

# KAIQI ZHAO

## RESEARCH INTERESTS

Efficient AI, Machine Learning/Deep Learning Model Compression (Knowledge Distillation, Pruning, Quantization), Distributed/Cloud/Edge Computing

## EDUCATION

**Ph.D. in Computer Science**, May 2024

**Arizona State University**, Tempe, AZ

Dissertation: Improved and Automated Machine Learning Model Compression

Advisor: Ming Zhao

**M.S. in Software Engineering**, **National Graduate Fellowship winner**, March 2018

**Beihang University**, Beijing, China

**B.S. in Software Engineering**, July 2015

**Inner Mongolia University**, Inner Mongolia, China

## PROFESSIONAL EXPERIENCE

**Oakland University**, Computer Science and Engineering Department, Rochester, MI

**Tenure-Track Assistant Professor**, August 2024 - Present

**Arizona State University**, School of Computing and Augmented Intelligence (SCAI), Tempe, AZ

**Research Assistant / Instructor / Teaching Assistant**, August 2018 - May 2024

**Amazon Web Services, Inc.**

- **Applied Scientist Intern**, Santa Clara, CA, May 2022 - August 2022
- **Applied Scientist Intern**, Santa Clara, CA, May 2021 - August 2021
- **Applied Scientist Intern**, East Palo Alto, CA, May 2020 - August 2020

**Venture Commune Investment Development Co. Ltd.**, Beijing, China

**Data Scientist Intern**, June 2016 - December 2017

**Huawei Beijing Research Institute**, Beijing, China

**Research Intern**, October 2014 - September 2015

## SELECTED PUBLICATIONS

[P1] **AISTATS'24: K. Zhao**, M. Zhao, "Self-Supervised Quantization-Aware Knowledge Distillation," *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics (AISTATS)*, May 2024. **(Acceptance Rate: 27.6%)**

[P2] **SEC'23: K. Zhao**, M. Zhao, "Poster: Self-Supervised Quantization-Aware Knowledge Distillation," *Proceedings of ACM / IEEE Symposium on Edge Computing (SEC)*, December 2023. **(Best Poster Award)**

[P3] **ICASSP'23: K. Zhao**, Y. Chen, M. Zhao, "A Contrastive Knowledge Transfer Framework for Model Compression and Transfer Learning," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, June 2023. **(Oral)**

[P4] **AISTATS'23**: K. Zhao, A. Jain, M. Zhao, "Automatic Attention Pruning: Improving and Automating Model Pruning using Attentions," *Proceedings of the 26th International Conference on Artificial Intelligence and Statistics (AISTATS)*, April 2023. **(Acceptance Rate: 29%)**

[P5] **Interspeech'22**: K. Zhao, H. D. Nguyen, A. Jain, N. Susanj, A. Mouchtaris, L. Gupta, M. Zhao, "Knowledge Distillation via Module Replacing for Automatic Speech Recognition with Recurrent Neural Network Transducer," *Proceedings of the 23rd Interspeech Conference*, September 2022. **(Oral)**

[P6] **HotEdge'19**: Y. Chen, K. Zhao, B. Li, and M. Zhao, "Exploring the Use of Synthetic Gradients for Distributed Deep Learning across Cloud and Edge Resources," *Proceedings of the 2nd USENIX Workshop on Hot Topics in Edge Computing (HotEdge)*, July 2019. **(Acceptance Rate: 29%)**

[P7] **OpML'19**: S. Biookaghazadeh, Y. Chen, K. Zhao, and M. Zhao, "KnowledgeNet: Disaggregated and Distributed Training and Serving of Deep Neural Networks," *Proceedings of the USENIX Conference on Operational Machine Learning (OpML)*, May 2019.

## **SELECTED AWARDS**

- **Faculty Research Fellowship**, Oakland University, 2025
- **Nomination of Dean's Dissertation Award**, Arizona State University, 2024
- **Graduate College Completion Fellowship**, Arizona State University, 2024
  - **The most prestigious award** given to graduate students in the School of Computing and Augmented Intelligence (**two recipients per year**)
- **ACM/IEEE Symposium on Edge Computing (SEC) NSF Best Poster Award**, 2023
- **ACM/IEEE Symposium on Edge Computing (SEC) NSF Student Travel Grant**, twice, 2023 & 2021
- **School of Computing and Augmented Intelligence (SCAI) Doctoral Fellowship**, Arizona State University, twice, 2023 & 2022
- **National Graduate Fellowship**, Beihang University, China, 2017
  - **The most prestigious award** given to graduate students (**five recipients per year**)
- **Excellence Award in the Global Competition of the Open Innovation Business Contest 5.0 of NTT DATA**, Japan, 2017
  - Represented **the sole champion team** from China, garnering media coverage from **prominent mainstream outlets**, e.g., PEOPLE.CN<sup>1</sup> and NTT DATA<sup>2</sup>.

## **SELECTED SERVICES**

- **Conference Chair**: **ACM/IEEE Symposium on Edge Computing (SEC) Ph.D. Forum Chair**, 2021
- **Program Committee Member**: **AAAI Conference on Artificial Intelligence**, 2025 & 2024 & 2023
- **Invited Reviewer**: **NeurIPS** 2022-2025, **ICML** 2022-2025, **ICLR** 2022-2025, **AAAI** 2023-2025, **AISTATS** 2024&2025, **ICASSP** 2024&2025, **InterSpeech** 2023-2025, **IEEE Internet of Things Journal** (Impact Factor: 11.1), **IEEE Transactions on Neural Networks and Learning Systems** (Impact Factor: 10.2), **IEEE Transactions on Circuits and Systems for Video Technology** (Impact Factor: 8.3)

<sup>1</sup>PEOPLE.CN: <http://japan.people.com.cn/BIG5/n1/2017/0316/c35421-29150003.html>

<sup>2</sup>NTT DATA: <https://oi.nttdata.com/en/contest/5th/result/#gf>, and <https://oi.nttdata.com/en/contest/5th/venue/c07/>