Haoyu Li

Phone: (424) 293-9235 | Website: haoyuli02.github.io | Email: haoyuli5@illinois.edu

Education

University of Illinois, Urbana-Champaign (UIUC) Aug 2024 - May 2029 (Expected)
PhD in Computer Science
Advisor: Huan Zhang

University of California, Los Angeles (UCLA) Sep 2020 - Jun 2024
*B.S. in Mathematics (*GPA: 3.93/4.00)

Research Interests

- Foundation Models, LLM Post-Training, LLM Reasoning
- Learning Based Control, Formal Verification of Neural Networks

Publications & Preprints

- Wei Shen*, Han Wang*, Haoyu Li*, Huan Zhang, "DecepChain: Inducing Deceptive Reasoning in Large Language Models", Under review 2025 [Paper][Code]
- Han Wang*, **Haoyu Li***, Brian Ko*, Huan Zhang, "On The Fragility of Benchmark Contamination Detection in Reasoning Models", Under review 2025 [Paper][Code]
- Kairun Zhang*, **Haoyu Li***, Yanjun Zhao*, Yifan Sun, Huan Zhang, "Learning to Learn a Zeroth-Order Optimizer for Fine-tuning LLMs", Under review 2025 [Paper][Code]
- **Haoyu Li***, Xiangru Zhong*, Bin Hu, Huan Zhang, "Two-Stage Learning of Stabilizing Neural Controllers via Zubov Sampling and Iterative Domain Expansion", **NeurIPS 2025 (Spotlight)** [Paper][Code]
- Mohamed Serry*, **Haoyu Li***, Ruikun Zhou*, Huan Zhang, Jun Liu, "Safe Domains of Attraction for Discrete-Time Nonlinear Systems: Characterization and Verifiable Neural Network Estimation", **CDC 2025** [Paper][Code]
- **Haoyu Li***, Xiangru Zhong*, Bin Hu, Huan Zhang, "Neural Contraction Metrics with Formal Guarantees for Discrete-Time Nonlinear Dynamical Systems", **L4DC 2025** [Paper]
- Derek Xu*, Tong Xie*, Botao Xia*, **Haoyu Li***, Yunsheng Bai, Yizhou Sun, Wei Wang, "Does Few-Shot Learning Help LLM Performance in Code Synthesis?", Preprint 2024 [Paper]
- **Haoyu Li***, Shichang Zhang*, Longwen Tang, Matheiu Bauchy, Yizhou Sun, "Predicting and Interpreting Energy Barriers of Metallic Glasses with Graph Neural Networks", **ICML 2024** [Paper][Code]
- Tong Xie*, **Haoyu Li***, Andrew Bai, Cho-Jui Hsieh, "Interpretability through Training Samples: Data Attribution for Diffusion Models", **TMLR 2024** [Paper][Code]

Research Experiences

UIUC **Champaign, IL**
Large Language Models Sep 2024 - Present
Advisor: Prof. Huan Zhang

- *DecepChain (co-first author)*: Introduced a backdoor that makes CoT look benign while flipping the final answer, by exploiting LLM's own hallucination with SFT on self-generated wrong rollouts and GRPO with a flipped verifiable reward; achieves >95% attack success and non-differentiable human trust compared to the benign case.

- *Reasoning model contamination (co-first author)*: Showed that even brief GRPO can conceal contamination signals introduced during SFT contamination; proposed theoretical results that pin the effect on PPO-style importance-sampling/clipping.
- *ZO-Finetuner (co-first author)*: Proposed a compact learned zeroth-order optimizer that learns perturbation strategies once per LLM and transfers across tasks; outperforms previous ZO baselines in 4 LLMs × 7 datasets across model sizes with minimal time/memory overhead.

UIUC **Champaign, IL**
Learning-Based Control & Formal Verification Sep 2024 - Present
Advisor: Prof. Huan Zhang, Prof. Bin Hu

- *Two-Stage Neural Controller (first author, NeurIPS'25 Spotlight)*: Proposed Zubov-inspired data sampling + iterative domain expansion for training, and a strengthened α,β-CROWN pipeline for fast continuous-time verification; yields ROA volumes 5-1.5*10^5 times larger than baselines and 40-10,000 times faster verification than dReal;
- *Neural Contraction Metrics (first author, L4DC'25)*: Proposed a new Jacobian-/LMI-free sufficient condition for contraction in discrete-time systems, enabling scalable certification with non-smooth neural network controllers;

Awards

- NeurIPS 2025 Scholar Award
- First place in the 6th International Verification of Neural Networks Competition (VNN-COMP 2025) for both the regular and extended tracks. Member of team alpha-beta-CROWN.
- L4DC 2025 Travel Grant

Services

- Reviewer for NeurIPS 2025, ICLR 2025-2026, L4DC 2025-2026

Skill Sets

Python, Pytorch, Hugging Face, verl, vLLM, DeepSpeed, FSDP, C/C++, Git