Updated Dec 2025

ZIYU "SYLVIA" ZHANG

GitHub : SylviaZiyuZhang • sylziyuz@csail.mit.edu

EDUCATION

MASSACHUSETTS INSTITUTE OF TECHNOLOGY Cambridge, MA

Ph.D. student, CSAIL, EECS Sep 2023 - present

Co-advised by Professor Julian Shun and Professor Michael Cafarella

Master of Science June 2025

Thesis: Graph-based Vector Search Algorithms for Retrieval-Augmented AI Systems

CARNEGIE MELLON UNIVERSITY Pittsburgh, PA

B.S., Computer Science Aug 2018 – Dec 2021

Concentration in Algorithms and Complexity

University Honors

AWARDS AND HONORS

NSF Graduate Research Fellow, 2023 – Present

Jacobs Presidential Fellowship, MIT, AY 2023-2024

Senior Leadership Recognition, Carnegie Mellon University, 2022

Phi Beta Kappa, Inducted 2022

RESEARCH EXPERIENCE

Vector set search for new challenges in information retrieval Sep 2025 – Present

Graduate research at MIT Advisors: Prof. Mike Cafarella, Prof. Julian Shun

Evaluating existing vector set search algorithms on different workloads. Exploring how vector set search could help solve challenging information retrieval problems and designing new efficient algorithms for vector set search.

Fully dynamic graph-based approximate nearest neighbor search Sep 2023 – Present

Graduate research at MIT Advisor: Prof. Julian Shun

Designed and implemented algorithms for adapting graph-based approximate nearest neighbor indexes to support concurrent search, insert, and delete queries efficiently and robustly.

Storage optimizations for approximate nearest neighbor search June 2025 – Present

Data Systems Group @ Microsoft Research / Azure Data Mentor: Phil Bernstein

Designed and implemented algorithms for optimizing graph-based nearest neighbor indexes for varying storage primitives.

Benchmarking AI systems for data science Oct 2024 – Oct 2025

Graduate research at MIT Data Systems Group

Lead the development of KramaBench, a benchmark for AI systems on end-to-end data science tasks involving data discovery, data retrieval, data wrangling, and data analysis.

Causal reasoning for systems diagnosis Sep 2023 – May 2024

Graduate research at MIT Advisor: Prof. Mike Cafarella

Contributed to a project on reliably quantitatively identifying causes of system anomalies using log data and causal inference. Designed algorithms for identifying potential causal graphs while minimizing the number of oracle queries.

Low-rank tensor decomposition Jan 2021 – Dec 2021

Undergraduate Independent Study, CMU Advisor: Prof. David Woodruff

Worked on research on sketch-based tensor approximation, including algorithms for general tensor networks and a few other common decompositions, as well as hardness results.

Mass spectrometry matching Aug 2019 – Mar 2020

Undergraduate Research Assistant, CMU Advisor: Prof. Hosein Mohimani

Worked on improving and benchmarking high performance algorithms on mass spectrometry matching in C++.

RESEARCH INTERESTS

I am interested in efficient data algorithms and data systems design that practically support the reliability of augmented data retrieval and discovery in aspects like freshness, truthfulness, and semantic awareness.

PUBLICATIONS

- **Near-Linear Time and Fixed-Parameter Tractable Algorithms for Tensor Decompositions.** Arvind V. Mahankali, David P. Woodruff, **Ziyu Zhang**. (In alphabetical order) *ITCS 2024.*
- **Sawmill: From Logs to Causal Diagnosis of Large Systems.** Markos Markakis, Anbo Chen, Brit Youngmann, Trinity Gao, **Ziyu Zhang,** Rana Shahout, Peter Baile Chen, Chunwei Liu, Ibrahim Sabek, Michael Cafarella**.** S*IGMOD 2024 Demo*.
- **Press ECCS to Doubt (Your Causal Graph).** Markos Markakis, **Ziyu Zhang**, Rana Shahout, Trinity Gao, Chunwei Liu**,** Ibrahim Sabek, Michael Cafarella**.** *SIGMOD 2024 GUIDE-AI*. **(Best Paper)**
- **From Logs to Causal Inference: Diagnosing Large Systems.** Markos Markakis, Brit Youngmann, Trinity Gao, **Ziyu Zhang,** Rana Shahout, Peter Baile Chen, Chunwei Liu, Ibrahim Sabek, Michael Cafarella**.** *To appear in VLDB 2025.*

