Mir Mahathir Mohammad - Curriculum Vitae

Contact

Location: Salt Lake City, UT **Email:** mirmahathir1@gmail.com **Portfolio:** https://mirmahathir1.github.io/

LinkedIn GitHub Google Scholar DBLP

Education

University of Utah, Salt Lake City, Utah
Ph.D. in Computer Science
Kahlert School of Computing | August 2023 – Present
Advisor: El Kindi Rezig

University of Utah, Salt Lake City, Utah
M.S. in Computer Science
Kahlert School of Computing | August 2023 – December 2025
CGPA: 3.978 / 4.0
Advisor: El Kindi Rezig

Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh
B.Sc. in Computer Science and Engineering
April 2017 – April 2022
Advisor: Muhammad Abdullah Adnan

Research Interests

I focus on data preparation in general and data discovery in particular for data lakes. I built SemDisc, the first end-to-end semantic join discovery system using a query-by-example interface. I also explored how discovered relationships can recommend meaningful ways to organize and sort data. I built NLDisc, which turns a user's natural language description of intent into join discovery over data lakes, and co-developed LORY-J, which treats spatial attributes as first-class signals to discover location-aware joins. I contributed to Buckaroo, a visual data wrangling system that enables users to interactively clean and repair data anomalies in visualizations. My interests include:

Data Systems • Data Lakes • Data Discovery & Integration • Data Wrangling • Data Cleaning • AI for Data Management

Publications

Publications since starting Ph.D. (2023-Present, ~3.3 years)

- ***SIGMOD'26***: Mir Mahathir Mohammad, El Kindi Rezig. "Qualitative Join Discovery in Data Lakes using Examples." Accepted at ACM SIGMOD 2026. [PDF] [Link] *a system for discovering hybrid join paths (combining semantic and equi-joins) in data lakes using query-by-example, supporting hidden tables and semantic tuple matching*
- ***SIGMOD'26***: Mir Mahathir Mohammad, El Kindi Rezig. "SemDisc: An End-to-End Query-by-Example Semantic Join Discovery System." Accepted at ACM SIGMOD 2026. [PDF] [Link] *an interactive system enabling users to discover hybrid join paths in schema-less data lakes through a query-by-example interface*
- ***SIGMOD'26***: Mir Mahathir Mohammad, El Kindi Rezig. "NLDisc: Intent-aware Join Discovery using Natural Language." Accepted at ACM SIGMOD 2026. [PDF] [Link] *a system for intent-aware join discovery over data lakes using natural language queries with interactive intent disambiguation and join path diversification*
- ***VLDB'26***: Lu Xing, Mir Mahathir Mohammad, Walid G. Aref, El Kindi Rezig. "LORY-J: Location-aware Join Discovery from Data Lakes." Accepted at VLDB 2026. [PDF] [Link] *a system for location-aware join discovery in data lakes that treats spatial attributes as first-class signals, composing hybrid join paths through an interactive, map-driven interface*
- ***CIDR'26***: El Kindi Rezig, Mir Mahathir Mohammad, Nicolas Baret, Ricardo Mayerhofer, Andrew McNutt, Paul Rosen. "Towards Scalable Visual Data Wrangling via Direct Manipulation." Accepted at CIDR 2026. [PDF] [Link] *a visual data wrangling system that enables users to clean and repair data anomalies through direct manipulation of interactive visualizations*
- ***VLDB'25***: Akash Khatri, Mir Mahathir Mohammad, El Kindi Rezig. "Sort it Like You Mean It: Discovering Semantically Interesting Attribute Augmentations to Sort Tables." Accepted at VLDB 2025. [PDF] [Link] *Recommends semantically meaningful ways to sort tables by automatically discovering and augmenting attributes from data lakes using LLMs.*

Undergraduate research

- ***IEEE FG'24***: Iftekhar E Mahbub Zeeon, Mir Mahathir Mohammad, Muhammad Abdullah Adnan. "BTVSL: A Novel Sentence-Level Annotated Dataset for Bangla Sign Language Translation." Accepted at IEEE FG 2024. *Introduces the first large-scale sentence-level dataset for Bangla Sign Language translation.* [PDF] [Link]
- ***Neurocomputing'22***: Md. Ashraful Islam, Mir Mahathir Mohammad, Sarkar Snigdha Sarathi Das, Mohammed Eunus Ali. "A survey on deep learning based Point-of-Interest (POI) recommendations." Accepted at Neurocomputing (Journal), 2022. [PDF] [Link] *Categorizes deep learning approaches for POI recommendation systems in location-based social networks.*

Research Experience

University of Utah, Kahlert School of Computing, Salt Lake City, UT

Graduate Research Assistant, August 2023 – Present

Advisor: El Kindi Rezig (Postdoc, MIT CSAIL; Ph.D., Purdue University)

- Developed algorithms for qualitative join discovery in data lakes using example-based queries
- Built systems for semantic table sorting, natural-language intent-aware join discovery, and location-aware join discovery over data lakes
- Implemented scalable data wrangling systems with direct manipulation interfaces, handling data transformations

Bangladesh University of Engineering and Technology, CSE, Dhaka, Bangladesh

Research Assistant, July 2022 – June 2023

Advisor: Muhammad Abdullah Adnan (Ph.D., University of California San Diego)

- Developed machine learning pipelines for processing and analyzing video datasets for sign language translation
- Built data collection and annotation systems for creating structured datasets, handling data cleaning

Teaching Experience

University of Utah, Kahlert School of Computing, Salt Lake City, UT

Teaching Assistant, Spring 2026

CS 6964: Manage Data for & with Machine Learning. Instructor: El Kindi Rezig

Awards and Honors

- SIGMOD Student Travel Grant, 2026
- Selected as 1 of 4 students to represent the University of Utah for the Google PhD Fellowship Program, 2026
- Fellowship, Kahlert School of Computing, University of Utah, 2023

Technical Skills

LLMs/ML: PyTorch, LangChain, Finetuning Language Models (Flan-T5-base, Qwen2.5-0.5B, Llama3.2-1B)

DevOps: Docker, Kubernetes

SWE: Full Stack Web Development (Vue.js, React.js, Next.js, Express.js), Streamlit, Progressive Web App (PWA) Development for Android

Additional Experience

Everforth, Tokyo, Japan (Remote)

Frontend Developer, April 2022 – June 2023

- Built e-commerce web applications using Vue.js and CakePHP