Jinjin Zhao

jinjinz.com | 312-358-4946 | j2zhao@uchicago.edu | github.com/j2zhao
5730 S Ellis Ave, Chicago, IL 60637

INTERESTS

Systems for Machine Learning, Data Provenance, Databases

EDUCATION

University of Chicago 2025 (est.)
Computer Science, PhD, Advisor: Sanjay Krishnan

University of Chicago August 2022
Computer Science, Masters

Princeton University, *Summa Cum Laude* June 2019
Computer Science, Bachelor of Science in Engineering
Statistics and Machine Learning, Minor

SELECTED PROJECTS

Behavior Based Semantic Capture for Data Tables. Annotating Python and Jupyter Notebook execution to automatically store code changes when the program is ran. Organizing experiments with University of Chicago undergraduate classes to capture a history of work for data science assignments with this annotation. Using such experiments to analyze behavior component in data analytics.

Fine Grained Provenance Storage and Query for Arrays. Presenting a query and storage framework for fine-grained provenance in data science. Fine-grained provenance is defined as cell-to-cell contribution lineage through a set of array operations. Introducing a new compression algorithm and query optimization techniques that improve storage space and query time by up to 700x and 200x respectively.

Object Recognition Error Classification in Complex Environments with Triplet Labels. Labeling machine learning object detection errors with crowd-sourcing "odd-one-out" triplets, and presenting practical algorithms for semi-supervised classification of errors. This problem is expressed as a MAX-SAT formulation with triplet labels as constraints. Presenting practical algorithms that minimize compute-time and human labels required for an approximate solution.

Domain-Agnostic Lineage Analysis for Data Containers. Designing a system to parse data pipeline graphs over black-box operations without access to intermediate states.

WORK EXPERIENCE

Research Intern **Linea Labs**
June 2023 - September 2023 San Fransisco, CA
Worked closely with a 6-person startup team to design and implement an initial product for Airflow pipeline reproducibility. Ownership of core lineage and large language model features.

Research Intern **Princeton Plasma Physics Lab**
June 2018 - July 2018 Princeton, NJ

Software Engineering Intern **Facebook (Meta)**
June 2017 - August 2017 Seattle, WA

Facebook University Intern **Facebook (Meta)**
June 2016 - August 2016 Menlo Park, CA

TEACHING EXPERIENCE

CDAC Lab Coordinator — **University of Chicago**
May 2020 - August 2020 — Chicago, IL
Coordinated data science research summer program.
Led machine learning team (12 students).

Teaching Assistant — **CMSC 16100, 21800**
University of Chicago — Autumn 2019, 2020
Honors Introduction to Programming, I.
Data Science For Computer Scientists

Teaching Assistant — **COS 397/497**
Princeton University — Fall 2018
Mobile Computing Design for Assistive Technology.

PAPERS

Jinjin Zhao, Avidgor Gal, Sanjay Krishnan. *Quantifying Variation in Data Science Workflows with Fine-Grained Procedural Logging* [Under Submission].

Jinjin Zhao, Sanjay Krishnan. *Compression and In-Situ Query Processing for Fine-Grained Array Lineage*, ICDE 2024. [Paper].

Jinjin Zhao, Avidgor Gal, Sanjay Krishnan. *Data Makes Better Data Scientists*, HILDA 2023. [Short Paper].

Shinan Liu, Tarun Mangla, Ted Shaowang, **Jinjin Zhao**, Sanjay Krishnan, Nick Feamster. *AMIR: Active Multimodal Interaction Recognition from Video and Network Traffic in Connected Environments*, UbiComp/IMWUT 2023. [Paper].

Siyuan Xia, Zhiru Zhu, Chris Zhu, **Jinjin Zhao**, Kyle Chard, Aaron J. Elmore, Ian Foster, Michael Franklin, Sanjay Krishnan, Raul Castro Fernandez. *Data Station: Delegated, Trustworthy, and Auditable Computation to Enable Data-Sharing Consortia with a Data Escrow.* VLDB 2022. [Paper].

Ted Shaowang*, **Jinjin Zhao***, Stavos Sintos, Sanjay Krishnan. *Towards Causal Query Answering for Debugging Video Analytics Systems.* HILDA 2022. [Paper].

Emi Zeger, Florian Laggner, Alessandro Bortolon, Cristina Rea, Orso Meneghini, Samuli Saarelma, Brian Sammuli, Sterling Smith, **Jinjin Zhao**. *Prediction of DIII-D Pedestal Structure From Externally Controllable Parameters.* IEEE Trans. Plasma Sci. 2021. [Paper].

POSTERS

Jinjin Zhao, Egemen Kolemen, Xiaoyan Li, Florian Laggner. *Experimental Based Pedestal Prediction using Machine Learning.* American Physical Society (APS) Division of Plasma Physics 2018. [Poster].

ACTIVITIES/ AWARDS

2022 University Unrestricted Fellowship, University of Chicago
OSDI '20 Diversity Grant, USENIX Association
2020 - 2022 Curriculum/Social Minister, UChicago CS Graduate Student Ministry
2019 Neubauer Graduate Scholarship, University of Chicago
ChatterWorks, 2016 YHacks 1&1 Prize Winner