Zizhao Wang

Contact zizhao.wang@utexas.edu
Website https://wangzizhao.github.io/
Google scholar https://tinyurl.com/zizhaowangscholar

Research focus

LLM Post-Training, World Models, Reinforcement Learning, Causal Reasoning

Skills

- LLM: RL post-training (PPO, GRPO), LLM agents, reasoning, safety, tool use
- decision making: model-based RL, offline RL, hierarchical RL, imitation learning, planning
- autonomous driving: motion prediction
- development: Python, ML frameworks (PyTorch, TensorFlow, Transformers), distributed training (deepspeed), efficient training (PEFT), deployment (vLLM), simulation (Mujoco), data structure, algorithms
- deep learning: representation learning, generalization and robustness, interpretability and explainability, probabilistic graphical models

Education

2020 - 25 **PhD**, Electrical and Computer Engineering, GPA: 4.00/4.00 — **University of Texas at Austin**
Expected graduation: 2025/12
2018 - 19 **MS**, Computer Science, GPA: 4.00/4.00 — **Columbia University**
2016 - 18 **BS**, Computer Engineering, GPA: 3.96/4.00 — **University of Michigan**
2014 - 18 **BS**, Electrical and Computer Engineering, GPA: 3.72/4.00 — **Shanghai Jiao Tong University**

Work Experience

2025/03 **Student Researcher** — **Google**

- Designed an adversarial **RL post-training** framework that enhance **LLM agent** security again prompt injections, by co-training two LLMs: an attacker that learns to create diverse prompt injections and an agent that learns to defend against them.
- Implemented the data collection pipeline with vLLM and parallel simulation environments for fast LLM agent rollout inference.
- Fine-tuned the LLM model with the GRPO algorithm, implemented with Transformer, deepspeed, and LoRA for fast and memory-efficient training.
- Reduced the attack success rate by 21% and improve task success rate by 18% compared to the untrained model.

2024/06 **Research Intern** — **Microsoft Research**

- Designed an **generative world model** that can synthesize images of novel scenarios, by using object-centric representations and disentangled representations.
- Enhanced the generalization of **reinforcement learning** policies by 30%, when learning with generated out-of-distribution data.

2024/01 **Research Intern** — **Honda Research Institute**

- Developed a **motion prediction** algorithm for **autonomous driving** that, reduced prediction error by 48%, by applying **causal reasoning** to vehicle interactions.
- Sped up model training with **distributed training** and **efficient CUDA implementations** for sparse attention.

Research Experience

2021 - 22 **Causal World Model (ICML oral, AAAI oral)** **University of Texas at Austin**

- Developed a **motion prediction** algorithm for **autonomous driving** that reduced prediction error by 48%, by applying **causal reasoning** to vehicle interactions.
- Built a novel **transformer**-based model for vehicle interaction reasoning, improving reasoning performance (vehicle interaction detection accuracy) by 10%.
- Sped up model training with **distributed training** and **efficient CUDA implementations** for sparse attention.

2022 - 23 **Unsupervised Skill Learning (NeurIPS)** **University of Texas at Austin**

- Proposed a skill discovery method for **structured decision-making** tasks, where reusable skills are learned to induce interactions between state factors.
- Implemented a novel **hierarchical RL** algorithm for skill learning in PyTorch – the high-level policy selects the interaction to induce and the low-level policy learns to induce it using primitive actions.
- Enhanced skill diversity and downstream task performance on long-horizon robotics tasks and structured decision-making tasks by 40%.

2018 - 19 **Reinforcement Learning from Human Feedback (ICRA, IROS)** **Columbia University**

- Proposed a skill discovery method for **structured decision-making** tasks, where reusable skills are learned to induce interactions between state factors.
- Implemented a novel **hierarchical RL** algorithm for skill learning in PyTorch – the high-level policy selects the interaction to induce and the low-level policy learns to induce it using primitive actions.
- Enhanced skill diversity and downstream task performance on long-horizon robotics tasks and structured decision-making tasks by 40%.

Selected Publications

See google scholar (https://tinyurl.com/zizhaowangscholar) for a complete list of publications.

- Adversarial Reinforcement Learning for LLM Agent Safety, *In submission*
  Z Wang, D Li, V Keshava, P Wallis, A Balashankar, P Stone, L Rutishauser.
- SkiLD: Unsupervised Skill Discovery Guided by Local Dependencies, *NeurIPS 2024*
  Z Wang*, J Hu*, C Chuck*, S Chen, R Martín-Martín, A Zhang, S Niekum, P Stone.
- Building Minimal and Reusable Causal State Abstractions for Reinforcement Learning, *AAAI 2024 (**oral**)*
  Z Wang*, C Wang, X Xiao, Y Zhu, and P Stone.
- ELDEN: Exploration via Local Dependencies, *NeurIPS 2023*
  Z Wang*, J Hu*, R Martín-Martín, and P Stone.
- Causal Dynamics Learning for Task-Independent State Abstraction (**Oral**), *ICML 2022 (**oral**)*
  Z Wang, X Xiao, Z Xu, Y Zhu, and P Stone.
- Learning to Correct Mistakes: Backjumping in Long-horizon Task and Motion Planning, *CoRL 2022*
  Y Sung*, **Z Wang***, and P Stone.
- From Agile Ground to Aerial Navigation: Learning from Learned Hallucination, *IROS 2021*
  Z Wang, X Xiao, A Nettekoven, K Umasankar, A Singh, S Bommakanti, U Topcu, and P Stone.
- APPLE: Adaptive Planner Parameter Learning from Evaluative Feedback, *RAL 2021*
  Z Wang, X Xiao, G Warnell, and P Stone.
- Maximizing BCI Human Feedback using Active Learning, *IROS 2020*
  Z Wang*, J Shi*, I Akinola*, and P Allen.
- Accelerated Robot Learning via Human Brain Signals, *ICRA 2020*.
  I Akinola*, **Z Wang***, J Shi, X He, P Lapborisuth, J Xu, D Watkins-Valls, P Sajda, and P Allen.