

# Hongyi Wang

Senior Project Scientist

Machine Learning Department  
Carnegie Mellon University  
✉ [hongyiwa@andrew.cmu.edu](mailto:hongyiwa@andrew.cmu.edu)  
📄 [hwang595.github.io](https://github.com/hwang595)

## Objective

I am applying for the Assistant Professor position in Computer Science.

## Positions

- 10/2023-  
Current **Senior Project Scientist** Machine Learning Department at CMU.  
Hosted by Eric P. Xing
- 02/2023-  
09/2023 **Senior Researcher** Machine Learning Department at CMU.  
Hosted by Eric P. Xing
- 09/2021-  
01/2023 **Postdoctoral Fellow** Machine Learning Department at CMU.  
Hosted by Eric P. Xing
- Summer 2020 **Research Intern** Microsoft, DeepSpeed Team.  
Hosted by Minjia Zhang and Yuxiong He
- Summer 2019 **Research Intern** IBM Research.  
Hosted by Mikhail Yurochkin and Yasaman Khazaeni

## Research interests

I am interested in co-designing systems and algorithms for efficient and trustworthy large-scale machine learning. I am particularly interested in applying my research in the development and deployment of foundation models, such as GPT and LLaMA.

## Education

- 2016–2021 **Ph.D. in Computer Science** University of Wisconsin–Madison.  
Advisor: Dimitris Papailiopoulos
- 2016–2019 **M.S. in Computer Science** University of Wisconsin–Madison.
- 2012–2016 **B.S. in Electrical Engineering** Hangzhou Dianzi University.

## Publications

\* stands for the joint first author. Here is my [Google Scholar Profile](#).

- [1] Junbo Li, Ang Li, Chong Tian, Qirong Ho, Eric Xing, and **Hongyi Wang**. Fednar: Federated optimization with normalized annealing regularization. *NeurIPS*, 2023.
- [2] **Hongyi Wang**, Saurabh Agarwal, Pongsakorn U-chupala, Yoshiki Tanaka, Eric Xing, and Dimitris Papailiopoulos. Cuttlefish: Low-rank model training without all the tuning. *MLSys*, 2023.
- [3] Dacheng Li\*, Rulin Shao\*, **Hongyi Wang\***, Han Guo, Eric Xing, and Hao Zhang. Mpcformer: fast, performant and private transformer inference with mpc. *ICLR (Spotlight)*, 2023.
- [4] Han Guo, Philip Greengard, **Hongyi Wang**, Andrew Gelman, Eric Xing, and Yoon Kim.

- Federated learning as variational inference: A scalable expectation propagation approach. *ICLR*, 2023.
- [5] Kai Zhang, Yu Wang, **Hongyi Wang**, Lifu Huang, Carl Yang, and Lichao Sun. Efficient federated learning on knowledge graphs via privacy-preserving relation embedding aggregation. *Findings of EMNLP*, 2022.
- [6] Kartik Sreenivasan, Jy-yong Sohn, Liu Yang, Matthew Grinde, Alliot Nagle, **Hongyi Wang**, Eric Xing, Kangwook Lee, and Dimitris Papailiopoulos. Rare gems: Finding lottery tickets at initialization. *NeurIPS*, 2022.
- [7] Dacheng Li, **Hongyi Wang**, Eric Xing, and Hao Zhang. Amp: Automatically finding model parallel strategies with heterogeneity awareness. *NeurIPS*, 2022.
- [8] Saurabh Agarwal, **Hongyi Wang**, Shivaram Venkataraman, and Dimitris Papailiopoulos. On the utility of gradient compression in distributed training systems. *MLSys*, 2022.
- [9] **Hongyi Wang**, Saurabh Agarwal, and Dimitris Papailiopoulos. Pufferfish: Communication-efficient models at no extra cost. *MLSys*, 2021.
- [10] Saurabh Agarwal, **Hongyi Wang**, Kangwook Lee, Shivaram Venkataraman, and Dimitris Papailiopoulos. Accordion: Adaptive gradient communication via critical learning regime identification. *MLSys*, 2021.
- [11] Chaoyang He, Songze Li, Jinhyun So, Mi Zhang, **Hongyi Wang**, Xiaoyang Wang, Pra-neeth Vepakomma, Abhishek Singh, Hang Qiu, Li Shen, Peilin Zhao, Yan Kang, Yang Liu, Ramesh Raskar, Qiang Yang, Murali Annavaram, and Salman Avestimehr. Fedml: A research library and benchmark for federated machine learning. *NeurIPS SpicyFL workshop (Best Paper Award, 467 citations)*, 2020.
- [12] **Hongyi Wang**, Kartik Sreenivasan, Shashank Rajput, Harit Vishwakarma, Saurabh Agarwal, Jy-yong Sohn, Kangwook Lee, and Dimitris Papailiopoulos. Attack of the tails: Yes, you really can backdoor federated learning. *NeurIPS (432 citations)*, 2020.
- [13] **Hongyi Wang**, Mikhail Yurochkin, YueKai Sun, Dimitris Papailiopoulos, and Yasaman Khazaeni. Federated learning with matched averaging. *ICLR (Oral, 913 citations)*, 2020.
- [14] Shashank Rajput\*, **Hongyi Wang**\*, Zachary Charles, and Dimitris Papailiopoulos. Detox: A redundancy-based framework for faster and more robust gradient aggregation. *NeurIPS (113 citations)*, 2019.
- [15] Lingjiao Chen, **Hongyi Wang**, Leshang Chen, Paraschos Koutris, and Arun Kumar. Demonstration of nimbus: Model-based pricing for machine learning in a data marketplace. In *SIGMOD 2019*, pages 1885–1888. ACM, 2019.
- [16] **Hongyi Wang**, Scott Sievert\*, Shengchao Liu, Zachary Charles, Dimitris Papailiopoulos, and Stephen Wright. Atomo: Communication-efficient learning via atomic sparsification. In *NeurIPS (343 citations)*, 2018.
- [17] Lingjiao Chen, **Hongyi Wang**, Jinman Zhao, Dimitris Papailiopoulos, and Paraschos Koutris. The effect of network width on the performance of large-batch training. In *NeurIPS*, 2018.
- [18] Lingjiao Chen, **Hongyi Wang**, Zachary Charles, and Dimitris Papailiopoulos. Draco: Byzantine-resilient distributed training via redundant gradients. In *ICML (270 citations)*, 2018.
- [19] Lingjiao Chen, **Hongyi Wang**, and Dimitris Papailiopoulos. Draco: Robust distributed training against adversaries. In *SysML*, 2018.
- [20] Guru Subramani, Daniel Rakita, **Hongyi Wang**, Jordan Black, Michael Zinn, and Michael Gleicher. Recognizing actions during tactile manipulations through force sensing. In *IROS*, pages 4386–4393. IEEE, 2017.

