Xiaotian (Max) Han

Ph.D. Student in Computer Science
Texas A&M University
435 Nagle St, College Station, TX

+1 (979) 820-0707
han@tamu.edu
https://ahxt.github.io

Research Interests & Highlights [Google Scholar]

My research focuses on artificial intelligence (AI), machine learning (ML), and data science (DS), including the following topics:

- **Large Language Models**: Long LLMs, Efficient LLM pre-training, Lite LLM Model, LLM for Healthcare.
- **Efficient Machine Learning**: Develop ML models with less data and constrained computational resources.
- **Trustworthy Machine Learning**: Fair Machine Learning, Fairness Evaluation.

The highlights of my research are as follows:

- **ICML2022 Outstanding Paper Award (first author)**.
- ~20 (11 first-author) peer-reviewed research papers published in ICML, ICLR, NeurIPS, WWW, KDD, AAAI, TMLR, etc.
- **1122** citations; h-index: 13; i10-index: 14.
- **Train an LLM (LiteLLaMa) from scratch** using 1 trillion tokens, 40000+ downloads to date.

Education

Texas A&M University, College Station, TX, USA — Sep 2019 – May 2024 (Expected)
Ph.D. in Computer Science, Advisor: Prof. Xia (Ben) Hu

Beijing University of Posts and Telecommunications, Beijing, CHINA, *M.S. in Computer Science* — Sep 2016 – June 2019
Shandong University, Jinan, Shandong, CHINA, *B.S. in Information and Communication Engineering* — Sep 2011 – June 2015

Awards & Honors

ICML2022 Outstanding Paper Award 2022
Excellent Ph.D. Student Award (One Per Year), Department of CSE, Texas A&M University 2023
NeurIPS 2023 Scholar Award 2023
Grad School Research and Presentation Travel Award, Texas A&M University 2023
Best Paper Awards, ADMA2018 2018
Travel Grant, Department of CSE, Texas A&M University 2022,2023
Travel Award, ICML2022 2022
Outstanding Reviewer Award, ICML2022 2022
Best Reviewer Award, CCF Transactions on Pervasive Computing and Interaction 2019
National Graduate Scholarship, Beijing University of Posts and Telecommunications 2018

Research Proposal Writing Experience (Selected)

NSF FAI: Towards Fairness in Deep Neural Networks with Learning Interpretation (funded), PI: Xia Hu — Sep 2019
My Role: Formulating the research topic and writing proposal task, *Fairness-Aware GNN via Subgraph Analysis*.

NSF III CORE: Medium: Towards Effective Interpretation of Deep Learning (funded), PI: Xia Hu — Sep 2020
My Role: Discussing research objectives and writing multiple proposal tasks in *interpretability of deep learning from prediction*.

NIH: R01: AI-enabled Learning Health Community for AD/ADRD Care and Equity (in progress), M-PI: Xia Hu — Oct 2023
My Role: Discussing research objectives, preparing preliminary experiments, and writing proposal task, *Develop a sustainable analytical pipeline to support an intelligent LHS-LHC for AD/ADRD care*.

Teaching & Mentoring

CS courses I would like to teach **Computer Architecture**, **Embedded Systems**, **Data Structures**, **Programming**
AI/ML courses I would like to teach: **Artificial Intelligence**, **Deep Learning**, **Data Science**, **Ethics in Machine Learning**

Teaching Assistant and Lecturer, Texas A&M University — Aug 2023 – present
CSCE636: Deep Learning, Instructor: Anxiao Jiang

Online Course Creator and Mentor — Sept 2021 - Nov 2021
Online Course: Image Classification with AutoKeras, Manning Publications Co.

Guest lecturer, Rice University — Fall 2023
COMP 640: Graduate Research Seminar in Machine Learning, Instructor: Xia Hu

Mentoring: *Yicheng Wang*; *Hongye Jin*; *Chia-Yuan Chang*; *Zhimeng Jiang*; *Ting-Hsiang Wang*; *Devi Sandeep Endluri*

Publications [Google Scholar]

~20 (11 first-author) peer-reviewed research papers;
1122 citations; h-index: 13; i10-index: 14
* indicates equal contribution;

[ICML2022] ***$\mathcal{G}$-Mixup: Graph Augmentation for Graph Classification***
Xiaotian Han, Zhimeng Jiang, Ninghao Liu, Xia Hu
International Conference on Machine Learning (ICML), 2022
ICML2022 Outstanding Paper Award

[ICLR2024] ***FFB: A Fair Fairness Benchmark for In-Processing Group Fairness Methods***
Xiaotian Han, Jianfeng Chi, Yu Chen, Qifan Wang, Han Zhao, Na Zou, Xia Hu
International Conference on Learning Representations (ICLR), 2024

