Jiashu Xu 徐家澍

Santa Clara, CA | 949-522-2936 | 1999j0615une@gmai.com | Website | Scholar | cnut1648 | jiashu-xu

Research Interests

My current research interests lie in **generative models' post-training**. Particularly,

1. Post-training diffusion model ([1], [2], [4], [11]) and VLM ([3], [5]) w/ SFT and RL ([6], [7])
2. Reliable model against malicious exploitation of LLM vulnerabilities ([8] and [12] to [15]).
3. Excels in low-resource regimes, through indirect supervision ([16], [20]) or synthetic data generation ([9], [10], [17], [19]).

Affiliation

NVIDIA Research — Santa Clara, CA
Research Scientist; Deep Imagination Research Group (Cosmos) — *Fall 2024 – Present*

Harvard University — Cambridge, MA
M.S. in Computational Science and Engineering; cross-registered at MIT; GPA: 4.0/4.0 — *Fall 2022 – Spring 2024*

University of Southern California — Los Angeles, CA
B.S. double major in Applied Math & Computer Science; Summa Cum Laude*; GPA: 3.97/4.0* — *Fall 2020 – Spring 2022*

Publications & Services

*=Equal Contribution

[1] **World Simulation with Video Foundation Models for Physical AI**
NVIDIA (core contributor)
Arxiv, 2025 — code paper project benchmark

[2] **Cosmos World Foundation Model Platform for Physical AI**
NVIDIA (core contributor)
Arxiv, 2025 (**CES'25 Best of AI, Best Overall**) — code paper project video Jensen's Keynote

[3] **Cosmos-Reason1: From Physical Common Sense To Embodied Reasoning**
NVIDIA (core contributor)
Arxiv, 2025 — code paper project Jensen's Keynote

[4] **Cosmos-Transfer1: Conditional World Generation with Adaptive Multimodal Control**
NVIDIA
Arxiv, 2025 — code paper project

[5] **GenUSD: 3D Scene Generation Made Easy**
Jiashu Xu, Chen-Hsuan Lin, Donglai Xiang, Yunhao Ge, Xiaohui Zeng, Qianli Ma, Yifan Ding, Zekun Hao, Fangyin Wei, Yin Cui, Zhaoshuo Li, J.P. Lewis, Qinsheng Zhang, Jingyi Jin, Pooya Jannaty, Seungjun Nah, Hanzi Mao, Tsung-Yi Lin, Arun Mallya, Yen-Chen Lin, Ming-Yu Liu
SIGGRAPH, 2024 (**Real-Time Live!**) — project video

[6] **Data-regularized Reinforcement Learning for Diffusion Models at Scale**
Haotian Ye, Kaiwen Zheng, **Jiashu Xu**, Puheng Li, Huayu Chen, Jiaqi Han, Sheng Liu, Qinsheng Zhang, Hanzi Mao, Zekun Hao, Prithvijit Chattopadhyay, Dinghao Yang, Liang Feng, Maosheng Liao, Junjie Bai, Ming-Yu Liu, James Zou, Stefano Ermon
Arxiv, 2025 — paper project

[7] **Training Large Language Models as Reward Models**
Jiashu Xu, Daniel Pressel, Prasoon Goyal, Luke Dai, Reza Ghanadan, Michael Johnston
Amazon Internal Project, 2024

[8] **Instructional Fingerprinting of Large Language Models**
Jiashu Xu, Fei Wang*, Mingyu Derek Ma*, Pang Wei Koh, Chaowei Xiao, Muchao Chen
NAACL, 2024 (**Oral**) — code paper project

[9] **BEHAVIOR Vision Suite: Customizable Dataset Generation via Simulation**
Yunhao Ge*, Yihe Tang*, **Jiashu Xu***, Cem Gokmen*, Chengshu Li, Wensi Ai, Benjamin Jose Martinez, Arman Aydin, Mona Anvari, Ayush K Chakravarthy, Hong-Xing Yu, Josiah Wong, Sanjana Srivastava,

Sharon Lee, Shengxin Zha, Laurent Itti, Yunzhu Li, Roberto Martín-Martín, Miao Liu, Pengchuan Zhang, Ruohan Zhang, Fei-Fei Li, Jiajun Wu
CVPR, 2024 (**Highlight**) code paper project

[10] **DreamDistribution: Prompt Distribution Learning for Text-to-Image Diffusion Models**
Brian Nlong Zhao, Yuhang Xiao*, **Jiashu Xu***, Xinyang Jiang, Yifan Yang, Dongsheng Li, Laurent Itti, Yunhao Ge, Vibhav Vineet
ICLR, 2025 code paper project

[11] **Edify 3D: Scalable High-Quality 3D Asset Generation**
NVIDIA
Arxiv, 2024 paper project video

[12] **Mitigating Backdoor Threats to Large Language Models: Advancement and Challenges**
Qin Liu, Wenjie Mo, Terry Tong, **Jiashu Xu**, Fei Wang, Chaowei Xiao, Muhao Chen
Allerton Conference, 2024 paper

