

YINMIN ZHONG

Yanyuan Building, Peking University, No.5 Yiheyuan Road, Haidian District, Beijing, China

Homepage: yinminzhong.com ◇ Email: zhongyinmin@pku.edu.cn ◇ Github: PKUFlyingPig

EDUCATION

Peking University
Ph.D. in Computer Science.
Advisor: Xin Jin

Sep 2022 - Present

Peking University
B.S. in Computer Science.

Sep 2018 - June 2022
GPA: 3.79/4.0 (top 5%)

RESEARCH INTEREST

My research interest lies in the intersection of Deep Learning and Distributed Systems. I use the insight from a system view to improve different aspects (efficiency, performance, scalability) of deep learning training and serving.

PUBLICATIONS

DistServe: Disaggregating Prefill and Decoding for Goodput-optimized Large Language Model Serving

Yinmin Zhong, Shengyu Liu, Junda Chen, Jianbo Hu, Yibo Zhu, Xuanzhe Liu, Xin Jin, Hao Zhang
In preprint

Fast Distributed Inference Serving for Large Language Models

Bingyang Wu^{*}, **Yinmin Zhong**^{*}, Zili Zhang^{*}, Gang Huang, Xuanzhe Liu, Xin Jin

(*Equal Contribution)

In preprint

Scaling Large Language Model Training to More Than 10,000 GPUs

Ziheng Jiang^{*}, Haibin Lin^{*}, **Yinmin Zhong**^{*}, Qi Huang, Yangrui Chen, Zhi Zhang, Yanghua Peng, Xiang Li, Cong Xie, Shibiao Nong, Yulu Jia, Sun He, Hongmin Chen, Zhihao Bai, Qi Hou, Shipeng Yan, Ding Zhou, Yiyao Sheng, Zhuo Jiang, Haohan Xu, Haoran Wei, Zhang Zhang, Pengfei Nie, Leqi Zou, Sida Zhao, Liang Xiang, Zherui Liu, Zhe Li, Xiaoying Jia, Jianxi Ye, Xin Jin, Xin Liu

(*Equal Contribution)

USENIX Symposium on Networked Systems Design and Implementation (NSDI 2024)

DistMind: Efficient Resource Disaggregation for Deep Learning Workloads

Xin Jin, Zhihao Bai, Zhen Zhang, Yibo Zhu, **Yinmin Zhong**, Xuanzhe Liu

IEEE/ACM Transactions on Networking (TON 2024)

AlpaServe: Statistical Multiplexing with Model Parallelism for Deep Learning Serving

Zhuohan Li^{*}, Lianmin Zheng^{*}, **Yinmin Zhong**^{*}, Vincent Liu, Ying Sheng, Xin Jin, Yanping Huang, Zhifeng Chen, Hao Zhang, Joseph E. Gonzalez, Ion Stoica (*Equal Contribution)

USENIX Symposium on Operating Systems Design and Implementation (OSDI 2023)

ElasticFlow: An Elastic Serverless Training Platform for Distributed Deep Learning

Diandian Gu, Yihao Zhao, **Yinmin Zhong**, Yifan Xiong, Zhenhua Han, Peng Cheng, Fan Yang, Gang Huang, Xin Jin, Xuanzhe Liu

ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS 2023)

EXPERIENCE

Peking University <i>Research Assistant</i>	Sep 2022 - Present
---	---------------------------

· Mentor: Xin Jin

ByteDance AML Team <i>Research Intern</i>	Aug 2023 - Present
---	---------------------------

· Mentor: Xin Liu

Alibaba DAMO Academy <i>Research Intern</i>	Sep 2021 - Sep 2022
---	----------------------------

· Mentor: Pengyu Zhang

AI Innovation Center, Peking University <i>Software Engineer Intern</i>	Sep 2020 - Mar 2021
---	----------------------------

· Mentor: Ming Lei

PROJECTS

Alpa

Alpa automates model-parallel training of large deep learning (DL) models by generating execution plans that unify data, operator, and pipeline parallelism. Alpa makes it simple to train and serve large models like GPT-3.

csdiy.wiki

csdiy.wiki collects various learning materials and online courses in computer science (CS) from top universities in the world. It covers almost all the areas in CS and provides a comprehensive guidance for anyone who would like to self-learn CS. This project has repeatedly ranked the top popular repository in Github Trending.

TEACHING

Head TA, Operating Systems (Honor Track)	2023 Spring
---	--------------------

TA, Introduction to Computing	2022 Fall
--------------------------------------	------------------

Head TA, Operating Systems (Honor Track)	2022 Spring
---	--------------------

TA, Introduction to Computer System	2020 Fall
--	------------------

AWARDS & HONORS

Outstanding Graduate of Peking University	2022
--	-------------

Third Prize Scholarship of Peking University	2020
---	-------------

Merit Student of Peking University	2019
---	-------------

Tian Chuang Scholarship	2019
--------------------------------	-------------

Zhongying Tang Scholarship	2019
-----------------------------------	-------------

Merit Student of Peking University	2018
---	-------------