Jucheng (Jack) Shen

jackshen-1215.github.io | jackshen-1215 | jack.shen@rice.edu | +1 (346) 448-4031

Education

Rice University B.A. in Computer Science, Mathematics, and Economics Expected May 2027

Selected Coursework: Algorithms and Data Structures, Discrete Mathematics, Computer Systems, Linear Algebra, Real Analysis, Probability and Statistics, Honors ODE

Research Experience

The University of Texas at Austin, Research Intern May 2025 – Sep 2025

- Proposed One-Shot Dynamic Thresholding (OSDT), a training-free adaptive decoding method for diffusion LLMs achieving 24–50% higher throughput without accuracy loss.
- Directed large-scale experiments and analyses, including all main results, ablations, and figure production for publication.
- Led a collaboration with Intel Labs, proposing CadLLM, a unified adaptive decoding system that jointly tunes step size, block size, vocabulary, and thresholds, achieving up to 2.3× speedup on diffusion LLMs.

Rice University, Research Intern Apr 2025 – Aug 2025

- Implemented *SuperGen*, a novel framework enabling generation of ultra-high resolution (2-4K) videos with limited GPU resources, on two baseline model architecture.
- Conducted end-to-end and ablations experiments, assisted with figure producion and paper writing for publication.

Publications and Preprints

[1] **Jucheng Shen**, Yeonju Ro. *Beyond Static Cutoffs: One-Shot Dynamic Thresholding for Diffusion Language Models*. NeurIPS 2025 *Efficient Reasoning Workshop*. Preprint.

[2] **Jucheng Shen**, Gaurav Sarkar, Yeonju Ro, Sharath Nittur Sridhar, Zhangyang Wang, Aditya Akella, Souvik Kundu. *Improving the Throughput of Diffusion-based Large Language Models via a Training-Free Confidence-Aware Calibration*. Preprint.

[3] Fanjiang Ye, Zepeng Zhao, Yi Mu, **Jucheng Shen**, Renjie Li, Kaijian Wang, Desen Sun, Saurabh Agarwal, Myungjin Lee, Triston Cao, Aditya Akella, Arvind Krishnamurthy, T. S. Eugene Ng, Zhengzhong Tu, Yuke Wang. *SuperGen: An Efficient Ultra-High-Resolution Video Generation System with Sketching and Tiling*. Preprint.