Zhenfei(Jeremy) Yin

✉ zhenfei.yin@sydney.edu.au ⌂ https://yinzhenfei.github.io

Research interests

Multi-Modal Representation Learning • Multi-Agent System • Embodied AI

Education and Working Experience

2022-Present	**Ph.D.** in the School of Computer Science at the University of Sydney
	Thesis Title: Multi-Modal Foundation Models empowered AI Agents
	Advisor: Wanli Ouyang and Zhiyong Wang
2022-Present	**Rising Star Research Fellow** at the Shanghai AI Lab
	Working on AI Agents
	Mentor: Amanda Shao
2018-2022	**Research Manager** at the AGI Group of SenseTime
	Working on Multi-Modal Foundation Models
	Mentor: Junjie Yan
2015-2019	**B.Sc.** in the School of Software Engineering at HUST
	ACM/ICPC Team Captain
	Advisor: Kun Hu

Publications

Equal contributions are denoted using *, while corresponding authors are indicated using ‡. Project leads are marked with †, which indicates co-advising role without being the corresponding author. As research on foundation models and AI agents often involves systematic collaborative efforts, resulting in some technical reports using alphabetical author ordering. When available, DOIs and links to preprints are provided. Publications are listed in reverse order of their first arXiv submission

2024	Rui Ye*, Xianghe Pang*, Jingyi Chai, Jiaao Chen, **Zhenfei Yin**, Zhen Xiang, Xiaowen Dong, Jing Shao, Siheng Chen‡. *Are We There Yet? Revealing the Risks of Utilizing Large Language Models in Scholarly Peer Review*. Preprint. arXiv: 2412.01708 [cs.CL]
	Ziyi Yang*, Zaibin Zhang*, Zirui Zheng, Yuxian Jiang, Ziyue Gan, Zhiyu Wang, Zijian Ling, Jinsong Chen, Martz Ma, Bowen Dong, Prateek Gupta, Shuyue Hu, **Zhenfei Yin**‡, Guohao Li‡, Xu Jia, Lijun Wang, Bernard Ghanem, Huchuan Lu, Wanli Ouyang, Yu Qiao, Philip Torr, Jing Shao‡. *OASIS: Open Agents Social Interaction Simulations on One Million Agents*. Neural Information Processing Systems Workshop on Open-World Agents, **NeurIPS-W 2024**, arXiv: 2411.11581 [cs.AI]
	Yiran Qin*, Zhelun Shi*, Jiwen Yu, Xijun Wang, Enshen Zhou, Lijun Li, **Zhenfei Yin**†, Xihui Liu, Lu Sheng, Jing Shao‡, Lei Bai‡, Wanli Ouyang, Ruimao Zhang‡. *WorldSimBench: Towards Video Generation Models as World Simulators*. Technical Report. arXiv: 2410.18072 [cs.CV]

Haoyang Su*, Renqi Chen*, Shixiang Tang‡, Xinzhe Zheng, Jingzhe Li, **<u>Zhenfei Yin</u>**, Wanli Ouyang, Nanqing Dong‡. *Two Heads Are Better Than One: A Multi-Agent System Has the Potential to Improve Scientific Idea Generation*. Preprint. arXiv: 2410.09403 [cs.AI]⌐

Yisong Xiao, Aishan Liu, QianJia Cheng, **<u>Zhenfei Yin</u>**, Siyuan Liang, Jiapeng Li, Jing Shao, Xianglong Liu, Dacheng Tao. *GenderBias-VL: Benchmarking Gender Bias in Vision Language Models via Counterfactual Probing*. Preprint. arXiv: 2407.00600 [cs.CV]⌐

Yongting Zhang*, Lu Chen*, Guodong Zheng, Yifeng Gao, Rui Zheng, Jinlan Fu, **<u>Zhenfei Yin</u>**, Senjie Jin, Yu Qiao, Xuanjing Huang, Feng Zhao, Tao Gui‡, Jing Shao‡. *SPA-VL: A Comprehensive Safety Preference Alignment Dataset for Vision Language Model*. Preprint. arXiv: 2406.12030 [cs.CV]⌐

Zeren Chen*, Zhelun Shi*, Xiaoya Lu*, Lehan He*, Sucheng Qian, **<u>Zhenfei Yin</u>**†, Wanli Ouyang, Jing Shao‡, Yu Qiao, Cewu Lu, Lu Sheng‡. *RH20T-P: A Primitive-Level Robotic Manipulation Dataset Towards Composable Generalization Agents in Real-world Scenarios*. Neural Information Processing Systems Workshop on Open-World Agents, **NeurIPS-W 2024**, arXiv: 2403.19622 [cs.RO]⌐

