

SHEN JUNYI

Singapore

j1shen@comp.nus.edu.sg [j1shen.github.io](https://github.com/j1shen) [Scholar](#) [GitHub](#)

EDUCATION

National University of Singapore

Aug. 2024–Present

Ph.D. in Computer Science

Singapore

Advisor: Prof. Yao Lu Research: LLM systems, agentic workflows, and inference optimization

President's Graduate Fellowship GPA: 4.75/5.00

Zhejiang University

Sept. 2020–June 2024

Bachelor of Engineering in Automation

Hangzhou, China

Minor in the Intensive Training Program of Innovation and Entrepreneurship, Chu Kochen Honors College

GPA: 3.98/4.00 Rank: 3/122

PUBLICATIONS

Peer-Reviewed Publications

Yuhao Shen, Tianyu Liu, **Junyi Shen**, Jinyang Wu, Quan Kong, Huan Li, and Cong Wang. *Double: Breaking the Acceleration Limit via Double Retrieval Speculative Parallelism*. ACL 2026 (Oral, Best Paper Candidate).

Noppanat Wadlom, **Junyi Shen**, and Yao Lu. *Efficient LLM Serving for Agentic Workflows: A Data Systems Perspective*. SIGMOD 2026 (accepted).

Yuhao Shen*, **Junyi Shen***, Quan Kong, Tianyu Liu, Yao Lu, and Cong Wang. *SpecBranch: Speculative Decoding via Hybrid Drafting and Rollback-Aware Branch Parallelism*. ICLR 2026 (Poster). *Equal contribution.

Preprints

Junyi Shen, Noppanat Wadlom, Lingfeng Zhou, Dequan Wang, Xu Miao, Lei Fang, and Yao Lu. *FlowMesh: A Service Fabric for Composable LLM Workflows*. arXiv preprint, 2025.

Junyi Shen, Noppanat Wadlom, and Yao Lu. *Halo: Batch Query Processing and Optimization for Agentic Workflows*. arXiv preprint, 2025. Submitted to The VLDB Journal.

Cheng Li, Ziang Leng, Chenxi Yan, **Junyi Shen**, et al. *ChatHaruhi: Reviving Anime Character in Reality via Large Language Model*. arXiv preprint, 2023.

RESEARCH EXPERIENCE

Research Intern, Shanghai Qi Zhi Institute

Nov. 2023–May 2024

Advisors: Huanchen Zhang and Yao Lu

Developed continuous-batching techniques for Stable Diffusion serving.

Student Research Assistant, The University of Hong Kong

July–Aug. 2023

Advisor: Hengshuang Zhao

Worked on cross-modal alignment and diffusion-model training.

HONORS

President's Graduate Fellowship, National University of Singapore

2024–Present

Outstanding Graduate of Zhejiang Province

2024

First-Class Scholarship, Zhejiang University

2021, 2022, 2023

National Scholarship, Ministry of Education of China

2021

TEACHING

Teaching Assistant, National University of Singapore

CS6216: Advanced Topics in Machine Learning

AY 2025/26 Semester 1

CS4262/5462: Machine Learning Systems

AY 2025/26 Semester 2

SELECTED TALK

MLSys Singapore Meetup #17

6 Nov. 2025

Batch Query Processing and Optimization for Agentic Workflows

National Supercomputing Centre Singapore

TECHNICAL SKILLS

Programming: Python, C/C++ **Tools:** PyTorch, Linux, Git, L^AT_EX, Codex, Claude Code

LEADERSHIP & SERVICE

League-Branch Secretary, Control 2001 / IT 2022

2020–2024

Executive Chairman, Student Union, College of Control Science and Engineering, Zhejiang University 2022

Committee Member, Student Committee, Zhejiang University

2022

Volunteer, 19th Asian Games, Hangzhou

2023