[TKDD2024] ***Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond***
Xiaotian Han*, Jingfeng Yang*, Hongye Jin*, Ruixiang Tang*, Qizhang Feng*, Haoming Jiang, Bing Yin, Xia Hu
Transactions on Knowledge Discovery from Data (TKDD), 2024

[NeurIPS2023] ***Chasing Fairness under Distribution Shift: a Model Weight Perturbation Approach***
Xiaotian Han*, Zhimeng Jiang*, Hongye Jin, Guanchu Wang, Rui Chen, Na Zou, Xia Hu.
Neural Information Processing Systems (NeurIPS), 2023

[ICLR2023] MLPInit***: Embarrassingly Simple GNN Training Acceleration with MLP Initialization***
Xiaotian Han, Tong Zhao, Yozen Liu, Xia Hu, Neil Shah
International Conference on Learning Representations (ICLR), 2023

[TMLR2023] ***Retiring ΔDP: New Distribution-Level Metrics for Demographic Parity***
Xiaotian Han*, Zhimeng Jiang*, Hongye Jin*, Zirui Liu, Na Zou, Qifan Wang, Xia Hu
Transactions on Machine Learning Research (TMLR), 2023

[WWW2022] ***Graph Representation Learning via Unsupervised Rate Reduction Maximization***
Xiaotian Han, Zhimeng Jiang, Ninghao Liu, Xia Hu
The Web Conference (WWW), 2022

[AMIA2023] ***Does Synthetic Data Generation of LLMs Help Clinical Text Mining?***
Xiaotian Han*, Ruixiang Tang*, Xiaoqian Jiang, Xia Hu
AMIA 2023 Annual Symposium, 2023

[IJCAI2018] ***Aspect-Level Deep Collaborative Filtering via Heterogeneous Information Networks***
Xiaotian Han, Chuan Shi, Senzhang Wang, Philip, S Yu, Li Song
International Joint Conference on Artificial Intelligence (IJCAI), 2018

[APWeb2018] ***Representation Learning with Depth and Breadth for Recommendation using Multi-view Data***
Xiaotian Han, Chuan Shi, Lei Zheng, Philip, S Yu, Jianxin Li, Yuanfu Lu
APWeb-WAIM International Joint Conference on Web and Big Data (APWeb-WAIM), 2018

[KDDExplo] ***Marginal Nodes Matter: Towards Structure Fairness in Graphs***
Xiaotian Han, Kaixiong Zhou, Ting-Hsiang Wang, Jundong Li, Fei Wang, Na Zou
SIGKDD Exploration, 2023

[Preprint] ***LLM Maybe LongLM: Self-Extend LLM Context Window Without Tuning***
Xiaotian Han*, Hongye Jin*, Jingfeng Yang, Zhimeng Jiang, Zirui Liu, Chia-Yuan Chang, Huiyuan Chen, Xia Hu
2024

[Preprint] ***On the Equivalence of Graph Convolution and Mixup***
Xiaotian Han, Hanqing Zeng, Yu Chen, Shaoliang Nie, Jingzhou Liu, Kanika Narang, Zahra Shakeri, Karthik Abinav Sankararaman, Song Jiang, Madian Khabsa, Qifan Wang, Xia Hu, *2023*

[Preprint] ***GrowLength: Accelerating LLMs Pretraining by Progressively Growing Training Length***
Xiaotian Han*, Hongye Jin*, Jingfeng Yang, Zhimeng Jiang, Chia-Yuan Chang, Xia Hu
2023

[Preprint] ***Do We Really Achieve Fairness with Explicit Sensitive Attributes?***

Xiaotian Han, Zhimeng Jiang, Ninghao Liu, Na Zou, Qifan Wang, Xia Hu
2023

[CPAL2024 Submission] ***You Only Debias Once: Towards Flexible Accuracy-Fairness Trade-offs at Inference Time***
Xiaotian Han, Tianlong Chen, Kaixiong Zhou, Zhimeng Jiang, Zhangyang Wang, Xia Hu
2023

[ICLR2022] *Generalized Demographic Parity for Group Fairness*
Zhimeng Jiang, **Xiaotian Han**, Chao Fan, Fan Yang, Xia Hu,
International Conference on Learning Representations (ICLR), 2022

[Recsys2020] *AutoRec: An Automated Recommender System*
Ting-Hsiang Wang, Qingquan Song, **Xiaotian Han**, Zirui Liu, Jin Haifeng, Xia Hu
ACM Conference on Recommender Systems (Recsys), 2020

