

# Shiwei Gao

Beijing | gsw23@mails.tsinghua.edu.cn | storage.cs.tsinghua.edu.cn/~gsw/ | github.com/gswxp2

## Education

---

**Tsinghua University** Sept 2023 – Present

Ph.D Student in Computer Science and Technology  
Advisor: Prof. Jiwu Shu (shujw@mail.tsinghua.edu.cn)  
Research focus: Machine Learning Systems, Storage Systems

**Beihang University** Sept 2019 – June 2023

BS in Computer Science and Technology  
GPA: 3.91/4.0 (rank 1st)

## Publications

---

**Efficient Multi-LLM Serving with Workload Weaving.** USENIX ATC'25

Shiwei Gao, Qing Wang, Shaoxun Zeng, Youyou Lu, Jiwu Shu

**GPreempt: GPU Preemptive Scheduling Made General and Efficient.** USENIX ATC'25

Ruwen Fan, Tingxu Ren, Minhui Xie, **Shiwei Gao**, Jiwu Shu, Youyou Lu

**Frugal: Efficient and Economic Embedding Model Training with Commodity GPUs.** ASPLOS'25

Minhui Xie, Shaoxun Zeng, Hao Guo, **Shiwei Gao**, Youyou Lu

**Medusa: Accelerating Serverless LLM Inference with Materialization.** ASPLOS'25

Shaoxun Zeng, Minhui Xie, **Shiwei Gao**, Youmin Chen, Youyou Lu

**Fast State Restoration in LLM Serving with HCache.** EuroSys'25

**Shiwei Gao**, Youmin Chen, Jiwu Shu

**Fast Core Scheduling with Userspace Process Abstraction.** SOSP'24

Jiazhen Lin, Youmin Chen, **Shiwei Gao**, Youyou Lu

## Honors and Awards

---

1st Place, ASPLOS /EuroSys 2025 Contest on LLM Inference 2025

Comprehensive Excellent Scholarship, Tsinghua University 2024

Future Scholar Scholarship, Tsinghua University 2023

Outstanding Graduates, Beihang University 2023

Nominated for the Shen Yuan Medal, Beihang University 2022

National First Prize of Contemporary Undergraduate Mathematical Contest in Modeling 2021

## Technologies

---

**Languages:** C++, C, Python, CUDA

**Technologies:** PyTorch, vLLM, SPDK, RDMA