MANUSCRIPTS

- **Using Deep Learning Sequence Models to Identify SARS-CoV-2 Divergence. [link]**
  Yanyi Ding, Zhiyi Kuang, Yuxin Pei, Jeff Tan, Ziyu Zhang, Joseph Konan. (Students in alphabetical order)
  Graduate Level course project: CMU 11-785 Introduction to Deep Learning
- **Locality Sensitive Hashing on Causal Graphs.** Joshua Engels, **Ziyu Zhang**. (Equal contribution) (In submission)
  Graduate Level course project: MIT 6.838 Geometric Computing
- **Clean-ANN: Robust and Efficient Dynamism of Graph-based Approximate Nearest Neighbor Search. Ziyu Zhang**, Yuanhao Wei, Joshua Engels, Julian Shun. (In submission)
- **KRAMABENCH: A Benchmark for AI Systems on Data Intensive Tasks.** Eugenie Lai*, Gerardo Vitagliano*, **Ziyu Zhang***, Om Chabra et al. (In submission)

PRESENTATIONS

Near-linear Time and Fixed Parameter Tractable Algorithms for Tensor Decompositions, ITCS, University of California Berkeley, Berkeley, CA. January 2024.
Press ECCS to Doubt (Your Causal Graph), Northeast Database Day, Boston University, Boston, MA. January 2025.

TEACHING

CARNEGIE MELLON UNIVERSITY Pittsburgh, PA
Undergraduate Teaching Assistant, School of Computer Science Jan 2019 – Dec 2021
15-451 Design and Analysis of Algorithms, Head Teaching Assistant Fall 2021
15-451 Design and Analysis of Algorithms 2020-2021
15-151 Mathematical Foundations for Computer Science Fall 2019
15-122 Principles of Imperative Programming Spring 2019

MASSACHUSETTS INSTITUTE OF TECHNOLOGY Cambridge, MA
Graduate Teaching Assistant, EECS Sep 2024 - Dec 2024
Database Systems

LEADERSHIP AND SERVICE

- Treasurer, MIT Ballroom Dance Team, 2025-2026
- Webmaster and Media Officer, MIT EECS, Graduate Women in Course 6, 2024
- Coaching assistant, Egyptian national team for European Girls' Olympiad in Informatics, 2023
- Student Advisory Committee, Carnegie Mellon University, School of Computer Science, 2019-2020

PROFESSIONAL EXPERIENCE

MICROSOFT RESEARCH Redmond, WA
Research Intern Summer 2025
Designed and implemented algorithms for optimizing graph-based nearest neighbor indexes for varying storage primitives.

OTTERTUNE Pittsburgh, PA
Software Engineer Summer 2021, Feb 2022 – Aug 2023
Conducted research on database health checks for PostgresSQL and MySQL. Developed a prototype end-to-end pipeline from database metrics collection to database health check result generation and storage. Designed and implemented several critical user management features. Developed some database related features. Helped with improving backend performance, architecture design, and technical maintenance. Mentored a summer intern on data engineering and analytics.

GOOGLE Mountainview, CA
STEP Intern Summer 2020

Developed a mentor-mentee matching web-app in Java and Javascript. Designed the matching algorithm and NoSQL storage layer.

BYTEDANCE Shanghai, China
Algorithms Engineering Intern, Audio Algorithms Team Summer 2019

Experimented with voice texture transformation framework research results, including GAN, variational encoder, and normalizing flow. Implemented chorus partn detection for pop songs in Python.

SKILLS

PROGRAMMING

Proficient: C++, golang, Python, Rust.
Familiar: React/Javascript, SML/OCaml, Java, PASCAL.

TECHNICAL

Adobe Photoshop, Illustrator, Premier, Audition. AWS, Google Cloud, Kubernetes.

LANGUAGES

Mandarin Chinese (native), English (bilingual proficiency), French (advanced), Russian (intermediate)