxiv][code]

– Propose a chunk-level RL optimization for flow-matching text-to-image generation. We shift GRPO from the step-level to the chunk-level, aggregating consecutive steps into a coherent "chunk" to mitigate the sparse rewards bottleneck.

[3] **Yifu Luo**, Yongzhe Chang, et al. "Wavelet Fourier Diffuser: Frequency-Aware Diffusion Model for Reinforcement Learning". In *International Joint Conference on Neural Networks* ***(IJCNN)***, 2025. [arxiv]

– Propose a frequency-aware decision diffuser for offline RL. We introduced frequency analysis into diffusion-based decision making for superior stability.

[4] Haoyuan Sun, Bo Xia, **Yifu Luo**, et al. "Calibration Enhanced Decision Maker: Towards Trustworthy Sequential Decision-Making with Large Sequence Models". In *Transactions on Machine Learning Research* ***(TMLR)***. [Openreview]

– Propose a decision-making method for minimizing the state-action marginal distribution distance and enhancing the agent's calibration.

[5] Bo Xia, **Yifu Luo**, et al. "D3D: Conditional Diffusion Model for Decision-Making Under Random Frame Dropping". In *International Conference on Robot and Human Interactive Communication* ***(ROMAN)***,2024. [IEEE]

– Propose a conditional diffusion method for decision making under random frame dropping.

[6] Sinan Du, Jiahao Guo, Bo Li, Shuhao Cui, Zhengzhuo Xu, **Yifu Luo**, et al. "VQRAE: Representation Quantization Autoencoders for Multimodal Understanding, Generation and Reconstruction". In *Computer Vision and Pattern Recognition* ***(CVPR)***, 2026. [arxiv][code]

– Propose a quantization version of autoencoders for unified visual tasks for both understanding and generation.

Under Review

★ [1] **Yifu Luo**, Zeyu Chen, et al. "Learning from the Self-future: On-policy Self-distillation for dLLMs". Under review at *Neural Information Processing Systems* ***(NeurIPS)***, 2026.

– Propose the first OPSD approach for dLLMs. We reframe the OPSD formulation from prefix conditioning to the self-suffix generation conditioning tailored specifically for dLLMs. We also shift dense supervision from the token-level to step-level, aligning training with the iterative denoising nature of dLLMs. [arxiv][code]

[2] Haoyuan Sun, Jing Wang, Yuxin Song, Yu Lu, Bo Fang, **Yifu Luo**, et al. "Power Reinforcement Post-Training of Text-to-Image Models with Super-Linear Advantage Shaping". Under review at *Neural Information Processing Systems* ***(NeurIPS)***, 2026.

– Proposed Super-Linear Advantage Shaping (SLAS) to mitigate the normalization issues in the text-to-image RL post-training. [arxiv]

[3] Xu Yang, Zhizhou Sha, Junbo Li, Jian Yu, Yifan Sun, Matthew Zhao, Jinrui Fang, Xinyue Guo, Yining Wu, Xu Hu, **Yifu Luo**, et al. "No Hidden Prompts Needed! You Can Game AI Peer Review with Presentation-Only Revisions". Under review at *Annual Meeting of the Association for Computational Linguistics* ***(EACL)***, 2027.

– Proposed Super-Linear Advantage Shaping (SLAS) to mitigate the normalization issues in the text-to-image RL post-training. [arxiv][code]

Preprint

[1] Haoyuan Sun, Jiaqi Wu, Bo Xia, **Yifu Luo**, et al. "Reinforcement Fine-Tuning Powers Reasoning Capability of Multimodal Large Language Models". [arxiv]
– Argue reinforcement fine-tuning powers the reasoning capability of multimodal large language models.

LANGUAGE: IELTS 8.0