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   *Beijing, China*
PH.D. IN COMPUTER SCIENCE AND ENGINEERING   *Sep. 2023 - Jun. 2028 (Expected)*
- Advisor: Prof. Xin Jin

Peking University   *Beijing, China*
B.S. IN COMPUTER SCIENCE AND ENGINEERING   *Sep. 2019 - Jun. 2023 (Expected)*
- Overall GPA: 3.71/4.0 (top 9%)

Experience

Software Engineering Institute, PKU.   *Beijing, China*
RESEARCH ASSISTANT   *Oct. 2020 - Present*
- Topic: Machine Learning System and Cloud Computing
- Advisor: Prof. Xin Jin

ByteDance Inc.   *Beijing, China*
RESEARCH INTERN OF DL COMPILER   *Jun. 2021 - Sep. 2021*
- Optimization for TVM compiler runtime

Moqi.   *Beijing, China*
RESEARCH INTERN OF INFORMATION RETRIEVAL SYSTEM   *Oct. 2021 - Feb. 2023*
- Optimization for Faiss runtime

Teaching Experience

Introduction to Computer System   *Peking University*
TEACHING ASSISTANT   *Sep. 2021 - Jan. 2022*
- Corrected students' homework and assisted teachers in the examination paper.
- Organized the seminar and prepared topics for discussion.

Introduction to Computer System (Honor Track)   *Peking University*
TEACHING ASSISTANT   *Sep. 2022 - Jan. 2023*
- Corrected students' homework and assisted teachers in the examination paper.
- Organized the seminar and prepared topics for discussion.

Publications

Fast Distributed Inference Serving for Large Language Models
IN PREPRINT   *May 2023*
- Bingyang Wu*, Yinmin Zhong*, **Zili Zhang***, Gang Huang, Xuanzhe Liu, Xin Jin (* indicates equal contribution)

Fast, Approximate Vector Queries on Very Large Unstructured Datasets   *Boston, U.S.*
NSDI'23   *Apr. 2023*
- **Zili Zhang**, Chao Jin, Linpeng Tang, Xuanzhe Liu, Xin Jin

Ditto: Efficient Serverless Analytics with Elastic Parallelism   *New York City, U.S.*
SIGCOMM'23   *Apr. 2023*
- Chao Jin, **Zili Zhang**, Xingyu Xiang, Songyun Zou, Gang Huang, Xuanzhe Liu, Xin Jin

Transparent GPU Sharing in Container Clouds for Deep Learning Training   *Boston, U.S.*
NSDI'23   *Apr. 2023*
- Bingyang Wu, **Zili Zhang**, Zhihao Bai, Xuanzhe Liu, Xin Jin

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

2021	**Third Prize**, Peking University Award	*Peking University*
2021	**Learning Excellence Award**, Peking University Award	*Peking University*
2022	**Outstanding Research Award**, Peking University Award	*Peking University*
2023	**Outstanding Research Award**, Peking University Award	*Peking University*
2023	**Academic Innovation Award**, Peking University Award	*Peking University*
2023	**Top 10 Bachelor Thesis of Peking University, EECS**, Graduation Ceremony	*Peking University*
2023	**Outstanding Graduation Thesis of Peking University**, Graduation Ceremony	*Peking University*
2023	**Outstanding Graduates of Peking University**, Graduation Ceremony	*Peking University*

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