

# YanJun Shao

 super-dainiu |  YanJun Shao |  super-dainiu.github.io |  yjshao19@fudan.edu.cn

## EDUCATION

---

### Fudan University

*Bachelor of Science in Data Science (GPA: 3.67/4.0; Rank: 11/79)*

Shanghai, China

Sep. 2019 – Jun. 2023

### Nanyang Technological University

(Exchange) *Bachelor of Science in Data Science (GPA: 5.0/5.0)*

Singapore

Sep. 2022 – Dec. 2022

## PUBLICATIONS

---

- [1] Y. Liu, S. Li, J. Fang, Y. Shao, B. Yao, and Y. You, “Colossal-Auto: Unified automation of parallelization and activation checkpoint for large-scale models,” 2023. arXiv: [2302.02599](https://arxiv.org/abs/2302.02599) [cs.LG].
- [2] Y. Liu, X. Tang, Z. Cai, J. Lu, Y. Zhang, Y. Shao, *et al.*, “ML-Bench: Large language models leverage open-source libraries for machine learning tasks,” 2023. arXiv: [2311.09835](https://arxiv.org/abs/2311.09835) [cs.CL].

## RESEARCH EXPERIENCE

---

### Yale University

*Research Assistant* (Advisor: Dr. Xiangru Tang & Prof. Mark Gerstein)

New Haven, CT

Jul. 2023 – Present

Topic: INVESTIGATING THE APPLICATION OF LARGE LANGUAGE MODELS IN SINGLE-CELL TRANSCRIPTOMES

- Conducted extensive experiments evaluating foundation models for single-cell transcriptomes.
- Investigated code comprehension and task completion capabilities of popular open-source models.
- Integrated LLaMA in FavorGPT and developed LLM Agents for biological database interactions.

### Fudan University

*Research Assistant* (Advisor: Prof. Meiyue Shao)

Shanghai, China

Oct. 2021 – Jun. 2023

Thesis: AGGRESSIVE EARLY DEFLATION ON NON-SYMMETRICAL QUATERNION MATRICES (Grade: A)

- Proposed a new multi-shift QR algorithm and deflation strategy for non-symmetrical quaternion matrices.
- Employed accelerated Schur decomposition on quaternion matrices.

## INTERNSHIP EXPERIENCE

---

### Microsoft Research Asia

*Machine Learning Research Intern* (Advisor: Mr. Ning Shang)

Beijing, China

Jan. 2023 – Jun. 2023

- Maintained the pipeline for neural network compression and acceleration
- Standardized robust `concrete_trace` functionality for PyTorch programs with diverse applications.
- Improved `torch.FX` graph tracing through CPU-GPU mixed trace

### HPC-AI Technology, Inc

*Machine Learning Research Intern* (Advisor: Mr. Shenggui Li & Prof. Yang You)

Singapore

Jul. 2022 – Jan. 2023

- Modelled an auto-parallel training pipeline for LLM with tensor parallelism and rematerialization.
- Leveraged `torch.FX` to enable AoT analysis and graph transformation on PyTorch programs.
- Achieved nearly the same performance as expert-designed strategies.

### Shanghai Tuli Technology Co., Ltd.

*Software Engineer Intern*

Shanghai, China

Jul. 2021 – Aug. 2021

- Built a full pipeline for a self-adaptive pedestrian crossing system with YOLOv5 on Jetson-Xavier.

## COMMUNITY & SERVICE

---

**Open Source Contributor** in Colossal-AI  (35.0k stars), NNI  (13.3k stars)

**Program Committee** for SIGDIAL & INLG Workshop on TamingLLM, 2023

## AWARDS

---

2023 *Star of Tomorrow* Excellent Internship Award at **Microsoft Research Asia**

2021, 2022 *Undergraduate Scholarship* for Academic Performance at **Fudan University**

2021 *First Prize* in College Students Mathematical Competition by **Chinese Mathematical Society**

Last updated: April 30, 2024