

HAITIAN JIANG

☎ +1 (332)-323-2919 ✉ haitian.jiang@nyu.edu 🌐 github.com/haitian-jiang 👤 haitian-jiang.github.io

🎓 EDUCATION

New York University Ph.D. in Computer Science Advisor: Prof. Jinyang Li & Prof. Aurojit Panda Research Interest: Machine Learning Systems, Reliability, Machine Learning	09/2023 – Present <i>New York, NY</i>
Fudan University B.S. in Data Science (Summa Cum Laude) Advisor: Prof. Zengfeng Huang	09/2019 – 06/2023 <i>Shanghai, China</i>
University of California, San Diego (UCSD) Exchange Student, Computer Science and Engineering	09/2021 – 12/2021 <i>La Jolla, CA</i>

📖 PUBLICATION & PREPRINTS

- **TTrace: Lightweight Error Checking and Diagnosis for Distributed Training.**
Jiang H., Zhu S., Zhang Z., Song Z., Fu X., Jia Z., Wang Y. and Li J. Submitted to OSDI.
- **DiskGNN: Bridging I/O Efficiency and Model Accuracy for Out-of-Core GNN Training.**
Liu R., Wang Y., Yan X., Jiang H., Cai Z., Wang M., Tang B. and Li J. SIGMOD 2025.
- **MuseGNN: Forming Scalable, Convergent GNN Layers that Minimize a Sampling-Based Energy.**
Jiang H., Liu R., Huang Z., Wang Y., Yan X., Cai Z., Wang M. and Wipf D. ICLR 2025.
- **FreshGNN: Reducing Memory Access via Stable Historical Embeddings for Graph Neural Network Training.**
*Huang K. *, Jiang H. * (Equal Contribution), Wang M., Xiao G., Wipf D., Song X., Gan Q., Huang Z., Zhai J. and Zhang Z.* VLDB 2024.
- **SGFormer: Simplifying and Empowering Transformers for Large-Graph Representations.**
Wu Q., Zhao W., Yang C., Zhang H., Nie F., Jiang H., Bian Y. and Yan J. NeurIPS 2023.

💻 RESEARCH & INTERN EXPERIENCES

Systems Group, New York University <i>Research Assistant</i> Machine Learning Systems Advisor: Prof. Jinyang Li & Prof. Aurojit Panda <ul style="list-style-type: none">• Automatic Searching for Kernel Fusion and Scheduling in Machine Learning Compilers	09/2023 – Present <i>New York</i>
Neuron Science Team, Amazon Web Services <i>Applied Scientist Intern</i> Machine Learning Systems Advisor: Yida Wang & Zhen Zhang <ul style="list-style-type: none">• Robust low-precision inference• TTrace: Lightweight Error Checking and Diagnosis for Distributed Training<ul style="list-style-type: none">◦ Developed the first testing system for silent bugs in distributed training.◦ Designed novel method to identify floating-point error tolerance with theoretical basis.◦ Identified 14 (3 new) bugs in production-ready framework Megatron-LM with TTrace.	06/2025 – 12/2025, 05/2024 – 12/2024 <i>Santa Clara</i>
DGL Group, Amazon Web Services AI Lab <i>Applied Scientist Intern</i> Machine Learning Theory and System Advisor: David Wipf & Minjie Wang <ul style="list-style-type: none">• MuseGNN: Forming Scalable, Convergent GNN Layers that Minimize a Sampling-Based Energy<ul style="list-style-type: none">◦ Expanded a widely-used unfolded GNN framework to incorporate offline sampling into the architecture-inducing energy function design.	06/2022 – 08/2023 <i>Shanghai</i>

- Demonstrated the model possess convergence properties of the bi-level optimization process.
- Empirically extended the unfolded GNN framework to large graphs and achieved SOTA performance for homogeneous graph models applied to the largest open graph dataset.
- **FreshGNN**: Graph Neural Network (GNN) training system with historical cache
 - Accelerated and scaled up GNN training on large graphs with a selective historical cache to reduce computation and data movement.
 - Designed managing policies based on gradients and staleness for cached nodes.

Lab of Intelligent Information Processing, Fudan University

01/2022 – 10/2022

Research Assistant | Graph Neural Network

Shanghai

Advisor: Prof. Zengfeng Huang

- **Scalable Graph Neural Networks using Subgraph Summarization**

McAuley Lab, University of California, San Diego

10/2021 – 08/2022

Research Assistant | Recommender System, Natural Language Processing

La Jolla

Advisor: Prof. Julian McAuley

- **Recommendation System with Faithful Textual Explanation**

Keen Lab, Tencent Technology

02/2021 – 07/ 2021

Software Engineer | Reverse Engineering

Shanghai, China

- Reverse-engineering on IoT device firmware; implementing an asynchronous crawler system.

TEACHING EXPERIENCE

TA: Natural Language Processing, CSCI-GA.2590, Spring 2025, New York University

TA: Honors Analysis of Algorithms, CSCI-GA.3110, Fall 2024, 2025, New York University

TA: Advanced Computer Graphics, CSCI-GA.2274, Spring 2024, New York University

TA: Big Data and Machine Learning Systems, CSCI-GA.3033(077), Spring 2024, New York University

TA: Advanced Big Data Analytics, DATA130014, Spring 2023, Fudan University

TA: Algorithm and Data Structures (Honor), DATA130023H, Fall 2022, Fudan University

ACADEMIC SERVICES

Reviewer: NeurIPS 2024, 2025; ICLR 2025; MLSys 2025; ICML 2025; SIMODS

Artifact Evaluation: ATC 2025

Organizer: NYU MLSys Seminar

HONORS & AWARDS

Henry M. MacCracken Fellowship **2023**

China National Scholarship (top 1%) **2021**

National Second Prize (top 3%), CUMCM **2020**

(Contemporary Undergraduate Mathematical Contest in Modeling)

SKILLS

Languages: Mandarin (Native speaker), English (Proficient)

Programming: Python, C/C++, CUDA, Triton, Go, MATLAB, R, SQL, PHP, Bash

Frameworks: Linux, Git, PyTorch, GPU programming, OpenMP, Spark, Docker, Web Crawler