Pranav Subramaniam

PhD Candidate, University of Chicago, Chicago, IL
psubramaniam@uchicago.edu — linkedin.com/in/pranavsubramaniam — github.com/primkey7607 — primkey7607.github.io

RESEARCH FOCUSES

AI Agents for Data Governance Workflows, Data Science, and Automated Database Deployment

SELECTED PROJECTS

Intent-Based Access Control for Database Deployments Devised novel access control model for synthesizing and auditing access control implementations from NL policies, Intent-based Access Control for Databases (IBAC-DB). A LLM-backed reference architecture for implementing IBAC-DB access control over database backends, NL2SQL systems, and LLMs. Novel automatic benchmark generation procedures to evaluate any IBAC-DB system. Accurate systems to synthesize and audit access control privileges, temporal constraints, and role hierarchies.

Numeric Related Table Search using LLMs Leveraged numeric distribution properties and knowledge graphs to discover strongly semantic joinable numeric tables to an input table. Devised algorithms connecting LLMs and knowledge graphs for related numeric table search.

LLMs-as-a-Crowd For Entity Resolution Experimented with state-of-the-art entity resolution systems, using novel embedding and fine-tuning strategies for tabular data. Benchmarked novel entity resolution strategies combining various LLM prompting strategies (e.g., chain-of-thought, self-consistency, persona-based prompting) and crowd truth inference (e.g., Dawid-Skene, Majority Vote, EBCC, etc.) including both prompting and fine-tuning, done using both pytorch and OpenAI libraries.

WORK EXPERIENCE

AI Research Intern, IBM Watson-Yorktown, NY June 2022 — September 2022
Synthesized a novel definition of strongly semantic join discovery over numeric data from the join discovery literature. Used SPARQL to automatically generate a benchmark from DBpedia labels and real-world datasets. Built a system, NumJoin, that leverages numeric distributions and knowledge graphs to discover such semantic joins. Devised a novel algorithm leveraging LLMs and knowledge graphs for numeric related table search.

Research Assistant, University of Chicago-Prof. Aaron Elmore 2018 — 2019
Analyzed the query performance benefits of various encodings in a columnar database system employing the Hadoop datafile system (HDFS) with the Parquet file format. Assisted in the implementation of a new type of dictionary encoding (the mostly-order-preserving dictionary, or MOP) in Scala and Spark. Implemented a query generator in java for testing the performance of MOP encoding that generated queries whose results satisfied a specific workload requirement.

Research Assistant, University of Chicago-Prof. Shan Lu 2017 — 2019
Analyzed performance problems in ORM (object-relational mapping)-backed web applications by manually reading and analyzing code related to the issues. Built a java plugin, Powerstation, that can statically analyze ruby-on-rails code for ORM-backed web applications and fix performance problems.

TECHNICAL SKILLS

Languages: Python (5 years experience), Java (3 years), C++ (2 years), SQL (5 years)
Prompting Frameworks: OpenAI (2 years), LangChain (1 year)
Large-Scale Data Analysis: Apache Spark (3 years), Pandas (5 years), Hadoop (2 years)
ML Frameworks: PyTorch (3 years), Tensorflow (1 year)

AWARDS/HONORS

The University of Chicago-NSF Graduate Research Fellow 2019 — 2024
The University of Chicago-Eckhardt Scholarship Recipient 2019 — 2025
NBER Young Scholars Workshop on the Economics of Artificial Intelligence Invitee Fall 2020
The University of Chicago-College Research Fellow 2018 — 2019
The University of Chicago-Stamps Leadership Merit Scholar 2015 — 2019

EDUCATION

University of Chicago, Chicago, IL 2019 — 2025
PhD in Computer Science, Advisor: Sanjay Krishnan
Thesis Committee: Michael Franklin, Grant Ho, Sanjay Krishnan

University of Chicago, Chicago, IL — September 2023
M.S. in Computer Science, — Cumulative GPA: 3.93/4.00
Thesis Title: LLM-as-a-Crowd: How To Use LLMs For Data Integration Tasks

University of Chicago, Chicago, IL — 2015 — 2019
B.S. in Computer Science(with Honors) & Mathematics, — Cumulative GPA: 3.80/4.00

