

Chao Jin

SENIOR UNDERGRADUATE STUDENT · PEKING UNIVERSITY

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

✉ chaojin@pku.edu.cn | 🏠 chaojin0310.github.io | 📧 chaojin0310 | 🎓 Chao Jin

Education

Peking University

PH.D. IN COMPUTER SOFTWARE AND THEORY

- Advisor: Prof. Xin Jin

Beijing, China

Sep. 2023 - Present

Peking University

B.SC. IN COMPUTER SCIENCE AND TECHNOLOGY

- GPA: 3.82/4.0, Ranking: 5/205

Beijing, China

Sep. 2018 - Jun. 2023

Experience

ByteDance Seed-Foundation

RESEARCH INTERN

- Mentor: Haibin Lin

Beijing, China

Jul. 2023 - Present

Software Engineering Institute, PKU

RESEARCH ASSISTANT

- Advisor: Prof. Xin Jin

Beijing, China

Sep. 2021 - Jun. 2023

Alibaba Cloud

RESEARCH INTERN

- Mentor: Rui Miao

Beijing, China

Mar. 2022 - Sep. 2022

Teaching Experience

Operating Systems (Honor Track)

TEACHING ASSISTANT

Peking University

Feb. 2024 - Jun. 2024

Introduction to Computer System

TEACHING ASSISTANT

Peking University

Sep. 2021 - Jan. 2022

Publications

MegaScale-MoE: Large-Scale Communication-Efficient Training of Mixture-of-Experts Models in Production

IN PREPRINT.

May 2025

- **Chao Jin***, Ziheng Jiang*, Zhihao Bai, Zheng Zhong, Juncai Liu, Xiang Li, Ningxin Zheng, Xi Wang, Cong Xie, Wen Heng, Yiyuan Ma, Wenlei Bao, Size Zheng, Yanghua Peng, Haibin Lin, Xuanzhe Liu, Xin Jin, Xin Liu
- (* Equal Contribution)

MegaScale-Infer: Serving Mixture-of-Experts at Scale with Disaggregated Expert Parallelism

IN PREPRINT.

Apr. 2025

- Ruidong Zhu*, Ziheng Jiang*, **Chao Jin***, Peng Wu, Cesar A. Stuardo, Dongyang Wang, Xinlei Zhang, Huaping Zhou, Haoran Wei, Yang Cheng, Jianzhe Xiao, Xinyi Zhang, Lingjun Liu, Haibin Lin, Li-Wen Chang, Jianxi Ye, Xiao Yu, Xuanzhe Liu, Xin Jin, Xin Liu
- (* Equal Contribution)

RAGCache: Efficient Knowledge Caching for Retrieval-Augmented Generation

IN PREPRINT.

Apr. 2024

- **Chao Jin**, Zili Zhang, Xuanlin Jiang, Fangyue Liu, Xin Liu, Xuanzhe Liu, Xin Jin

Jolteon: Unleashing the Promise of Serverless for Serverless Workflows

Santa Clara, CA, U.S.

USENIX SYMPOSIUM ON NETWORKED SYSTEMS DESIGN AND IMPLEMENTATION (NSDI)

Apr. 2024

- Zili Zhang, **Chao Jin**, Xin Jin

Ditto: Efficient Serverless Analytics with Elastic Parallelism

New York City, NY, U.S.

ACM SPECIAL INTEREST GROUP ON DATA COMMUNICATION (SIGCOMM)

May 2023

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

Fast, Approximate Vector Queries on Very Large Unstructured Datasets

Boston, MA, U.S.

USENIX SYMPOSIUM ON NETWORKED SYSTEMS DESIGN AND IMPLEMENTATION (NSDI)

Apr. 2023

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

Melon: Breaking the Memory Wall for Resource-Efficient On-Device Machine Learning

Portland, OR, U.S.

ACM INTERNATIONAL CONFERENCE ON MOBILE SYSTEMS, APPLICATIONS, AND SERVICES (MOBISys)

Jun. 2022

- Qipeng Wang, Mengwei Xu, **Chao Jin**, Xinran Dong, Jinliang Yuan, Gang Huang, Xin Jin, Yunxin Liu, Xuanzhe Liu

Honors & Awards

UNIVERSITY AWARDS

- | | | |
|------|--|--------------------------|
| 2024 | Merit Student , Peking University Award | <i>Peking University</i> |
| 2024 | Huawei Scholarship , Peking University Award | <i>Peking University</i> |
| 2023 | Presidential Scholarship , Peking University Award | <i>Peking University</i> |
| 2021 | Learning Excellence Award , Peking University Award | <i>Peking University</i> |
| 2020 | Merit Student , Peking University Award | <i>Peking University</i> |
| 2020 | Xiaomi Scholarship , Peking University Award | <i>Peking University</i> |