Shichang Zhang

Contact Information

Address: 150 Western Ave. SEC 6.220, Boston, MA 02134
Email: shzhang@hbs.edu
Website: https://shichangzh.github.io/

Work Experience

Harvard University — Cambridge, MA
Postdoctoral Fellow — Aug. 2024 - Present

Education

University of California, Los Angeles — Los Angeles, CA
Ph.D. in Computer Science — June 2024

Stanford University — Stanford, CA
M.S. in Statistics — Apr. 2019

University of California, Berkeley — Berkeley, CA
B.A. in Statistics — May 2017
Honors: Honors in Statistics, High Distinction

Research Interests

Explainable AI, Trustworthy AI, Data Attribution, Mechanistic Interpretability, Large Language Models, Graph Data Mining

Honors and Awards

NENLP **Outstanding Paper**	2025
KDD Outstanding Reviewer (Top 10%, twice, in Aug and Feb)	2025
Amazon PhD Fellowship	2023
J.P.Morgan Chase AI PhD Fellowship	2023
KDD Excellence in Reviewing (Top 30 of 1551)	2023
Snap Research Fellowship Honorable Mention	2022
ICML Top Reviewer (Top 10%)	2022
UCLA Graduate Division Fellowship	2021

Publications

Refereed Publications:

1. How Post-Training Reshapes LLMs: A Mechanistic View on Knowledge, Truthfulness, Refusal, and Confidence
   Hongzhe Du*, Weikai Li*, Min Cai, Karim Saraipour, Zimin Zhang, Himabindu Lakkaraju, Yizhou Sun, **Shichang Zhang** (*equal contribution)
   Conference on Language Modeling (**COLM**), (**NENLP Outstanding Paper**), 2025
2. Automated Molecular Concept Generation and Labeling with Large Language Models
   Zimin Zhang*, Qianli Wu*, Botao Xia*, Fang Sun, Ziniu Hu, Yizhou Sun, **Shichang Zhang** (*equal contribution)
   International Conference on Computational Linguistics (**COLING**), 2025
3. An Explainable AI Approach using Graph Learning to Predict ICU Length of Stay
   Tianjian Guo, Indranil Bardhan, Ying Ding, **Shichang Zhang**
   Information Systems Research (**ISR**), 2024

4. Motif-driven Contrastive Learning of Graph Representations
   Shichang Zhang[*], Ziniu Hu[*], Arjun Subramonian, Yizhou Sun ([*]equal contribution)
   IEEE Transactions on Knowledge and Data Engineering (**TKDE**), 2024
5. Predicting and Interpreting Energy Barriers of Metallic Glasses with Graph Neural Networks
   Haoyu Li[*], **Shichang Zhang**[*], Longwen Tang, Yizhou Sun ([*]equal contribution)
   International Conference on Machine Learning (**ICML**), 2024
6. PaGE-Link: Graph Neural Network Explanation for Heterogeneous Link Prediction
   Shichang Zhang, Jiani Zhang, Xiang Song, Soji Adeshina, Da Zheng, Christos Faloutsos, Yizhou Sun
   The Web Conference (**WWW**), 2023
7. GStarX: Explaining Graph Neural Networks with Structure-Aware Cooperative Games
   Shichang Zhang, Yozen Liu, Neil Shah, Yizhou Sun
   Advances in Neural Information Processing Systems (**NeurIPS**), 2022
8. Graph-less Neural Networks, Teach Old MLPs New Tricks via Distillation
   Shichang Zhang, Yozen Liu, Yizhou Sun, Neil Shah
   International Conference on Learning Representations (**ICLR**), 2022
9. Weak Models Can be Good Teachers: A Case Study on Link Prediction with MLPs
   Zongyue Qin, **Shichang Zhang**, Mingxuan Ju, Tong Zhao, Neil Shah, Yizhou Sun
   Learning on Graphs Conference (**LOG**), 2025
10. FUSE: Measure-Theoretic Compact Fuzzy Set Representation for Taxonomy Expansion
    Fred Xu, Song Jiang, Zijie Huang, Xiao Luo, **Shichang Zhang**, Yuanzhou Chen, Yizhou Sun
    Findings of the Association for Computational Linguistics (**ACL Findings**), 2024
11. SciBench Evaluating College-Level Scientific Problem-Solving Abilities of Large Language Models
    Xiaoxuan Wang[*], Ziniu Hu[*], Pan Lu[*], Yanqiao Zhu[*], Jieyu Zhang, Satyen Subramaniam, Arjun R Loomba, **Shichang Zhang**, Yizhou Sun, Wei Wang ([*]equal contribution)
    International Conference on Machine Learning (**ICML**), 2024
12. Laplacian Score Benefit Adaptive Filter Selection for Graph Neural Networks
    Yewen Wang, **Shichang Zhang**, Junghoo Cho, Yizhou Sun
    SIAM International Conference on Data Mining (**SDM**), 2024
13. Linkless Link Prediction via Relational Distillation
    Zhichun Guo, William Shiao, **Shichang Zhang**, Yozen Liu, Nitesh Chawla, Neil Shah, Tong Zhao
    International Conference on Machine Learning (**ICML**), 2023
14. Graph Condensation for Graph Neural Networks
    Wei Jin, Lingxiao Zhao, **Shichang Zhang**, Yozen Liu, Jiliang Tang, Neil Shah.
    International Conference on Learning Representations (**ICLR**), 2022

