CURRICULUM VITAE

Zhiyuan (Aaron) Feng

Ph.D. Candidate in Computer Science · Tsinghua University

fengzy24@mails.tsinghua.edu.cn aaronfengzy.github.io Google Scholar

EDUCATION

Tsinghua University, Beijing, China 2024 - Present

Ph.D. Candidate in Computer Science, currently in the third year
Chen-Ning Yang Institute for Advanced Study
Advisor: Prof. Baining Guo

Xi'an Jiaotong University, Xi'an, China 2020 - 2024

Bachelor's degree in Computer Science · Qian Xuesen Honors College
Advisors: Bin Shi, Hua Wei, and Qinghua Zheng

RESEARCH INTERESTS

Embodied intelligence through complementary advances in systems, models, and evaluation.

- **Agent harnesses:** task-relevant context organization, grounded reasoning, planning, and executable action generation.
- **Foundation models:** multimodal and vision-language-action learning, scalable pretraining and post-training, spatial reasoning, and generalization.
- **Benchmarks and evaluation:** assessing multimodal understanding, spatial reasoning, and embodied capabilities to identify gaps and guide improvements.

RESEARCH EXPERIENCE

Microsoft Research Asia · Research Intern Sep 2026 - Present

Spatial Intelligence Group · Group lead: Baining Guo
Advisors: Jianfeng Xiang, Zhenhua Liu, Sicheng Xu, and Jiaolong Yang

Research on embodied AI.

Kling AI · Research Intern Jun 2026 - Sep 2026

Foundation Models and Agent Algorithms Department · Department lead: Han Li
Advisors: Zhongxia Chen and Shuo Yang

Research on agent harnesses.

Microsoft Research Asia · Research Intern Jul 2024 - Jun 2026

Spatial Intelligence Group · Group lead: Baining Guo
Advisors: Yu Deng and Jiaolong Yang

Research on embodied AI.

Microsoft Research Asia · Research Intern Dec 2023 - Jun 2024

Vision Computing Group · Group lead: Baining Guo
Advisor: Zheng Zhang

Research on reasoning with large language models.

PROFESSIONAL SERVICE

- **Conference reviewer:** ICLR, NeurIPS, ECCV, CVPR, ICRA, and AAAI.
- **Journal reviewer:** International Journal of Computer Vision (IJCV).
- IEEE Student Member.

SELECTED PUBLICATIONS · 2026

Full author lists are retained; my name is bold. First-author papers are marked explicitly.
* Equal contribution. Includes accepted papers.

[1] **TaskGround: Ground-Infer-Execute for Full-Scene Household Reasoning**

Zhiyuan Feng, Yu Deng, Ruichuan An, Zhenhua Liu, Qixiu Li, Keming Wu, Zhiying Du, Weijie Wang, Haoxiao Wang, Shuang Chen, Sicheng Xu, Yaobo Liang, Jiaolong Yang, Baining Guo.

NeurIPS 2026 First author

[2] **Seeing across views: Benchmarking spatial reasoning of vision-language models in robotic scenes**

Zhiyuan Feng, Zhaolu Kang, Qijie Wang, Zhiying Du, Jiongrui Yan, Shubin Shi, Chengbo Yuan, Huizhi Liang, Yu Deng, Qixiu Li, Rushuai Yang, Arctanx An, Leqi Zheng, Weijie Wang, Shawn Chen, Sicheng Xu, Yaobo Liang, Jiaolong Yang, Baining Guo.

ICLR 2026 First author

[3] **From Human Videos to Robot Manipulation: A Survey on Scalable Vision-Language-Action Learning with Human-Centric Data**

Zhiyuan Feng, Qixiu Li, Huizhi Liang, Rushuai Yang, Yichao Shen, Zhiying Du, Zhaowei Zhang, Yu Deng, Li Zhao, Hao Zhao, Zongqing Lu, Oier Mees, Marc Pollefeys, Jiaolong Yang, Baining Guo.

IJCAI 2026 Survey Track First author

[4] **AD-MIR: Bridging the Gap from Perception to Persuasion in Advertising Video Understanding via Structured Reasoning**

Binxiao Xu, Junyu Feng, Xiaopeng Lin, Haodong Li, **Zhiyuan Feng**, Bohan Zeng, Shaolin Lu, Ming Lu, Qi She, Wentao Zhang.

ICML 2026

[5] **WorldReasonBench: Human-Aligned Stress Testing of Video Generators as Future World-State Predictors**

Keming Wu, Yijing Cui, Wenhan Xue, Qijie Wang, Xuan Luo, **Zhiyuan Feng**, Zuhao Yang, Sudong Wang, Sicong Jiang, Haowei Zhu, Zihan Wang, Ping Nie, Wenhu Chen, Bin Wang.

NeurIPS 2026

[6] **VolSplat: Rethinking Feed-Forward 3D Gaussian Splatting with Voxel-Aligned Prediction**

Weijie Wang*, Yeqing Chen*, Zeyu Zhang, Hengyu Liu, Haoxiao Wang, **Zhiyuan Feng**, Wenkang Qin, Feng Chen, Jia-Wang Bian, Zheng Zhu, Donny Y. Chen, Bohan Zhuang.

