Zhihe Yang (4th year Ph.D. Candidate in CUHK)

RL for LLM&LMM; Trustworthy offline RL; Diffusion model/Transformer in RL

☎ +852 5930-7245/+86 130-6785-6089 ✉ zhyang@link.cuhk.edu.hk
⌂ Room.302, Academic Building No.1 in CUHK, Sha Tin, Hong Kong SAR, 999077
🎓 Google Scholar 👤 Personal Web Github LinkedIn

Education

Zhejiang University, China **2017 - 2021**
B.E. in Mechanical Engineering *3.87/4.00 (Rank 7/84)*

National University of Singapore, Singapore **2020 - 2022**
M.S. in Mechanical Engineering (Joint Research Program with ZJU) *4.90/5.00 (Rank 1/150)*

The Chinese University of Hong Kong, Hong Kong SAR **2022 - Now**
Ph.D. in Mechanical and Automation Engineering *3.91/4.00 till now*

In-Campus Research Experience

RLH(AI)F Algorithms for LLMs and LMMs **Jan 2025 - Now**

In cooperation with MSRA

- Development of RL algorithms for enhancing reasoning ability of LLMs based on DPO/PPO/GRPO.
- Development of RL algorithms for improving LMMs in radiology report generation (Microsoft-MAIRA).
- Outcomes: Advancing RL4LLMs [C4].

Trustworty Offline Reinforcement Leanring **Aug 2022 - Now**

Supervised by Prof.Yunjian XU *CUHK Automation Engineering*

- Development of algorithms for robust offline reinforcement learning against perturbations on state observations.
- Development of algorithms for safe offline reinforcement learning with hard cost constraints (constrained-MDP).
- Applications of the above algorithms in robot control, smart grid, and EV routing
- Outcomes: Robust Offline RL [C1,P1], Safe Offline RL [C3], Multi-agent RL [P2]

3D Reconstruction of Composites Reinforced by Short Carbon Fibers **Aug 2020 - July 2022**

Supervised by Prof.Wentao YAN *NUS AM Centers*

- Computer vision-based algorithms for the accurate 3D reconstruction of the fiber structure in 3D printed composites.
- Computational Fluid Dynamic (CFD) simulation on Fused-Filament Fabrication (FFF) printing process.
- Outcomes: First paper reporting QA/QLA carbon fiber breakage during 3D printing [J1]

Enterprise Intern Experience

Tencent Hunyuan, Project-Up Talent Program (腾讯混元-青云) **July 2025 - Now**

Supervised by Xin Li *Research Intern at Multimodal Model Department*

- Independently responsible for post-training (RL) of multimodal (video+audio) understanding and captioning. The first version was adopted for pre-training the next-generation movie generation model in Hunyuan.
- Independently responsible for post-training (RL) for AR instruction generation in image-to-video (I2V) tasks.

Microsoft Research Asia (MSRA) **June 2024 - Feb 2025**

Supervised by Xufang LUO, Dongqi HAN, Dongsheng LI *Research Intern at Shanghai AI/ML Group*

- (Principal) Development of RLAI(H)F algorithms to mitigate hallucinations for Multi-modal Large Language Models in general domain (based on DPO, PPO, and GRPO).
- (Assisting) Acceleration for multi-card/node RLAI(H)F training, serving as algorithmic support member for MLSYS.
- Outcomes: On-policy alignment DPO [C2] (**CVPR-Oral**), Co-author paper SortedRL [C5]

Academic Activities

Invited Talks

MegVII Foundation Model Group Seminar: Trustworthy offline RL for embodied AI *June.18 2024*

Serving as Reviewer for

NeurIPS, ICLR, ICML, CVPR, ICCV (starting from 2024/2025)

Publications

* for Co-first author.