Preprints:

1. Who Gets Credit or Blame? Attributing Accountability in Modern AI Systems
   Shichang Zhang, Hongzhe Du, Jiaqi W. Ma, Himabindu Lakkaraju
   (Under PNAS Review), 2025
2. Towards Unified Attribution in Explainable AI, Data-Centric AI, and Mechanistic Interpretabil ity
   Shichang Zhang, Tessa Han, Usha Bhalla, Himabindu Lakkaraju
   (Under Nature Machine Intelligence Review), 2025

3. Generalized Group Data Attribution
   Dan Ley*, **Shichang Zhang***, Suraj Srinivas, Gili Rusak, Himabindu Lakkaraju (*equal contribution)
   (Under Management Science Review), 2024
4. From Indirect Object Identification to Syllogisms: Exploring Binary Mechanisms in Transformer Circuits
   Karim Saraipour, **Shichang Zhang**
   (ORIGen@COLM), 2025
5. On the Retention of Edited Knowledge in Fine-Tuned Language Models
   Fufang Wen, **Shichang Zhang**
   (ORIGen@COLM), 2025
6. Hierarchical Compression of Text-Rich Graphs via Large Language Models
   Shichang Zhang, Da Zheng, Jiani Zhang, Qi Zhu, Xiang Song, Soji Adeshina, Christos Faloutsos, George Karypis, Yizhou Sun
   (Preprint), 2024
7. Self-Control of LLM Behaviors by Compressing Suffix Gradient into Prefix Controller
   Min Cai, Yuchen Zhang, **Shichang Zhang**, Fan Yin, Difan Zou, Yisong Yue, Ziniu Hu
   (MI@ICML), 2024
8. Parameter-Efficient Tuning Large Language Models for Graph Representation Learning
   Qi Zhu, Da Zheng, Xiang Song, **Shichang Zhang**, Bowen Jin, Yizhou Sun, George Karypis
   (Preprint), 2024
9. Efficient Ensembles Improve Training Data Attribution
   Junwei Deng*, Ting-Wei Li*, **Shichang Zhang**, Jiaqi Ma (*equal contribution)
   (DMLR@ICML), 2024
10. A Survey on Graph Neural Network Acceleration: Algorithms, Systems, and Customized Hardware
    Shichang Zhang, Atefeh Sohrabizadeh, Cheng Wan, Zijie Huang, Ziniu Hu, Yewen Wang, Yingyan (Celine) Lin, Jason Cong, Yizhou Sun
    (Under ACM Computing Review, Major Revision), 2023

Media Coverage

Unifying AI Attribution: A New Frontier in Understanding Complex Systems — June 2025
D ^3 Insights & Experiences

ChatGPT has entered the classroom: how LLMs could transform education — Nov. 2023
Nature News Feature

Teaching Experience

Instructor, University of California, Los Angeles
CS97: Introduction to Data Science — Summer 2024

Teaching Assistant, University of California, Los Angeles
CS145: Introduction to Data Mining — Fall 2020, Fall 2021
CS32: Introduction to Computer Science II — Spring 2021

Mentorship

Arjun Subramonian (UCLA Undergrad → UCLA PhD) — Mar. 2020 - Mar. 2021
Qianli Wu (UCLA Undergrad → Amazon SDE) — Mar. 2023 - Mar. 2024
Haoyu Li (UCLA Undergrad → UIUC PhD) — Mar. 2023 - July 2024
Gaotang Li (UMich Undergrad → UIUC PhD) — Oct. 2023 - June 2024
Botao Xia (UCLA Undergrad → UCLA Master) — Oct. 2023 - Aug. 2024