Zhelun Shi*, Zhipin Wang*, Hongxing Fan*, Zaibin Zhang, Lijun Li, Yongting Zhang, **<u>Zhenfei Yin</u>**, Lu Sheng‡, Yu Qiao, Jing Shao‡. *Assessment of Multimodal Large Language Models in Alignment with Human Values*. Preprint. arXiv: 2403.17830 [cs.CV]⌐

Enshen Zhou*, Yiran Qin*, **<u>Zhenfei Yin</u>**, Yuzhou Huang, Ruimao Zhang‡, Lu Sheng‡, Yu Qiao, Jing Shao†. *Chain-of-Imagination for Reliable Instruction Following in Decision Making*. Neural Information Processing Systems Workshop on Open-World Agents, **NeurIPS-W 2024**, arXiv: 2403.12037 [cs.CV]⌐

Chen Qian*, Jie Zhang*, Wei Yao*, Dongrui Liu, **<u>Zhenfei Yin</u>**, Yu Qiao, Yong Liu‡, Jing Shao‡. *Towards Tracing Trustworthiness Dynamics: Revisiting Pre-training Period of Large Language Models*. Findings of the Association for Computational Linguistics, **ACL 2024**, pp. 4864-4888. doi: 10.18653/v1/2024.findings-acl.290⌐. arXiv: 2402.19465 [cs.CL]⌐

Chaochao Lu, Chen Qian, Guodong Zheng, Hongxing Fan, Hongzhi Gao, Jie Zhang, Jing Shao‡, Jingyi Deng, Jinlan Fu, Kexin Huang, Kunchang Li, Lijun Li, Limin Wang, Lu Sheng, Meiqi Chen, Ming Zhang, Qibing Ren, Sirui Chen, Tao Gui, Wanli Ouyang, Yali Wang, Yan Teng, Yaru Wang, Yi Wang, Yinan He, Yingchun Wang, Yixu Wang, Yongting Zhang, Yu Qiao‡, Yujiong Shen, Yurong Mou, Yuxi Chen, Zaibin Zhang, Zhelun Shi, **<u>Zhenfei Yin</u>**†, Zhipin Wang. *From GPT-4 to Gemini and Beyond: Assessing the Landscape of MLLMs on Generalizability, Trustworthiness and Causality through Four Modalities*. Technical Report. arXiv: 2401.15071 [cs.CV]⌐

2023

Zhiyuan You*, Zheyuan Li*, Jinjin Gu*, **<u>Zhenfei Yin</u>**, Tianfan Xue‡, Chao Dong‡. *Depicting beyond scores: Advancing image quality assessment through multi-modal language models*. European Conference on Computer Vision, **ECCV 2024**, pp. 259-276. doi: 10.1007/978-3-031-72970-6_15⌐. arXiv: 2312.08962 [cs.CV]⌐

Yiran Qin*, Enshen Zhou*, Qichang Liu*, **<u>Zhenfei Yin</u>**, Lu Sheng‡, Ruimao Zhang‡, Yu Qiao, Jing Shao†. *MP5: A Multi-Modal Open-Ended Embodied System in Minecraft via Active Perception*. IEEE/CVF Conference on Computer Vision and Pattern Recognition, **CVPR 2024**, pp. 16307-16316. doi: 10.1109/CVPR52733.2024.01543⌐. arXiv: 2312.07472 [cs.CV]⌐

Zhelun Shi*, Zhipin Wang*, Hongxing Fan*, **<u>Zhenfei Yin</u>**, Lu Sheng‡, Yu Qiao, Jing Shao‡. *ChEF: A Comprehensive Evaluation Framework for Standardized Assessment of Multimodal Large Language Models*. Technical Report. arXiv: 2311.02692 [cs.CV]⌐

Zeren Chen*, Ziqin Wang*, Zhen Wang, Huayang Liu, **Zhenfei Yin**[†], Si Liu, Lu Sheng[‡], Wanli Ouyang, Jing Shao[‡]. ***Octavius: Mitigating Task Interference in MLLMs via LoRA-MoE***. International Conference on Learning Representations, **ICLR 2024**, arXiv: 2311.02684 [cs.CV]⧉