[AAAI2020] *FlowScope: Spotting Money Laundering Based on Graphs*,
Xiangfeng Li, Shenghua Liu, Zifeng Li, **Xiaotian Han**, Chuan Shi, Bryan Hooi, He Huang, Xueqi Cheng
AAAI Conference on Artificial Intelligence (AAAI), 2020

[WWWJ2020] *Embedding Geographic Information for Anomalous Trajectory Detection*
Ding Xiao, Li Song, Ruijia Wang, **Xiaotian Han**, Yanan Cai, Chuan Shi
World Wide Web, 2020

[KDD2020] *Metapath-guided Heterogeneous Graph Neural Network for Intent Recommendation*
Shaohua Fan, Junxiong Zhu, **Xiaotian Han**, Chuan Shi, Linmei Hu, Biyu Ma, Yongliang Li
SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2019

[TKDE 2019] *Deep Collaborative Filtering with Multi-aspect Information in Heterogeneous Networks*
Chuan Shi, **Xiaotian Han**, Li Song, Xiao Wang, Senzhang Wang, Junping Du, Philip, S Yu
IEEE Transactions on Knowledge and Data Engineering (TKDE), 2019

[ADMA2018] *Anomalous Trajectory Detection Using Recurrent Neural Network*
Li Song, Ruijia Wang, Ding Xiao, **Xiaotian Han**, Yanan Cai, Chuan Shi
International Conference on Advanced Data Mining and Applications (ADMA), 2018
Best Paper Award

[NeurIPS2023W] Auto-PINN: Understanding and Optimizing Physics-Informed Neural Architecture
Yicheng Wang, **Xiaotian Han**, Chia-Yuan Chang, Daochen Zha, Ulisses Braga-Neto, Xia Hu
NeurIPS2023 AI4Science Workshop, 2023

[AAAI2024-SRRAI] *Chasing Fairness in Graphs: A GNN Architecture Perspective*
Zhimeng Jiang, **Xiaotian Han**, Chao Fan, Zirui Liu, Na Zou, Ali Mostafavi, Xia Hu
AAAI2024, Special Track on Safe, Robust and Responsible AI

[ICLR2024 Submission] *Topology Matters in Fair Graph Learning: a Theoretical Pilot Study*
Zhimeng Jiang, **Xiaotian Han**, Chao Fan, Zirui Liu, Xiao Huang, Na Zou, Ali Mostafavi, Xia Hu
2023

[AAAI2024 Submission] *Towards Assumption-free Bias Mitigation*
Chia-Yuan Chang, Yu-Neng Chuang, Kwei-Herng Lai, **Xiaotian Han**, Xia Hu, Na Zou
2023

[AAAI2024 Submission] *Gradient Rewiring for Editable Graph Neural Network Training*
Zhimeng Jiang, Zirui Liu, **Xiaotian Han**, Qizhang Feng, Hongye Jin, Qiaoyu Tan, Kaixiong Zhou, Na Zou, Xia Hu
2023

[SDM2024 Submission] *Beyond Fairness: Age-Harmless Parkinson's Detection via Voice*
Yicheng Wangang, **Xiaotian Han**, Leisheng Yu, Na Zou
2023

[ACL2024 Submission] *PokeMQA: Programmable knowledge editing for Multi-hop Question Answering*
Hengrui Gu, Kaixiong Zhou, **Xiaotian Han**, Ninghao Liu, Ruobing Wang, Xin Wang
2023

[ICLR2024 *Reducing Communication Overhead in Distributed GNN Training via Client-Server Knowledge Distillation*

Submission] Song Jiang, **Xiaotian Han**, Yinglong Xia, Qifan Wang, Yizhou Sun
2023

Professional Experiences

Texas A&M University, College Station, TX — Sept 2019 – present
Graduate Research Assistant
Advisor: Prof. Xia (Ben) Hu

Amazon, Palo Alto, CA — May 2023 – Aug 2023
Applied Scientist Intern, Large Language Model Alignment
Mentor: Jingfeng Yang

Meta, Menlo Park, CA — Sept 2022 – April 2023
Research Intern, Understanding graph neural networks
Mentor: Qifan Wang

Snap Research, Seattle, WA — May 2022 – Aug 2022
Research Intern, Efficient large-scale graph neural networks
Mentor: Neil Shah

Microsoft Research Asia, Beijing, China — March 2019 - Jun 2019
Research Intern
Hyperparameter search with Bayesian optimization

Open-Source Contributions

Lite LLaMa: Reduced-Scale, Experimental Versions of LLaMa 2, Sole Contributor, 10000+ downloads — 2023
an open-source reproduction of Meta AI's LLaMA and LLaMa 2 with significantly reduced model sizes
https://huggingface.co/ahxt/llama2