Conference Papers

- **[C1]** **Zhihe Yang**, Yunjian Xu. DMBP: Diffusion model-based predictor for robust offline reinforcement learning against state observation perturbations. In *Twelfth International Conference on Learning Representations* (ICLR), 2024. (Poster Presentation 30.9% 2260/7304) [paper] [code]
- **[C2]** **Zhihe Yang**, Xufang Luo, Dongqi Han, Yunjian Xu, Dongsheng Li. Mitigating Hallucinations in Large Vision Language Models via DPO: On-Policy Data Hold the Key. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (CVPR), 2025. (**Oral Presentation 0.74% 96/13008**) [project] [arXiv] [code] [models] [dataset] [MSRA-Feature-Article]
- **[C3]** **Zhihe Yang**, Yunjian Xu, Yang Zhang. Q-Supervised contrastive representation: A state decoupling framework for safe offline reinforcement learning. In *Forty-Second International Conference on Machine Learning* (ICML), 2025. (Poster Presentation 26.9% 3260/12107) [paper] [code]
- **[C4]** **Zhihe Yang**, Xufang Luo, Zilong Wang, Dongqi Han, Zhiyuan He, Dongsheng Li, Yunjian Xu. Do Not Let Low-Probability Tokens Over-Dominate in RL for LLMs. AI4MATH II@ICML2025: *2nd AI for Math Workshop.* [OpenReview (Score 9-7-7-6)] [arXiv] [code]
- **[C5]** Zeyuan Liu*, **Zhihe Yang***, Jiawei Xu*, Rui Yang, Yunjian Xu, Xiu Li. AGD: Ambient Diffusion-Guided Dataset Recovery for Corruption-Robust Offline Reinforcement Learning. In *Thirty-Ninth Annual Conference on Neural Information Processing Systems* NeurIPS, (2025). [arXiv] [OpenReview]
- **[C6]** Yiqi Zhang, Huiqiang Jiang, Xufang Luo, **Zhihe Yang**, Chengruidong Zhang, Yifei Shen, Dongsheng Li, Yuqing Yang, Lili Qiu, Yang You. SortedRL: Accelerating RL Training for LLMs through Online Length-aware Scheduling. ES-FoMo III@ICML2025: *3rd Workshop on Efficient Systems for Foundation Models.* [paper]

Journal Article

- **[J1]** **Zhihe Yang**, Zeshi Yang, Hui Chen, Wentao Yan . 3D Printing of composites reinforced with chopped carbon fiber via material extrusion: fiber breakage. *Additive Manufacturing* 2022: 103067. (IF11.0, CAS-Q1-TOP) [paper]

Preprints

- **[P1]** Yang Zhang, Yunjian Xu, Chengwei Zhang, Chao Wang, **Zhihe Yang**. Policy Consistency in Multi-Agent Reinforcement Learning with Mixed Rewards. *Under review* as a Journal article.

Honors and Awards

Tencent “Project Up” Talent Internship Program (腾讯混元-青云计划)	Year 2025
MSRA “Stars of Tomorrow” Internship Award of Excellence (**Top 10%** Research Intern)	Year 2025
CUHK Postgraduate Scholarship	Year 2022-2026
Singapore PEB Gold Medal (**Best graduate in NUS-ME**, **Highest Distinction**)	Year 2022
Zhejiang University Outstanding Graduate Award (**Top 10%** graduate in ZJU)	Year 2021
Zhejiang University Academic Excellence Award (three times)	Year 2018, 2019, 2020
Jingsheng Electromechanics Scholarship - First Class	Year 2020
Zhejiang University International Engagement Award (MIT summer lab-visiting)	Year 2020
Zhejiang University Scholarship - Second Prize (twice)	Year 2019, 2020
Zhejiang Provincial Government Scholarship	Year 2019
Zhejiang University Scholarship - Third Prize	2018

Teaching Assistant

EEEN2030 Energy Utilization and Human Behavior **2022 term1, 2023 term1, 2024 term1**
Game Theory and Smart Grid

EEEN4060 Energy Distribution **2023 term2, 2024 term2**
Electric Circuit and Foundation of Optimization in Electricity Market

ENGG1130 Multi-variable Calculus for Engineers **2025 term1**
Calculus and its Application in Engineering

Technical Skills

Languages: Mandarin (native proficiency) and English (professional working proficiency)
Programming Languages: Python (PyTorch, Numpy, Pandas, vllm, ray, transformers, nltk, OpenCV, Scikit-learn), MATLAB
Hardware Design: SolidWorks, AutoCAD, Pro/E
Mechanics Analysis: Flow3D, ANYSYS, Avizo