[13] **Securing Multi-turn Conversational Language Models Against Distributed Backdoor Triggers**
Terry Tong, **Jiashu Xu**, Qin Liu, Muhao Chen
EMNLP, 2024 code paper

[14] **Instructions as Backdoors: Backdoor Vulnerabilities of Instruction Tuning for Large Language Models**
Jiashu Xu, Mingyu Derek Ma, Fei Wang, Chaowei Xiao, Muhao Chen
NAACL, 2024 paper project

[15] **Test-time Backdoor Mitigation for Black-Box Large Language Models with Defensive Demonstrations**
Wenjie Mo, **Jiashu Xu**, Qin Liu, Jiongxiao Wang, Jun Yan, Chaowei Xiao, Muhao Chen
NAACL, 2025 paper

[16] **Can NLI Provide Proper Indirect Supervision for Low-resource Biomedical Relation Extraction?**
Jiashu Xu, Mingyu Derek Ma, Muhao Chen
ACL, 2023 (**Oral**) code paper

[17] **Dall-e for detection: Language-driven context image synthesis for object detection**
Yunhao Ge*, **Jiashu Xu***, Brian Nlong Zhao, Neel Joshi, Laurent Itti, Vibhav Vineet
arXiv, 2022 code paper extension

[18] **X-Norm: Exchanging Normalization Parameters for Bimodal Fusion**
Yufeng Yin*, **Jiashu Xu***, Tianxin Zu, Mohammad Soleymani
ICMI, 2022 paper

[19] **Neural-Sim: Learning to Generate Training Data with NeRF**
Yunhao Ge, Harkirat Behl*, **Jiashu Xu***, Suriya Gunasekar, Neel Joshi, Yale Song, Xin Wang, Laurent Itti, Vibhav Vineet
ECCV, 2022 code paper

[20] **Unified Semantic Typing with Meaningful Label Inference**
James Y. Huang, Bangzheng Li*, **Jiashu Xu***, Muhao Chen
NAACL, 2022 code paper

[21] **Dissection Gesture Sequence during Nerve Sparing Predicts Erectile Function Recovery after Robot-Assisted Radical Prostatectomy**
Runzhuo Ma, **Jiashu Xu**, Ivan Rodriguez, Gina DeMeo, Aditya Desai, Loc Trinh, Jessica H. Nguyen, Anima Anandkumar, Jim C. Hu, Andrew J. Hung
NPJ Digital Medicine, 2022 paper

[22] **Dissection Assessment for Robotic Technique (DART) to Evaluate Nerve-Spare of Robot-Assisted Radical Prostatectomy**
Runzhuo Ma, Alvin Hui, **Jiashu Xu**, Aditya Desai, Michael Tzeng, Emily Cheng, Loc Trinh, Jessica H. Nguyen, Anima Anandkumar, Jim C. Hu, Andrew J. Hung
American Urological Association Annual Conference (AUA), 2022 paper

[23] **SalKG: Learning From Knowledge Graph Explanations for Commonsense Reasoning**
Aaron Chan, **Jiashu Xu**, Boyuan Long, Soumya Sanyal, Tanishq Gupta, Xiang Ren
NeurIPS, 2021 code paper

Research Experience

NVIDIA Research Santa Clara, USA
Research Scientist | Manager: Ming-Yu Liu *Summer 2024 – Present*

- Pretrained and Posttrained diffusion models and VLMs that excel at Physical AI domains [1] to [4], [6], [11].
- Created agentic pipeline to create 3D layout generation [5].

Amazon Science New York, USA
Applied Scientist Intern | Manager: Daniel Pressel, Michael Johnston *Summer 2023*

- Collaborated closely with the LLM team on the reward modeling side.
- Finetuned LLMs directly as reward models such that models learn to align with human preferences implicitly. Further benefits included zero-shot generalization to unseen dimensions and domains, high-quality data filtering, rationale generation to explain decisions, and synthetic conversation curation for AI self-improvement (RLAIF) [7].

Stanford SVL Group Palo Alto, USA
Research Assistant | Advisor: Prof. Jiajun Wu *Fall 2023 – Present*

- Developed BEHAVIOR Vision Suite, a customizable dataset generator featuring photorealistic assets and physically plausible annotations. Demonstrated applications include holistic benchmarks for 2D and 3D vision models, robustness evaluation through parametric out-of-distribution evaluation (*e.g.* low lighting, extreme camera pose), and synthetic dataset generation to bolster performance in low-resource scenarios [9].

Microsoft Research Los Angeles, USA
Student Collaborator | Manager: Prof. Laurent Itti, Vibhav Vineet *Spring 2022 – Present*

- Proposed prompt distribution learning for text-to-image and text-to-3D diffusion models to lightweight control image quality and diversity [10].
- Utilized diffusion models and object cut-and-paste to create coherent synthetic training datasets for enhancing low-resource object detection and segmentation [17]. And proposed differentiable synthetic dataset generation with NeRF to improve out-of-distribution object detection of varying views [19].