Zhenfei Yin*, Jiong Wang*, Jianjian Cao*, Zhelun Shi*, Dingning Liu, Mukai Li, Xiaoshui Huang, Zhiyong Wang, Lu Sheng, Lei Bai[‡], Jing Shao[‡], Wanli Ouyang. ***LAMM: Language-Assisted Multi-Modal Instruction-Tuning Dataset, Framework, and Benchmark***. Advances in Neural Information Processing Systems, **NeurIPS 2023**, Vol. 36. arXiv: 2306.06687 [cs.CV]⧉

2022   Yuan Yao, Yuanhan Zhang, **Zhenfei Yin**, Jiebo Luo, Wanli Ouyang, Xiaoshui Huang[‡]. ***3D Point Cloud Pre-Training with Knowledge Distilled from 2D Images***. IEEE International Conference on Multimedia and Expo, **ICME 2024**, pp. 1-6. doi: 10.1109/ICME57554.2024.10687626⧉. arXiv: 2212.08974 [cs.CV]⧉

Yuanhan Zhang, **Zhenfei Yin**[†], Jing Shao[‡], Ziwei Liu. ***Benchmarking omni-vision representation through the lens of visual realms***. European Conference on Computer Vision, **ECCV 2022**, pp. 594-611. doi: 10.1007/978-3-031-20071-7_35⧉. arXiv: 2207.07106 [cs.CV]⧉

Yinan He*, Gengshi Huang*, Siyu Chen*, Jianing Teng*, Kun Wang, **Zhenfei Yin**, Lu Sheng, Ziwei Liu, Yu Qiao, Jing Shao[‡]. ***X-learner: Learning cross sources and tasks for universal visual representation***. European Conference on Computer Vision, **ECCV 2022**, pp. 509-528. doi: 10.1007/978-3-031-19809-0_29⧉. arXiv: 2203.08764 [cs.CV]⧉

Yuanhan Zhang*, Qinghong Sun*, Yichun Zhou*, Zexin He*, **Zhenfei Yin**[†], Kun Wang, Lu Sheng, Yu Qiao, Jing Shao[‡], Ziwei Liu. ***Bamboo: Building Mega-Scale Vision Dataset Continually with Human-Machine Synergy***. Preprint. arXiv: 2203.07845 [cs.CV]⧉

2021   Yujie Wang*, Junqin Huang*, Mengya Gao*, Yichao Wu*, **Zhenfei Yin**, Ding Liang, Junjie Yan. ***One to Transfer All: A Universal Transfer Framework for Vision Foundation Model with Few Data***. Technical Report. arXiv: 2111.12386 [cs.CV]⧉

Jing Shao*, Siyu Chen*, Yangguang Li*, Kun Wang*, **Zhenfei Yin***, Yinan He*, Jianing Teng*, Qinghong Sun*, Mengya Gao*, Jihao Liu*, Gengshi Huang*, Guanglu Song, Yichao Wu, Yuming Huang, Fenggang Liu, Huan Peng, Shuo Qin, Chengyu Wang, Yujie Wang, Conghui He, Ding Liang, Yu Liu, Fengwei Yu, Junjie Yan, Dahua Lin, Xiaogang Wang, Yu Qiao[‡]. ***INTERN: A New Learning Paradigm Towards General Vision***. Technical Report. arXiv: 2111.08687 [cs.CV]⧉

Honours

2022   Rising Star at the Shanghai AI Lab

2020   4th Place Winner in Meta's DFDC Challenge

2018   Student coach for the team that reached the ACM-ICPC World Finals, finishing 31st globally

Invited Talks & Teaching

2024   1. *Building AI Society with Foundation-Model Agents.*

Workshop on Open-World Agents⧉ at NeurIPS 2024

2. *Multi-Modal Foundation Models and Applications as Embodied Agents*

Guest Lecture, ELEC5304: Intelligent Visual Signal Understanding, USYD

| 2024 Spring | 3. *Multimodal Applications* |
| | Teaching Assistant, COMP 5425: Multimedia Retrieval, USYD |

Workshop & Academic Event Organization

2024	1. Workshop Organizer, ICML 2024 workshop on Multi-modal Foundation Model meets Embodied AI, MFM-EAI
	2. Workshop Organizer, ICML 2024 workshop on Trustworthy Multi-modal Foundation Models and AI Agents, TiFA
Since 2023	3. Academic Talks Organizer, Echo AI Talk

Programme committee memberships

Conferences

International Conference on Learning Representations (ICLR)	Since 2023
Neural Information Processing Systems (NeurIPS)	Since 2023
International Conference on Machine Learning (ICML)	Since 2024
Association for the Advancement of Artificial Intelligence (AAAI)	Since 2023
European Conference on Computer Vision (ECCV)	Since 2022
IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)	Since 2022

Journals

IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)	Since 2023