	Zimin Zhang (UCLA Undergrad → UIUC Master)	Oct. 2023 - Present
	Min Cai (Shenzhen University Master → UAlberta PhD)	Nov. 2023 - Present
	Hongzhe Du (UCLA Master)	Mar. 2024 - Present
	Karim Saraipour (UCLA Master)	Apr. 2024 - Present
	Fufang Wen (Columbia Master)	June 2024 - Present
	Weikai Li (UCLA Ph.D.)	Sept. 2024 - Present
	Dan Ley (Harvard Ph.D.)	Sept. 2024 - Present
	Ethan Ji (UCLA Master)	June 2025 - Present
	Terry Zhou (Harvard Master)	Sept. 2025 - Present

Talks

How Post-Training Reshapes LLMs	
New England NLP Meeting	Apr 2025
Peering into The Mind of AI	
Seminar at Georgia Institute of Technology	Apr 2025
Interpreting AI Systems Through Features, Data, and Model Components	
Data Mining Seminar at Emory	Apr 2025
Explainable AI for Graph Data and More	
AI4LIFE Group at Harvard	Feb 2024
Graph Neural Network Explanation for Heterogeneous Link Prediction	
Amazon Trans.AI Research Talks	July 2023
International World Wide Web Conference	May 2023
Structure-Aware Graph Neural Network Explanation	
AI Time NeurIPS Talk Series	Feb 2023
Graph-less Neural Networks	
NVIDIA GNN Reading Group	May 2022

Academic Service

Conference Area Chair:

ACL ARR - Association for Computational Linguistics Rolling Review	2025

Conference Reviewer/Program Committee:

NeurIPS - Advances in Neural Information Processing Systems	2021 - 2025
ICML - International Conference on Machine Learning	2022 - 2025
ICLR - International Conference on Learning Representations	2024 - 2026
KDD - ACM SIGKDD Knowledge Discovery and Data Mining	2020, 2023 - 2025
AAAI - AAAI Conference on Artificial Intelligence	2023 - 2025
WSDM - ACM International Web Search and Data Mining Conference	2023 - 2025
CIKM - ACM Conference on Information and Knowledge Management	2022 - 2023
SDM - SIAM International Conference on Data Mining	2024
LOG - Learning on Graphs Conference	2023
ICDM - IEEE International Conference on Data Mining	2021

Journal Reviewer:

TPAMI - IEEE Transactions on Pattern Analysis and Machine Intelligence
TKDD - ACM Transactions on Knowledge Discovery from Data

TKDE - IEEE Transactions on Knowledge and Data Engineering
TNNLS - IEEE Transactions on Neural Networks and Learning Systems
TAI - IEEE Transactions on Artificial Intelligence
Management Science

Workshop Organizer:
Workshop on Regulatable Machine Learning @ NeurIPS 2024 - 2025

Reading Group Organizer:
UCLA Data Mining Reading Group 2022 - 2024

Industry Work Experience

Amazon Web Service (AWS) Santa Clara, CA
Applied Scientist Intern, Graph Machine Learning Team June 2023 - Nov. 2023

- Proposed a framework for applying LLMs to text-rich graph data with hierarchical neighborhood compression, which allows LLMs to leverage the graph structure and handle long input text features gathered in a rich neighborhood.
- The proposed method outperformed traditional graph ML models on node classification benchmarks and will be incorporated into the Amazon DGL project.

Amazon Web Service (AWS) Santa Clara, CA
Applied Scientist Intern, Graph Machine Learning Team June 2022 - Oct. 2022

- Proposed a new framework to explain GNN link prediction for recommendation on graph data, which improves user trust in the model and helps developers debug the model. Work published in WWW 2023.
- The implemented framework will be incorporated into the Amazon Neptune ML project in production.

Snap Research Los Angeles, CA
Research Intern, Computational Social Science Team June 2021 - Sept. 2021

- Proposed a cross-model distillation framework to transfer knowledge from GNNs to MLPs, which speeds up model inference by 179 times and facilitates model deployment on latency-constraint applications. Work published in ICLR 2022.
- Worked on condensing large-scale training graphs to small synthetic graphs by over 90% reduction rate while maintaining competitive model performance for GNNs trained from scratch, which significantly saves storage space and achieves efficient continual learning. Work published in ICLR 2022.

WeWork Inc. Palo Alto, CA
Data Scientist Intern, Research and Applied Science Team June 2019 - Sept. 2019

- Implemented a data processing pipeline in SQL and Python for data querying, data cleaning, and feature engineering.
- Trained a Gradient Boosted Tree model on two million customer data to predict occupancy rate for WeWork buildings and achieved 0.093 MAE on the test set.
- Presented the pricing model as a selected outstanding project to the Research and Applied Science team including the VP.

Skills

Programming: Python (PyTorch, Hugging Face, DGL), C++, R, Java, Linux, Git
Natural Languages: Mandarin Chinese (Native), English (Proficient)