

## Mir Mahathir Mohammad - Curriculum Vitae

---

### Contact

Location: Salt Lake City, UT    Phone: [+1 385 210 8554](tel:+13852108554)    Email: [mirmahathir1@gmail.com](mailto:mirmahathir1@gmail.com)

Portfolio: <https://mirmahathir1.github.io/>    [LinkedIn](#)    [GitHub](#)    [Google Scholar](#)

---

### Education

University of Utah, Salt Lake City, Utah

Ph.D. in Computer Science

Kahlert School of Computing | August 2023 – Present

CGPA: 3.978 / 4.0 (current)

Advisor: Dr. El Kindi Rezig

Bangladesh University of Engineering and Technology

(BUET), Dhaka, Bangladesh

B.S. in Computer Science

April 2017 – April 2022

Advisor: Dr. 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. Additionally, I contributed to Buckaroo, a visual data wrangling system that enables users to interactively clean and repair data anomalies through direct manipulation of 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, ~2.3 years)

- **SIGMOD'26:** Mir Mahathir Mohammad, El Kindi Rezig. "Qualitative Join Discovery in Data Lakes using Examples." Accepted at ACM SIGMOD International Conference on Management of Data (SIGMOD'26), 2026. [\[PDF\]](#) *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*
- **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. [\[PDF\]](#) *a visual data wrangling system that enables users to clean and repair data anomalies through direct manipulation of interactive visualizations*
- **VLDB'25 (Demo):** 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 (Demo Track). [\[PDF\]](#) *Recommends semantically meaningful ways to sort tables by automatically discovering and augmenting attributes from data lakes using LLMs.*

#### Undergraduate research

- **IEEE FG'24:** Iftekhhar 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, derived from 60 hours of YouTube news content with professional signers.* [\[PDF\]](#) [\[Link\]](#)
  - **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: Dr. El Kindi Rezig

- Developed algorithms for qualitative join discovery in data lakes using example-based queries, enabling efficient

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

Research Assistant, July 2022 – June 2023

Advisor: Dr. Muhammad Abdullah Adnan

- Developed machine learning pipelines for processing and analyzing large-scale video datasets for sign language

- dataset integration across heterogeneous tabular data
- Built systems for semantic attribute augmentation and table sorting, improving data discovery workflows for analysts working with complex datasets
- Implemented scalable data wrangling systems with direct manipulation interfaces, handling data transformations efficiently

- translation
- Built data collection and annotation systems for creating structured datasets, handling data cleaning

---

## Additional Experience

Everforth, Tokyo, Japan (Remote)

**Frontend Developer**, April 2022 – June 2023

- Built scalable web applications using Vue.js and CakePHP

---

## Technical Skills

**Data Systems & Databases:** PostgreSQL, MySQL, MongoDB, Query Optimization, Data Indexing

**Data Processing & ML:** Pandas, NumPy, PyTorch, TensorFlow, Scikit-learn, Data Wrangling, ETL Pipelines

**Cloud & Infrastructure:** Docker, Google Cloud Platform, Azure (familiar)

**Programming Languages:** Python (advanced), JavaScript/TypeScript, C++, Java, SQL

**Development Tools:** Node.js, Express.js, React.js, Vue.js, Git, Streamlit

---

## Selected Projects

- **Badhan Blood Donation Management System:** Designed and implemented a full-stack blood donation platform with MongoDB backend, serving users across BUET campus with real-time donor matching and request management [[GitHub](#)]
- **CNN-Based Object Detection:** Developed deep learning models for real-time object detection using PyTorch [[GitHub](#)]
- **Automated Robotic Arm:** Built computer vision and control systems for robotic manipulation tasks using MATLAB, integrating sensor data processing and motion planning algorithms [[GitHub](#)]
- **Comparative Analysis of AI Agents for Othello:** Compared 12 Othello AI agents, from heuristic search baselines to reinforcement learning approaches, using a round-robin tournament to analyze their relative performance [[GitHub](#)]