## Preprints

- [21] Zhengzhong Liu, Aurick Qiao, Willie Neiswanger, **Hongyi Wang**, Bowen Tan, Tianhua Tao, Junbo Li, Yuqi Wang, Suqi Sun, Omkar Pangarkar, et al. Llm360: Towards fully transparent open-source llms. *arXiv preprint arXiv:2312.06550*, 2023.
- [22] Bowen Tan, Yun Zhu, Lijuan Liu, **Hongyi Wang**, Yonghao Zhuang, Jindong Chen, Eric Xing, and Zhiting Hu. Redco: A lightweight tool to automate distributed training of llms on any gpu/tpus. *arXiv preprint arXiv:2310.16355*, 2023.
- [23] Minghao Yan, **Hongyi Wang**, and Shivaram Venkataraman. Polythrottle: Energy-efficient neural network inference on edge devices. *arXiv preprint arXiv:2310.19991*, 2023.
- [24] Zhiqiang Shen, Tianhua Tao, Liqun Ma, Willie Neiswanger, Zhengzhong Liu, **Hongyi Wang**, Bowen Tan, Joel Hestness, Natalia Vassilieva, Daria Soboleva, and Eric Xing. Slimpajama-dc: Understanding data combinations for llm training. *arXiv:2309.10818*, 2023.
- [25] **Hongyi Wang**, Felipe Maia Polo, Yuekai Sun, Souvik Kundu, Eric Xing, and Mikhail Yurochkin. Fusing models with complementary expertise. *arXiv:2310.01542*, 2023.
- [26] Song Bian, Dacheng Li, **Hongyi Wang**, Eric Xing, and Shivaram Venkataraman. Does compressing activations help model parallel training? *arXiv:2301.02654*, 2023.
- [27] with Jianyu Wang, Zachary Charles, Zheng Xu, Gauri Joshi, H Brendan McMahan, et al. A field guide to federated optimization. *arXiv:2107.06917 (285 citations)*, 2021.
- [28] Lingjiao Chen\*, Leshang Chen\*, **Hongyi Wang\***, Susan Davidson, and Edgar Dobriban. Solon: Communication-efficient byzantine-resilient distributed training via redundant gradients. *arXiv:2110.01595*, 2021.
- [29] **Hongyi Wang**, Zachary Charles, and Dimitris Papailiopoulos. Erasurehead: Distributed gradient descent without delays using approximate gradient coding. *arXiv:1901.09671*, 2019.

---

## Grants

NSF IIS2311990 (Senior Personnel, PI: Eric P. Xing) “III: Small: Multiple Device Collaborative Learning in Real Heterogeneous and Dynamic Environments”, 09/01/2023-08/31/2026.

Semiconductor Research Corp. Artificial Intelligence Hardware Program (Project Co-lead, PI: Eric P. Xing) “Co-designing Distributed ML Systems and Algorithms for Foundation Models for AI-for-Science”, 01/01/2024-12/31/2026.

