

Zili Zhang

SENIOR UNDERGRADUATE STUDENT · PEKING UNIVERSITY

Yanyuan Building 818, No.5 Yiheyuan Road, Haidian District, Beijing, Republic of China

☎ (+86) 18810396278 | ✉ zzlcs@pku.edu.cn | 🏠 www.zilizhang.site | 🏆 Gold-Sea

Education

Peking University

PH.D. IN COMPUTER SCIENCE AND ENGINEERING

- Advisor: Prof. Xin Jin

Beijing, China

Sep. 2023 - Jun. 2028 (Expected)

Peking University

B.S. IN COMPUTER SCIENCE AND ENGINEERING

- Overall GPA: 3.71/4.0 (top 9%)

Beijing, China

Sep. 2019 - Jun. 2023 (Expected)

Experience

Software Engineering Institute, PKU.

RESEARCH ASSISTANT

- Topic: Machine Learning System and Cloud Computing
- Advisor: Prof. Xin Jin

Beijing, China

Oct. 2020 - Present

ByteDance Inc.

RESEARCH INTERN OF DL COMPILER

- Optimization for TVM compiler runtime

Beijing, China

Jun. 2021 - Sep. 2021

Moqi.

RESEARCH INTERN OF VECTOR SEARCH SYSTEM

- Optimization for Faiss runtime

Beijing, China

Oct. 2021 - Feb. 2023

Alibaba.

RESEARCH INTERN OF SERVERLESS COMPUTING

- Optimization for Alibaba Serverless Platform

Beijing, China

Jun. 2023 - Present

Teaching Experience

Introduction to Computer System

TEACHING ASSISTANT

- Corrected students' homework and assisted teachers in the examination paper.
- Organized the seminar and prepared topics for discussion.

Peking University

Sep. 2021 - Jan. 2022

Introduction to Computer System (Turing Class)

TEACHING ASSISTANT

- Corrected students' homework and assisted teachers in the examination paper.
- Organized the seminar and prepared topics for discussion.

Peking University

Sep. 2022 - Jan. 2023

Publications

Fast Distributed Inference Serving for Large Language Models

IN PREPRINT

- Bingyang Wu*, Yinmin Zhong*, **Zili Zhang***, Gang Huang, Xuanzhe Liu, Xin Jin (* indicates equal contribution)

May 2023

Fast, Approximate Vector Queries on Very Large Unstructured Datasets

NSDI'23

- **Zili Zhang**, Chao Jin, Linpeng Tang, Xuanzhe Liu, Xin Jin

Boston, U.S.

Apr. 2023

Ditto: Efficient Serverless Analytics with Elastic Parallelism

SIGCOMM'23

- Chao Jin, **Zili Zhang**, Xingyu Xiang, Songyun Zou, Gang Huang, Xuanzhe Liu, Xin Jin

New York City, U.S.

Apr. 2023

Transparent GPU Sharing in Container Clouds for Deep Learning Training

NSDI'23

- Bingyang Wu, **Zili Zhang**, Zhihao Bai, Xuanzhe Liu, Xin Jin

Boston, U.S.

Apr. 2023

Rise of Distributed Deep Learning Training in the Big Model Era: From A Software Engineering Perspective

TOSEM'23

May. 2023

• Xuanzhe Liu , Diandian Gu, Zhenpeng Chen, Jinfeng Wen, **Zili Zhang**, Yun Ma, Haoyu Wang, Xin Jin

Optimizing Half Precision Winograd Convolution on ARM Many-Core Processors

Online

APSYS'22

Aug. 2022

• Dedong Xie, Zhen Jia, **Zili Zhang**, Xin Jin

Honors & Awards

UNIVERSITY AWARDS

2020	The Third Prize of Peking University Scholarship , Peking University Award	<i>Peking University</i>
2020	Award for Academic Excellents , Peking University Award	<i>Peking University</i>
2021	Award for Scientific Research , Peking University Award	<i>Peking University</i>
2022	Award for Scientific Research , Peking University Award	<i>Peking University</i>
2022	Exceptional Award for Academic Innovation (5/408) , Peking University Award	<i>Peking University</i>
2022	Lee Wai Wing Scholarship , Peking University Award	<i>Peking University</i>
2023	Top 10 Bachelor Thesis of Peking University, EECS (10/408) , Graduation Ceremony	<i>Peking University</i>
2023	Outstanding Graduation Thesis of Peking University (33/4239) , Graduation Ceremony	<i>Peking University</i>
2023	Excellent Graduate (613/4239) , Graduation Ceremony	<i>Peking University</i>
2023	Representor of Excellent Graduates (14/4239) , Graduation Ceremony	<i>Peking University</i>

Skills & Interests

Programming	AWS, Docker, Kubernetes, C++, Python, CUDA, Java, Node.js, Latex
Languages	Chinese (native), English
Interests	Running (www.itra.run/RunnerSpace/ZHANG.Zili/4938114), Climbing, Traveling, Video Games