SHAI MOHAN

(503) 801-7847 shaimohan23@gmail.com https://github.com/shaimohan23 https://www.linkedin.com/in/shaimohan

The University of Chicago

March 2025; Chicago, IL

M.S. Computer Science; Specialization: Data Analytics (Student Ambassador)

Relevant Coursework:

Advanced Algorithms, Computer Systems, NLP, ML, Data Analysis, Distributed Systems, Research Practicum, Reinforcement Learning

Merit Scholar 2023 - 2025

Purdue University

May 2019; W. Lafayette, IN

B.S. Computer Science; Concentration: Machine Learning; Minor: Mathematics

Awards/Honors:

Presidential Merit Scholarship	2015 - 2019
Purdue Dean's List	2015, 2016, 2019
EPICS (Engineering Projects in Community Service) Outstanding Team Technical Design	Dec 2017
EPICS Social Entrepreneurship Award	May 2017

Relevant Coursework:

Data Mining & ML, Data Structures & Algorithms, Intro to AI, Discrete Mathematics, Probability, Web Information Retrieval, Relational DB Systems, Computer Architecture

RESEARCH & PROFESSIONAL EXPERIENCE

Globus Labs (Prof Kyle Chard) UChicago, Il

Graduate Research Assistant June 2024 - Current

- Developed a Globus connector tool utilizing the Filesystem in User space (FUSE) interface to enable seamless integration with Globus data transfer systems. Tool allows third party company to transfer large datasets across distributed systems and use in ML queries. Project involved low-level file system management, network protocol integration, and optimizing data flow across various platforms.
- Leading the research and analysis of 3D photon data from the Advanced Photon Source (APS) and Advanced Light Source (ALS), focusing on data preprocessing, transformation, and visualization. Applied ML techniques for handling and interpreting large-scale scientific datasets, collaborating with Argonne National Labs to ensure accurate insights.
- Utilized tunable U-Net models to preprocess and standardize data from APS and ALS, facilitating effective cross-source comparison and analysis. Using HPC to implement federated learning model to enable advanced analysis across both datasets while preserving data privacy and security. Aiming to submit a paper on this work in Winter 2025

Research Assistant (Prof Haryadi Gunawi) UChicago, Il

Research Practicum Program January - March 2024

- Analyzed recent approaches to optimizing file storage efficiency using slow disk latency detection and recreated the results of the lab's recent work on this project
- Experimented with different ML models to see if other approaches could be used in conjunction with current approach and if not, analyze why for related works context

Polaris Wireless Mountain View, CA

Research & Development Software Engineer February 2020 - July 2023

- Designed and implemented back end server for 3D elevation product for emergency location services; coded, tested and delivered full software solution, created original efficient datastore algorithm based on storage requirements, and built custom data visualization models for raw data using python, Grafana, influxDB, and Docker
- Developed and upgraded technical solution for prison geofence product suite to support 5+ million users, which was then used as foundation for new line of company products; built Python servers to connect app to product engine allowing customers to use service and created language parser to maximize testing efficiency
- Led a cross-functional team and collaborated with DataWalk on a data fusion optimization project to design profile generator tool; conducted research and developed POCs for integrating GPT to optimize named entity recognition and sentiment analysis aspects; built and presented live demos to company management and customers
- Conducted code reviews and provided mentorship to new hires and a direct report. Created internal tools to enhance software efficiency, including Python-to-Java converter and location data cleaner and analyzer for behavior classification

Intuit Inc; SWE Intern; Mountain View, CA May - August 2018

- Built internal tools for testing TurboTax and Mint; conducted a technical tool analysis and developed multiple POCs for new QE tools to optimize the number of active users these products can support; created API and datastore Docker images and linked them to a UI within a container.

AVI Networks (A VMware Co.); SWE Intern; Palo Alto, CA May - July 2017

- Wrote python scripts to collate and analyze software logs for comparing the company's product to competing products; improved existing script's efficiency by 90%.

Lindamar Capital, LLC; SWE Intern; Palo Alto, CA May - July 2016

- Led a team of engineers in developing Python scripts utilizing Monte Carlo simulations for stock value and options prediction. Designed a script to extract pertinent data from PDF stock reports for analysis and algorithm development, resulting in a robust prediction model for stock option values

TECHNICAL EXPERIENCE

NLP Research Project June - August 2024

- Conducted research on political bias detection in news articles using machine learning and NLP techniques. Built data gathering, cleaning, and pre-processing pipelines, and trained models using random forest, neural net, BERT, and Llama3.
- Fine-tuned transformer-based models and evaluated their performance against baseline models, demonstrating expertise in hyper parameter optimization, tokenization, and classification pipelines. Employed advanced techniques such as sentiment-based filtering to enhance fine-tuned model's accuracy

Purdue EPICS (Deaf Kids Code); Back-end Developer, Archivist, Project Partner Liaison Jan - Dec 2017

Worked on a backend engineering team of 6 to create an ordering system for Jamaican coffee shop to help bridge communication gap between the hearing impaired staff and customers. Helped develop a hybrid recommendation system combining collaborative filtering and content-based approaches to enhance accuracy of suggested items for ordering.

Other

Built different models such as a fault tolerant Raft REST distributed message queue, PageRank, collaborative filters on meal and movie recommendations, clustering models to optimize cell tower placement, and Bayesian Monte Carlo simulations for analyzing the "anchoring effect"

SKILLS

Programming Languages: Python (NumPy, NLTK, SpaCy, PyTorch, SciKit, Transformers, Beautiful Soup, Fuse, and more), Java, Matlab, SQL

Databases/Tools: MongoDB, ClickHouse, mySQL, influxDB, neo4j, Grafana, Docker, Spring Boot, Autodesk Inventor

Certifications/Courses: Machine Learning w/ Andrew Ng, DataWalk Technical Pro, DataWalk App Center Pro