Benhao Huang | 黄奔皓

Benhao Huang

benhaoh@andrew.cmu.edu https://huskydoge.github.io

EDUCATION

Carnegie Mellon University, Pittsburgh, PA, USA **2025.7 - Present**
MS in Machine Learning (Advisor: Prof. Zico Kolter)

Shanghai Jiao Tong University, Shanghai, China **2021 - 2025**
B.Eng. in Computer Science and Technology (IEEE Honor Class)

- **GPA: 93.14/100.00, 4.08/4.30**

RESEARCH INTERESTS

My primary research goal is to build autonomous AI agents with human-level reasoning, planning, and memory that can **flexibly allocate computation** under varying budget constraints and adapt to evolving environments through **closed-loop interaction**. I am particularly interested in **Scalable Reasoning** (methods that improve both capability and efficiency), and **World Modeling** (**video generative models for long horizon prediction, control, and planning**).

SELECTED PROJECTS

[1] Equilibrium Reasoners: Learning Attractors Enables Scalable Reasoning

Preprint [**Paper**] [**Project**]

Benhao Huang, *Zhengyang Geng, Zico Kolter*

- Proposed an attractor-based framework for iterative reasoning, **casting reasoning as convergence to task-conditioned fixed points** in a latent dynamical system.
- Demonstrated reliable two-axis test-time scaling via depth (solver steps) and width (stochastic restarts), boosting accuracy from **8% to over 99%** on Sudoku-Extreme, enabled by **training interventions that reshape attractor landscapes** for stable convergence through randomness injections.

[2] Flow Equivariant World Models: Structured Dynamics Outside the Field of View

NeurIPS 2025 Workshops: NeurReps, SpaVLE [Paper] [Project]

Hansen Lillemark, **Benhao Huang***, Fangneng Zhan, Yilun Du, T. Anderson Keller† (*Equal Contribution)*

- Introduced **FloWM**, an action-conditioned video world model that **predicts partially observable dynamics** within its memory representation using **flow and self-motion equivariance**.
- Outperformed DFoT and SSM on long horizon video prediction, maintaining consistent memory of dynamics over 200 frames.

[3] PAN: A World Model for General, Interactable, and Long-Horizon World Simulation

Technical Report [Paper] [Project]

Jiannan Xiang, Yi Gu*, Zihan Liu*, Zeyu Feng*, Qiyue Gao*, Yiyan Hu*, **Benhao Huang***, Guangyi Liu*, Yichi Yang*, Kun Zhou*, Davit Abrahamyan, Arif Ahmad, Ganesh Bannur, Junrong Chen, Kimi Chen, Mingkai Deng, Ruobing Han, Xinqi Huang, Haoqiang Kang, Zheqi Li, Enze Ma, Hector Ren, Yashowardhan Shinde, Rohan Shingre, Ramsundar Tanikella, Kaiming Tao, Dequan Yang, Xinle Yu, Cong Zeng, Binglin Zhou, Hector Liu†, Zhiting Hu†, Eric P. Xing† (***Core Contributor**)*

- PAN is a **large-scale interactive** video world model for **open-domain, action-conditioned simulation**, combining an autoregressive backbone for **long-horizon temporal consistency** with a diffusion-based decoder for **short-term fidelity**, enabled by sliding-window causal chunk-wise attention.
- Achieved **state of the art performance** on planning, prediction, and controllability benchmarks, significantly outperforming prior world models including V-JEPA, Cosmos, and WAN by up to **20%**.
- **Co-led large-scale data curation, filtering, and captioning pipelines**; accelerated early **training and evaluation**; influenced key architecture and experiment choices.

[4] DCA-Bench: A Benchmark for Dataset Curation Agents

KDD 2025 [Paper]

Benhao Huang, *Yingzhuo Yu, Jin Huang, Xingjian Zhang, Jiaqi Ma†*

- Introduced a novel task for LLM agents: **detecting dataset quality issues for the purpose of automating AI training data curation.** We collected numerous instances and designed a four-level difficulty for fine-grained analysis of frontier models.
- Developed a LLM-based automatic evaluator for scalable evaluations, which is shown to be **robust to self-preference or length bias** through comprehensive experiments.

HONOURS AND AWARDS

- **National Scholarship.** ***Highest honor for undergraduates*** (**2024**)
- **Ruiyuan-Hongshan Scholarship.** (**Top 2% university-wide, 2023**)
- **Shaoqiu Scholarship.** (**2022**)

LEADERSHIP AND SERVICES

Reviewer, ICML 2026, NeurIPS 2025, KDD 2025
Student Mentor, Shanghai Jiao Tong University, CS2612 and CS2601, 2023
Volunteer, Shanghai Marathon, 2022 - 2024

SKILLS

Programming: Python, PyTorch, C/C++, TorchXLA, Slurm
Languages: English (**TOEFL**: 107, S24, R27, L29, W27, **GRE**: V157, Q170, AW4.0)