

PROGRAMMING SKILLS

Languages: Python, SQL, Go, Java, C++, JavaScript, Rust

Technologies: CUDA, AWS, Pytorch, MongoDB, Boost-Python, React, Faiss, TensorFlow, PostgreSQL, Azure

EDUCATION

University of Rochester

Ph.D. Computer Science

Rochester, NY

August 2023 - Present

University of Rochester

M.Sc. Computer Science (3.8/4.0)

Rochester, NY

Full Scholarship & Research Assistantship

August 2021 - August 2023

University of Washington

B.Sc. Mathematics (3.0/4.0)

Seattle, WA

September 2017 - June 2021

EXPERIENCE

MIT Computer Science and Artificial Intelligence Laboratory (CSAIL)

Remote

Graduate Summer Researcher - Advisors : Quanguan Liu & Julian Shun

June 2023 - Present

- Working with the ParAlg group to implement a benchmark suite for distributed privacy-preserving local graph algorithms.

Paris Lodron Universität Salzburg - Database Group

Salzburg, Austria

Graduate Summer Researcher - Advisor : Martin Schäler

June 2023 - August 2023

- Developing a comprehensive research plan to identify language patterns and semantic similarities in biblical texts across multiple languages.
- Collaborating with a team of linguists and researchers to integrate natural language processing and set similarity search algorithms into the BOSS project to improve search accuracy.

University of Rochester

Rochester, NY

Graduate Research Assistant - Advisor : Fatemeh Nargesian

July 2021 - May 2023

- Developed and implemented **KOIOS**, a **novel, exact, efficient, and generic** filter verification system for **top-k** set similarity search using semantic overlap that achieves at least **5.5x faster performance** than the current standard, as it can prune up to **95%** of the sets using inexpensive filters based on the bounds on the graph matching score. (Paper accepted at **IEEE ICDE 2023**)
- Developed a novel algorithm to select a coreset based on coverage and fairness criteria that achieves a **speedup** of at least **400x** compared to the current state of the art machine learning based data synthesis technique across three datasets: MNIST, FashionMNIST, CIFAR10. Conducted experiments to evaluate the performance of proposed coreset selection approach, achieving an average accuracy of at least **70%** with only **24%** of the data.
- Developing a **scalable system** utilizing data integration techniques to support **fast aggregate query search** over open world data. (Workshop Paper accepted at **HILDA 2023**)

Caltech - Anima AI Lab

Pasadena, CA

Undergraduate ML Researcher - Advisors : Forough Arabshahi & Animashree Anandkumar

June 2020 - April 2021

- Developed a novel recursive neural network architecture, Tree Stack Memory Units (Tree-SMU), to enable compositional generalization in the domain of mathematical reasoning.
- Evaluated the generalization of Tree-SMU on four different compositionality tests. We showed that Tree-SMU consistently outperforms the compositional generalization of powerful baselines such as transformers, tree transformers and Tree-LSTMs.

University of Washington Database Group

Seattle, WA

Undergraduate Research Assistant - Advisor: Brandon Haynes

January - December 2020

- Developed a high-performance Python API for **LightDB**, **enhancing query speed** and **enabling access** to all forms of video data including VR and AR videos.
- **Reduced device transfer time** through optimized mapping of the Python API to low-level constructs in LightDB, resulting in faster query execution for users.
- Implemented the **boost-python** framework to enable a wider audience to use the query expression feature in LightDB **increasing user adoption**.

Undergraduate Research Assistant - Advisors : Batya Kenig & Dan Suciu

April - September 2019

- Developed Maimon, the first system for discovering approximate **MultiValued Dependencies** and acyclic schemas in the data, and defined principled notion of approximate data dependencies based on information theory, and study its properties.
- Optimized Maimon to **reduce the number of file scans** by taking advantage of **Information Theory** to prune out MVDs and calculate the entropies using the already discovered MVDs; compared the performance tradeoff between in-memory database system (H2) and MySQL. (Paper accepted at **ACM SIGMOD 2020**)

PUBLICATIONS

Approximate Query Answering over Open Data: Mengqi Zhang, **Pranay Mundra**, Chukwubikem Chikweze, Fatemeh Nargesian, Gerhard Weikum. (HILDA 2023)

KOIOS : Top- k Semantic Overlap Set Search: **Pranay Mundra**, Jianhao Zhang, Fatemeh Nargesian, and Nikolaus Augsten. (IEEE ICDE 2023)[Paper]

Mining approximate acyclic schemes from relations: Batya Kenig, **Pranay Mundra**, Guna Prasaad, Babak Salimi, and Dan Suciu. (ACM SIGMOD 2020)[Paper]

Compositional Generalization with Tree Stack Memory Units: Forough Arabshahi, Zhichu Lu, **Pranay Mundra**, Sameer Singh, Animashree Anandkumar. (arXiv Preprint)[Paper]

RELEVANT COURSES

Computer Science: Computer Programming I, II; Introduction to Database Systems; Database Systems Internals; Data Structures & Algorithms; Linux Fundamentals; Introduction to Artificial Intelligence; Advanced Algorithms; Analytical Methods in Computer Science; Machine Learning, Parallel & Distributed Systems; Computer Networks; Data Mining; Computational Complexity

Mathematics: Honors Calculus I, II, III; Real Analysis I, II; Linear Analysis; Probability I, II; Differential Equations, Linear Algebra, Numerical Analysis I, II; Modern Algebra I, II; Combinatorial Theory I, II.

TEACHING EXPERIENCE

University of Rochester

Rochester, NY

Department of Computer Science : Graduate Teaching Assistant

- CSC 261/461 - **Database Systems** (Spring 2023)
- CSC 263/463 - **Data Management Systems** (Spring 2022)
- CSC 244/444 - **Knowledge Representation in AI**, (Fall 2022)

University of Washington

Seattle, WA

Paul G. Allen School of Computer Science & Engineering : Undergraduate Teaching Assistant

- CSE 444 - **Database Systems Internals**, (Winter 2021)
- CSE 414/344 - **Introduction to Database Systems**, (Fall 2020, Winter 2020, & Spring 2019)

PROJECTS

AquaDB & SimpleDB: Implemented a multi-user transactional database server written in Go and Java respectively.

Husky Map Server: Created a google map for the University of Washington campus, which shows the shortest path between two locations.

Flight Booking Application: Implemented a flight booking service with user management, transaction support, itinerary search & reservations.

Spotify Song Explorer: Web Application that allows visualization of different audio features for Top 50 songs, fetched using the Spotify API.

Gene Regulatory System: Optimized the code to leverage GPU parallelism using the CUDA framework for the following paper: McMurray et al. Gene network modeling via TopNet reveals functional dependencies between diverse tumor-critical mediator genes.