FFB: Fair Fairness Benchmark, Sole Contributor — 2023
A PyTorch-based framework for evaluating the fairness of machine learning models.
https://github.com/ahxt/fair_fairness_benchmark

LLMs Servery Repo, Main Contributor, 7100+ stars — 2023
A curated list of practical guide resources of LLMs.
https://github.com/Mooler0410/LLMsPracticalGuide

AutoRec: An Automated Recommender System, Main Contributor — 2020
A Keras-based implementation of automated recommendation algorithms for both rating prediction and Click Through Rate task.
https://github.com/datamllab/AutoRec

AutoKeras: An AutoML System Based on Keras, Contributor, 9000+ stars — 2020
An AutoML system based on Keras with the goal to make machine learning accessible to everyone.
https://autokeras.com/

Mentored Students

Yicheng Wang, *Ph.D. student at TAMU* — 2020-current
Auto-PINN: Understanding and Optimizing Physics-Informed Neural Architecture -> NeurIPS2023 AI4Science Workshop
Beyond Fairness: Age-Harmless Parkinson's Detection via Voice -> SDM2024 submission

Hongye Jin, *Ph.D. student at TAMU* — 2021-current
GrowLength: Accelerating LLMs Pretraining by Progressively Growing Training Length -> ICLR2024 submission
Project on LLMs context window extension

Chia-Yuan Chang, *Ph.D. student at TAMU* — 2021-current
Towards Assumption-free Bias Mitigation -> AAAI2024 submission
Project on the effect of position embedding in LLMs

Zhimeng Jiang, *Ph.D. student at TAMU* — 2020-2023
Generalized Demographic Parity for Group Fairness -> ICLR2022
Topology Matters in Fair Graph Learning: a Theoretical Pilot Study -> AAAI2024 submission

Ting-Hsiang Wang, *Master student at TAMU* 2019-2020
AutoRec: An Automated Recommender System -> Recsys2020 Demo
Currently Data Scientist at Goldman Sachs

Devi Sandeep Endluri, *Master student at TAMU* 2019-2020
Project on Named Entity Recognition for Material Science
Currently Data Scientist at Microsoft

Talks & Presentations

Invited Talks

Graph Learning with Low Resources, VISA Research	2023
Algorithmic Fairness in Finance: Problems, Methods and Benchmarks, MLF@KDD2023	2023
Graph Neural Network Training Acceleration, LOGS	2023
Graph Neural Network Training Acceleration, AI-TIME	2023
Graph Augmentation for Graph Classification, Meta AI	2022
Graph Augmentation for Graph Classification, Wormpex AI Research	2022
Graph Augmentation for Graph Classification, Central South University	2022

Conference Presentations

$\mathcal{G}$-Mixup: Graph Augmentation for Graph Classification, ICML, Baltimore, MD	2022
Graph Representation Learning via Unsupervised Rate Reduction Maximization, TheWebConf, Online	2022
Aspect-Level Deep Collaborative Filtering via HIN, IJCAI, Stockholm, Sweden	2018

Professional Services

Program Committee Member

International Conference on Learning Representations (**ICLR**)	2024
ACM International Conference on Web Search and Data Mining (**WSDM**)	2024
Conference on Information and Knowledge Management (**CIKM**)	2023
International Conference on Machine Learning (**ICML**)	2022,2023
Annual Conference on Neural Information Processing Systems (**NeurIPS**)	2022,2023
AAAI Conference on Artificial Intelligence (**AAAI**)	2021,2022,2023,2024
International Joint Conference on Artificial Intelligence(**IJCAI**)	2021,2023
The ACM Web Conference (**WWW**)	2023
Empirical Methods in Natural Language Processing (**EMNLP**)	2023
International Conference on Data Mining (**ICDM**)	2022
ACM SIGKDD Conference on Knowledge Discovery and Data Mining (**KDD**)	2023
TinyPaper Track @ International Conference on Learning Representations	2023
MLG Workshop @ International Conference on Data Mining	2022

Reviewer

ACM Transactions on Intelligent Systems and Technology (**TIST**)	2023
Transactions on Machine Learning Research (**TMLR**)	2023
IEEE Transactions on Knowledge and Data Engineering (**TKDE**)	2023
IEEE Transactions on Neural Networks and Learning Systems (**TNNLS**)	2023
Neurocomputing	2023
CCF Transactions on Pervasive Computing and Interaction (**TCPI**)	2019

Volunteer

International Conference on Machine Learning (**ICML**)	2022
The North American Chapter of the Association for Computational Linguistics (**NAACL**)	2022

Session Chair

The ACM Web Conference (**WWW**)	2023
International Conference on Machine Learning (**ICML**)	2022

Last updated on January 25, 2024