Chenxi Yang

cxyang@cs.utexas.edu | chenxi-yang.github.io | linkedin.com/in/chenxi-yang-ut | +1 (512) 960-6965

Education

The University of Texas at Austin

Aug 2019 – Mar 2025 (expected)

PhD in Computer Science, Advisor: Swarat Chaudhuri

Austin, TX

Fudan University

Sep 2015 – Jul 2019

BSc in Computer Science (Honor Program), Advisor: Prof. Yang Chen

Shanghai

Expertise

- Programming Languages: Python, C/C++, SQL, Java, Javascript, ACL2
- Machine Learning: PyTorch, TensorFlow, PyTorch Lightning, Keras, Scikit-Learn
- Technical: Algorithms & Data Structures, ML Systems (Software & Hardware), Artificial Intelligence, RL, Formal Verification

Work Experience

Google

PhD Intern, System Research Group (Hosts: Yawen Wang, Martin Maas)

May 2024 - Aug 2024, Seattle, WA

- Designed and implemented **the first lightweight Tensor Processing Unit (TPU) scheduling simulator** supporting various existing TPU scheduling algorithms. The simulator automated ML job scheduling and explored TPU design spaces.
- Designed a **job-lifetime-aware ML algorithm** to enhance TPU chip scheduling with improved bin packing efficiency and a **50**% **TPU utilization increase** on production workloads. Implemented with 10k lines of Python code.

Student Researcher, Storage Analytics Team (Hosts: Yan Li, Mustafa Uysal, Martin Maas) May 2023 – Jan 2024, Sunnyvale, CA

- Designed and implemented a storage tiering solution for high I/O-density workloads in Google's planet-scale storage system.
- Achieved a **2.48x** total cost savings compared to existing solutions estimated to save \$12 million upon full deployment.
- The solution is being **rolled out to production**. Implemented with 20k lines of Python code.
- Paper: A Practical Cross-Layer Approach for ML-Driven Storage Placement in Warehouse-Scale Computers. C. Yang, Y. Li, M. Maas, M. Uysal, U. Hafeez, A. Merchant, R. McDougall. Under Review.

Goldman Sachs

Summer Analyst, Engineering (Host: Gang Wang)

Jun 2018 - Aug 2018, Hong Kong

- Built a workload generation tool simulating trading orders through OSI layers to test ultra-low-latency trading gateway.
- The tool identified > 5 critical bugs during the trading system development phase.

Selected Projects

Certifiably Performant, Safe, and Robust ML Systems, Project lead

Jul 2020 - Sep 2024, UT-Austin

- Certified Learning for Networked Systems.
 - C. Yang, D. Saxena, R. Dwivedula, S. Chaudhuri, A. Akella. Under Review.
 - Designed and implemented the first ML-driven congestion control systems integrating learning with formal certification.
 - ► Achieved a 78% delay reduction and improved worst-case satisfaction with formal performance and robustness properties.
- Safe Neurosymbolic Learning with Differentiable Symbolic Execution [Paper]
 - C. Yang, S. Chaudhuri. ICLR 2022.
 - Created an approach for end-to-end, worst-case-safe learning in neural networks within symbolic programs.
 - Integrated symbolic execution and stochastic gradient estimators, enabling applications in autonomous driving.
- Certifiably Robust Reinforcement Learning (RL) through Model-Based Abstract Interpretation. [Paper]
 - C. Yang, G. Anderson, S. Chaudhuri. SaTML 2024.
 - Designed an RL framework with certifiable adversarial robustness, combining model-based learning and abstract interpretation, validated through control benchmarks.

Edge Server DNN Video Processing Acceleration, Project contributor

Aug 2019 - Jun 2020, UT-Austin

- Developed a batching-aware algorithm to enhance edge DNN request scheduling and enable collaborative DNN executions.
- Achieved 400% reduction in completion time. Increased on-time ratio by 22% compared over Earliest Deadline First batching.
- Paper: Adaptive Scheduling for Edge-Assisted DNN Serving. [Paper] J. He, C. Yang, Z. He, G. Baig, L. Qiu. MASS 2023.

For a complete list of my publications, please visit my website and my Google Scholar profile.