

Yifei He | Curriculum Vitae

✉ yifeihe3@illinois.edu • 🌐 yifei-he.github.io

My research interest is centered around trustworthy machine learning. I currently work on: i) Improving the multi-task capabilities of foundation models (e.g, multilingual LLMs, model merging). ii) Enhancing the efficiency of LLMs (e.g., semi-supervised learning, MoE models). Previously, I have worked on multi-objective optimization, domain adaptation/generalization and multimodal learning.

Education

University of Illinois Urbana-Champaign (UIUC)

Ph.D. in Computer Science

M.S. in Computer Science

Advisor: Prof. Han Zhao

Urbana, IL, USA

May 2023 - Present

Aug 2021 - May 2023

University of Michigan (UM)

B.S.E. in Data Science, minor in Mathematics

Summa Cum Laude

Ann Arbor, MI, USA

Aug 2019 - Apr 2021

Shanghai Jiao Tong University (SJTU)

B.S.E. in Electrical and Computer Engineering

Shanghai, China

Sept 2017 - Aug 2021

Industry Experience

Microsoft

Research Intern, GenAI

Manager: Dr. Yang Liu

Redmond, WA, USA

Aug 2024 - Feb 2025

- Worked on distillation of Mixture-of-Experts (MoE) models.

Microsoft

Applied Scientist Intern, Turing

Manager: Dr. Alon Benhaim

Redmond, WA, USA

May 2024 - Aug 2024

- Worked on scaling law for multilingual language models.

Amazon

Applied Scientist Intern, Search Science and AI

Manager: Dr. Alejandro Mottini

Seattle, WA, USA

May 2023 - Aug 2023

- Improved large-scale multi-task tuning of foundation models.
- Developed a vision-language retrieval foundation model with instruction tuning.

Publications (* denotes equal contribution)

[1] Efficiently Editing Mixture-of-Experts Models with Compressed Experts.

Yifei He, Yang Liu, Chen Liang, Hany Awadalla.

Under review.

[2] Scaling Laws for Multilingual Language Models.

Yifei He, Alon Benhaim, Barun Patra, Praneetha Vaddamanu, Sanchit Ahuja, Parul Chopra, Vishrav Chaudhary, Han Zhao, Xia Song.

Under review.

[3] Efficient Model Editing with Task Vector Bases: A Theoretical Framework and Scalable Approach.

Siqi Zeng, Yifei He, Weiqiu You, Yifan Hao, Yao-Hung Hubert Tsai, Makoto Yamada, Han Zhao.

Under review.

- [4] **Towards Understanding the Fragility of Multilingual LLMs against Fine-Tuning Attacks.**
Samuele Poppi, Zheng-Xin Yong, **Yifei He**, Bobbie Chern, Han Zhao, Aobo Yang, Jianfeng Chi.
Findings of the Association for Computational Linguistics: NAACL 2025. (NAACL 2025 Findings)
- [5] **Localize-and-Stitch: Efficient Model Merging via Sparse Task Arithmetic.**
Yifei He, Yuzheng Hu, Yong Lin, Tong Zhang, Han Zhao.
Transactions of Machine Learning Research. (TMLR)
- [6] **Semi-Supervised Reward Modeling via Iterative Self-Training.**
Yifei He*, Haoxiang Wang*, Ziyang Jiang, Alexandros Papangelis, Han Zhao.
Findings of the Association for Computational Linguistics: EMNLP 2024. (EMNLP 2024 Findings)
- [7] **Robust Multi-Task Learning with Excess Risks.**
Yifei He, Shiji Zhou, Guojun Zhang, Hyokun Yun, Yi Xu, Belinda Zeng, Trishul Chilimbi, Han Zhao.
International Conference on Machine Learning. (ICML 2024)
- [8] **Gradual Domain Adaptation: Theory and Algorithms.**
Yifei He*, Haoxiang Wang*, Bo Li, Han Zhao.
Journal of Machine Learning Research. (JMLR)
- [9] **Efficient Modality Selection in Multimodal Learning.**
Yifei He*, Runxiang Cheng*, Gargi Balasubramaniam*, Yao-Hung Hubert Tsai, Han Zhao.
Journal of Machine Learning Research. (JMLR)
(Extended version of publication [10].)
- [10] **Greedy Modality Selection via Approximate Submodular Maximization.**
Runxiang Cheng*, Gargi Balasubramaniam*, **Yifei He***, Yao-Hung Hubert Tsai, Han Zhao.
Conference on Uncertainty in Artificial Intelligence. (UAI 2022)
- [11] **Conformer-RL: A Deep Reinforcement Learning Library for Conformer Generation.**
Runxuan Jiang, Tarun Gogineni, Joshua Kammeraad, **Yifei He**, Ambuj Tewari, Paul Zimmerman.
Journal of Computational Chemistry. (JCC)
- [12] **A Hierarchical Approach to Multi-Event Survival Analysis.**
Donna Tjandra, **Yifei He**, Jenna Wiens.
AAAI Conference on Artificial Intelligence. (AAAI 2021)

Professional Service

Reviewer: UAI, NeurIPS, ICLR, AISTATS, ICML, TMLR, ACL

Teaching Experience

Teaching assistant at UIUC

- CS 357 Numerical Methods I 2022 Fall, 2022 Spring
- CS 441 Applied Machine Learning 2021 Fall

Teaching assistant at UM

- EECS 445 Intro to Machine Learning 2020 Fall

Skills

Programming: Python, Java, C++, Matlab, R, \LaTeX , Mathematica

Framework: PyTorch, DeepSpeed, TensorFlow, Keras, Gym