TEACHING EXPERIENCE

The University of Chicago, Advanced Database Systems
CMSC 23530 — Winter 2025
The University of Chicago, Computer Architecture for Scientists
CMSC 22240 — Winter 2024
The University of Chicago, Introduction to Databases
CMSC 23500 — Winter 2023
The University of Chicago, Data Science for Computer Scientists
CMSC 21800 — Fall 2022

PUBLICATIONS

- **Pranav Subramaniam** and Sanjay Krishnan. *Agentic Workflows for Extraction of Access Control Matrices from Policy Documents.* To Appear In 2025 International Conference on Data Engineering Workshop on Data-AI Systems. 2025.
- **Pranav Subramaniam** and Sanjay Krishnan. *Intent-Based Access Control: Using LLMs to Intelligently Manage Access Control.* arXiv preprint arXiv:2402.07332. under review. 2025.
- **Pranav Subramaniam**, Udayan Khurana, Kavitha Srinivas, and Horst Samulowitz. 2023. *NumJoin: Discovering Numeric Joinable Tables with Semantically Related Columns.* In Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM '23).
- **Pranav Subramaniam**, Udayan Khurana, Kavitha Srinivas, and Horst Samulowitz. *Related Table Search for Numeric data using Large Language Models and Enterprise Knowledge Graphs.* In ACM International Conference on Information and Knowledge Management. 2023.
- Zezhou Huang, **Pranav Subramaniam**, Raul Castro Fernandez, and Eugene Wu. *Kitana: Efficient Data Augmentation Search for AutoML.* arXiv preprint arXiv:2305.10419 (2023).
- Javen Kennedy, **Pranav Subramaniam**, Sainyam Galhotra, and Raul Castro Fernandez. 2022. *Revisiting Online Data Markets in 2022: A Seller and Buyer Perspective.* SIGMOD Rec. 51, 3 (September 2022), 30–37.
- **Pranav Subramaniam**, Yintong Ma, Chi Li, Ipsita Mohanty and Raul Castro Fernandez. *Comprehensive and Comprehensible Data Catalogs: The What, Who, Where, When, Why, and How of Metadata Management.* ArXiv abs/2103.07532 (2021): n. pag.
- Raul Castro Fernandez, **Pranav Subramaniam**, and Michael J. Franklin. 2020. *Data market platforms: trading data assets to solve data problems.* Proc. VLDB Endow. 13, 12 (August 2020), 1933–1947.
- **Pranav Subramaniam**. 2019. Generating Selective Filters for Access Method and Physical Design Evaluation. In Proceedings of the 2019 International Conference on Management of Data (SIGMOD '19). Association for Computing Machinery, New York, NY, USA, 1853–1855.
- Chunwei Liu, McKade Umbenhower, Hao Jiang, **Pranav Subramaniam**, Jihong Ma, and Aaron J. Elmore. *Mostly order preserving dictionaries.* In 2019 IEEE 35th International Conference on Data Engineering (ICDE), pp. 1214-1225. IEEE, 2019.
- Junwen Yang, Cong Yan, **Pranav Subramaniam**, Shan Lu, and Alvin Cheung. 2018. *PowerStation: automatically detecting and fixing inefficiencies of database-backed web applications in IDE.* In Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE 2018).
- Junwen Yang, **Pranav Subramaniam**, Shan Lu, Cong Yan, and Alvin Cheung. 2018. *How not to structure your database-backed web applications: a study of performance bugs in the wild.* In Proceedings of the 40th International Conference on Software Engineering (ICSE '18). Association for Computing Machinery, New York, NY, USA, 800–810.
- **Pranav Subramaniam** and Abhishek Parakh, *A Quantum Diffie-Hellman Protocol.* 2014 IEEE 11th International Conference on Mobile Ad Hoc and Sensor Systems, Philadelphia, PA, USA, 2014, pp. 523-524.
- **Pranav Subramaniam** and Abhishek Parakh, *Limits on detecting eavesdropper in QKD protocols.* 2014 IEEE International Conference on Advanced Networks and Telecommuncations Systems (ANTS), New Delhi, India, 2014, pp. 1-3.
- Parvathi Chundi and **Pranav Subramaniam**. *An approach to analyze web privacy policy documents.* In KDD Workshop on Data Mining for Social Good. 2014.