

# Yiqi Zhang (张懿麒)

Updated August 15, 2025

**Email:** limitviscent@gmail.com

**GitHub:** @ExtremeViscent

**Phone:** +65 98979657

**Department:** Institute of Data Science

**Research interests** High Performance Computing, Machine Learning Systems, Large Language Models

**Education**

<b>National University of Singapore</b>	Singapore
Data Science PhD	Jan 2023 - Present
PI: Dr. Yang You.	
<b>King's College London</b>	London, United Kingdom
Neuroscience MSc	Sept 2021 – Sept 2022
PI: Prof. Rosalyn Moran.	
<b>Southern University of Science and Technology</b>	Shenzhen, China
BSc in BioScience	Sept 2018 – June 2022 <sup>1</sup>
PI: Prof. Shengtao Hou.	

**Honors and scholarships**

President's Graduate Scholarship (National University of Singapore)	2023
Summa Cum Laude (Southern University of Science and Technology)	2022
3 <sup>rd</sup> Place Winner of ISC22 Student Cluster Competition (HPC-AI Advisory Council)	2022
1 <sup>st</sup> Prize of 4 <sup>th</sup> APAC HPC-AI Competition (HPC-AI Advisory Council)	2021
AI Special Prize of 4 <sup>th</sup> APAC HPC-AI Competition (HPC-AI Advisory Council)	2021
Highest Linpack Benchmark Winner in SC21 Virtual Student Cluster Competition (ACM SIGHPC/IEEE)	2021
1 <sup>st</sup> Prize of ASC20-21 Student Supercomputer Challenge (ASC committee)	2021
3 <sup>rd</sup> Prize of 3 <sup>rd</sup> APAC HPC-AI Competition (HPC-AI Advisory Council)	2020
19 <sup>th</sup> Place in SC20 Virtual Student Cluster Competition (IEEE)	2020
Certificated for Accelerated Computing – Python (NVIDIA DLI)	2020
Certificated for Accelerated Computing – C/C++ (NVIDIA DLI)	2020
Extraordinary Freshman Scholarship	2018

---

<sup>1</sup>Joint program of SUSTech and KCL

## Research experience

### **Anonymous AI Enterprise**

Jun 2025 – Present

- Redefine RL framework for best customizability and optimal performance on scalable platforms.
- Maximizing cluster-level MFU for online model serving and updating.

### **Microsoft Research Asia**

Mentor: Dr. Yuqing Yang

Jan 2025 – Jun 2025

- Proposed a high-efficiency rollout scheduling scheme for LLM RL training.
- By batch-centric scheduling, eliminated 70% scheduling bubbles in rollout.
- Reduced 50% steps in reasoning model training.

### **HPC-AI Lab**

Supervisor: Dr. Yang You

Nov 2021 – Present

- Redesigned computation graph for LLM training with offloading.
- Speed-up 10x over SOTA on LLM finetuning tasks and 3x on inference.
- Minimized memory fragmentation.

### **Institute of Psychology, Psychiatry, and Neuroscience**

Supervisors: Dr. Rosalyn Moran, Dr. Frantisek Vasa

Apr 2022 – Jan 2023

- Designed a explainable super-resolution scheme for ultra-low-field MRI.
- By Bayesian model selection, models with best confidence are chosen for inference.
- over 11% performance improvement in medical imaging tasks.

### **MEDical Image Computing and Analysis Lab (MEDICAL)**

Supervisor: Dr. Xiaoying Tang

Jan 2021 – Jan 2022

- Participated in a project aiming at medical image augmentation.
- Using statistical evidences to maximize data variety, enhancing model generalizability.
- Also, accelerated LDDMM registration algorithm by 2x on GPU.

### **Centre for Computational Science and Engineering (CCSE)**

Supervisors: Prof. Lianping Wang and Dr. Jing Fan

Jan 2019 – Sept 2023

- Benchmarking and test of new devices.
- Optimization of scientific computing software for competition, including but not limited to Relion, Presto, DLRM.

## Lianping Wang Lab

Supervisor: Prof. Lianping Wang

Jun 2019 – Sept 2020

- Ported discrete unified gas kinetic scheme to heterogeneous platform, bringing 16x speedup.

## Publications

### **SortedRL: Accelerating RL Training for LLMs through Online Length-aware Scheduling**

**Yiqi Zhang**, Huiqiang Jiang, Xufang Luo, Zhihe Yang, Chengruidong Zhang, Yifei Shen, Dongsheng Li, Yuqing Yang, Lili Qiu, Yang You

*In Submission*

### **Region-adaptive sampling for diffusion transformers**

Ziming Liu, Yifan Yang, Chengruidong Zhang, **Yiqi Zhang**, Lili Qiu, Yang You, Yuqing Yang

*arXiv preprint arXiv:2502.10389*

### **SpeedLoader: An I/O efficient scheme for heterogeneous and distributed LLM operation**

**Yiqi Zhang**, Yang You

*The 38<sup>th</sup> Annual Conference on Neural Information Processing Systems, 2024*

### **AADG: Automatic Augmentation for Domain Generalization on Retinal Images**

Junyan Lyu; **Yiqi Zhang**; Pujin Cheng; Xiaoying Tang

*IEEE Transactions on Medical Imaging, 2022.*<sup>2</sup>

### **40 Hz light flicker alters human brain EEG microstates and complexity with implications in brain diseases**

**Yiqi Zhang**; Zhenyu Zhang; Lei Luo; Huaiyu Tong; Fei Chen; Sheng Tao Hou

*Frontiers in Neuroscience, 2021*<sup>3</sup>

## Patents

### **Signal processing system**

ZHANG MINGMING; **ZHANG YIQI**; LIN WEICONG

*China National Intellectual Property Administration, 2020.*

## Talks and tutorials

### **Recipe of Champion: Award-winning HPC team 101**

Yiqi Zhang

*Supercomputing Asia, 2022.*

### **Ramp it up: From Server Room to International Challenges**

Yiqi Zhang

*Supercomputing Asia, 2021.*

---

<sup>2</sup>Full text available [here](#)

<sup>3</sup>Full text available [here](#)