ECCV 2026

[7] **HiSpatial: Taming Hierarchical 3D Spatial Understanding in Vision-Language Models**

Huizhi Liang, Yichao Shen, Yu Deng, Sicheng Xu, **Zhiyuan Feng**, Tong Zhang, Yaobo Liang, Jiaolong Yang.

CVPR 2026

[8] **VITRA: Scalable vision-language-action model pretraining for robotic manipulation with real-life human activity videos**

Qixiu Li, Yu Deng, Yaobo Liang, Lin Luo, Lei Zhou, Chengtang Yao, Lingqi Zeng, **Zhiyuan Feng**, Huizhi Liang, Sicheng Xu, Yizhong Zhang, Xi Chen, Hao Chen, Lily Sun, Dong Chen, Jiaolong Yang, Baining Guo.

ICRA 2026

[9] **HSSBench: Benchmarking Humanities and Social Sciences Ability for Multimodal Large Language Models**

Zhaolu Kang, Junhao Gong, Jiaxu Yan, Wanke Xia, Yian Wang, Ziwen Wang, Huaxuan Ding, Zhuo Cheng, Wenhao Cao, **Zhiyuan Feng**, Siqi He, Shannan Yan, Junzhe Chen, Xiaomin He, Chaoya Jiang, Wei Ye, Kaidong Yu, Xuelong Li.

ICLR 2026

SELECTED PUBLICATIONS · EARLIER WORK

[10] **MIRA: Medical Time Series Foundation Model for Real-World Health Data**
Hao Li, Bowen Deng, Chang Xu, **Zhiyuan Feng**, Viktor Schlegel, Yu-Hao Huang, Yizheng Sun, Jingyuan Sun, Kailai Yang, Yiyao Yu, Jiang Bian.
NeurIPS 2025 · Oral presentation at AI4X-AC 2026

[11] **TransDiff: Diffusion-Based Method for Manipulating Transparent Objects Using a Single RGB-D Image**
Haoxiao Wang, Kaichen Zhou, Binrui Gu, **Zhiyuan Feng**, Weijie Wang, Peilin Sun, Yicheng Xiao, Jianhua Zhang, Hao Dong.
ICRA 2025

[12] **Deep evidential learning in diffusion convolutional recurrent neural network**
Zhiyuan Feng, Kai Qi, Bin Shi, Hao Mei, Qinghua Zheng, Hua Wei.
DIT Workshop at CIKM 2022 First author

SELECTED PREPRINTS

[13] **HarnessPro: Recursive Harness Self-Improvement with Pareto Optimization for Performance-Cost Trade-offs**
Zhuoshi Pan*, **Zhiyuan Feng***, Yukuo Cen, Honglin Lin, Junru Lu, Zhongxia Chen, Shuo Yang, H. Vicky Zhao, Jie Tang.
arXiv:2610.26770, 2026 Co-first author

[14] **EnactOpt: Diagnosis-Guided Harness Evolution for Embodied Agents**
Zhiyuan Feng, Zhuoshi Pan, Zhaolu Kang, Zhongxia Chen, Shuo Yang, Sen Ye, Heng Wang, Han Li.
arXiv:2610.14977, 2026 First author

[15] **Discover, Learn, and Reinforce: Scaling Vision-Language-Action Pretraining with Diverse RL-Generated Trajectories**
Rushuai Yang, **Zhiyuan Feng**, Tianxiang Zhang, Kaixin Wang, Chuheng Zhang, Li Zhao, Xiu Su, Yi Chen, Jiang Bian.
arXiv:2511.19528, 2025

[16] **HiMoE-VLA: Hierarchical Mixture-of-Experts for Generalist Vision-Language-Action Policies**
Zhiying Du, Bei Liu, Yaobo Liang, Yichao Shen, Haidong Cao, Xiangyu Zheng, **Zhiyuan Feng**, Zuxuan Wu, Jiaolong Yang, Yu-Gang Jiang.
arXiv:2512.05693, 2025

[17] **Beyond human demonstrations: Diffusion-based reinforcement learning to generate data for vla training**
Rushuai Yang, Hangxing Wei, Ran Zhang, **Zhiyuan Feng**, Xiaoyu Chen, Tong Li, Chuheng Zhang, Li Zhao, Jiang Bian, Xiu Su, Yi Chen.
arXiv:2509.19752, 2025

AWARDS & HONORS

All-Round Excellence Award, Tsinghua University	2026
Microsoft "Stars of Tomorrow" Internship	2024
Outstanding Bachelor Graduate	2024
China Mobile Outstanding Scholarship	2023
National Scholarship	2021, 2022
Asia and Pacific Informatics Olympiad (APIO), Gold Medal	2018
National Olympiad in Informatics in Provinces (NOIP), First Prize	2017, 2018