---

## Honors & Awards

- 2024 **The Rising Stars Award at the Conference on Parsimony and Learning (CPAL).**
- 2018-2022 **Student Travel Award ICML 2018, NeurIPS 2018, 2019, MLSys 2022.**
- 2020 **The Baidu Best Paper Award SpicyFL workshop at NeurIPS 2020.**
- 2020 **Top Reviewer Award ICML 2020.**
- 2019 **Top Reviewer Award NeurIPS 2019.**
- 2015 **National Scholarship of China (Top 2%).**

---

## Open-Source Projects

**LLM360** An initiative to fully open-source LLMs, which advocates for all training code and data, model checkpoints, and intermediate results to be made available. The goal of LLM360 is to support open and collaborative AI research by making the end-to-end LLM pre-training process transparent and reproducible by everyone.

---

## Mentoring

Zheyu Shen (Ph.D. student at University of Maryland ECE), co-advised with Prof. Ang Li.

Han Guo (Ph.D. student at CMU LTI).

Kartik Sreenivasan (Ph.D. student at UWisconsin CS).

Jinyu Hou (M.Sc. student at CMU MLD).

Junbo Li (Research intern at MBZUAI).

Alliot Nagle (M.Sc. student at UW-Madison), now a Ph.D. student of ECE at UT Austin.

Dacheng Li (M.Sc. student at CMU), now a Ph.D. student at EECS at UC Berkeley.

Rulin Shao (M.Sc. student at CMU), now a Ph.D. student at UWashington CS.

Youngmin Kim (M.Sc. student at CMU), now a Machine Learning Scientist at Flexport.

Aditya Bhagwat (M.Sc. student at CMU), now an AI Framework Engineer at Modular.

Heather Jia (undergraduate student at UW-Madison).

---

## Invited Talks

- Oct. 2023 **INFORMS Annual Meeting 2023**, “*On the System Efficiency and Backdoor Robustness of Federated Learning*”.
- Aug. 2023 **MOPTA 2023**, “*On the System Efficiency and Backdoor Robustness of Federated Learning*”.
- Jun. 2023 **CVPR 2023 Tutorial**, “*ML Systems for Large Models and Federated Learning*”.
- Nov. 2022 **Qualcomm AI Research DistributedML Seminar**, “*On the Utility of Gradient Compression in Distributed Training Systems*”.
- Oct. 2022 **1st CASL Workshop**, “*On the Utility of Gradient Compression in Distributed Training Systems*”.
- Jul. 2022 **MBZUAI Center of Integrative Artificial Intelligence Colloquium**, “*On the Utility of Gradient Compression in Distributed Training Systems*”.
- Jan. 2022 **MLOPT Idea Seminar**, “*On the Utility of Gradient Compression in Distributed Training Systems*”.
- Jun. 2021 **Federated Learning One World Seminar**, “*On the efficiency and robustness of FL*”.
- Jan. 2021 **vITAL Research Lab Seminar**, “*Communication-efficient and robust distributed ML*”.

---

## Professional Service

**Program committee:** DAC 2024, EuroSys 2024, SOSP 2023 (light PC), MLSys 2023-24, MLSys 2022 (Artifact Evaluation Committee), SIGKDD 2022-23, AAAI 2021-22.

**Reviewer (journal):** JMLR, TMLR, IEEE TNNLS, IEEE IoT-J, IEEE Transactions on Pattern Analysis and Machine Intelligence.

**Reviewer (conference):** ICML 2019-23, NeurIPS 2019-23, ICLR 2021-24, CVPR 2021-24, ICCV 2021-23.

**Conference session chair:** Tutorial session ICML 2022, Federated learning session MLSys 2023.

**Workshop Organizer:** Federated Learning Systems (FLSys) Workshop @ MLSys 2023.

---

## Teaching Experience

Spring 2023 **Guest Lecturer MBZUAI ML710:** Parallel and Distributed ML Systems.

Fall 2022 **Guest Lecturer MBZUAI ML710:** Parallel and Distributed ML Systems.

Spring 2022 **Guest Lecturer UW-Madison ECE826:** Theoretical Foundations of Large-scale ML.

---

## References

### **Eric P. Xing (CMU supervisor)**

Professor  
School of Computer Science  
Carnegie Mellon University  
President & Professor  
Mohamed bin Zayed University of Artificial Intelligence  
epxing@cs.cmu.edu  
**Please request letter from:** aprotos@cs.cmu.edu

### **Dimitris Papailiopoulos (Ph.D. advisor)**

Jay & Cynthia Ihlenfeld Associate Professor  
Department of Electrical and Computer Engineering  
University of Wisconsin-Madison  
dimitris@ece.wisc.edu

### **Shivaram Venkataraman**

Assistant Professor  
Department of Computer Science  
University of Wisconsin-Madison  
shivaram@cs.wisc.edu

### **Kangwook Lee**

Assistant Professor  
Department of Electrical and Computer Engineering  
University of Wisconsin-Madison  
kangwook.lee@wisc.edu