ANNA FARIHA

Homepage: afariha.github.io
Email: anna.fariha.bd@gmail.com

RESEARCH INTEREST

The aim of my research is to democratize data and data-driven systems towards boosting productivity and enhancing transparency. To this end, my research focuses on three key aspects of data system democratization: enhancing *usability* of data systems for both non-experts and experts, providing *explanation* frameworks to enable understanding of system behavior, and achieving *trust* and *fairness* in data-driven machine learning.

- **Data systems usability:** query by example, data summarization, data discovery, data cleaning.
- **Metadata management and its applications:** data constraints, data drift, trusted machine learning.
- **Causal reasoning and explanation frameworks:** root cause analysis of nondeterministic bugs, interventional debugging, data debugging, explaining data non-conformance, fair classification.

EDUCATION

- Doctor of Philosophy in Computer Science (2021)
  College of Information and Computer Sciences, University of Massachusetts, Amherst
  Dissertation title: Enhancing Usability and Explainability of Data Systems
    - Advisor: Alexandra Meliou
    - Committee members: Emery Berger (UMass), Peter Haas (UMass), Suman Nath (MSR)
    - CGPA: 4.00 (in the scale of 4.00)
- Master of Science in Computer Science (2020)
  College of Information and Computer Sciences, University of Massachusetts, Amherst
    - CGPA: 4.00 (in the scale of 4.00)
- Master of Science in Computer Science and Engineering (2011)
  University of Dhaka, Bangladesh
    - CGPA: 4.00 (in the scale of 4.00)
- Bachelor of Science in Computer Science and Engineering (2010)
  University of Dhaka, Bangladesh
    - CGPA: 3.83 (in the scale of 4.00)

EMPLOYMENT HISTORY

- **Researcher** (June 2021 – present)
  PROSE (Programming by Examples and Natural Language) Research & Engineering Group
  Microsoft, Redmond, WA
- **Research Assistant** (September 2016 – May 2021)
  Data Systems Research for Exploration, Analytics, and Modeling (DREAM) Lab
  University of Massachusetts Amherst, MA
- **Research Intern** (Summer 2019 & 2018)
    - PROSE Research & Engineering, Microsoft (Mentor: Ashish Tiwari)
    - DMX Lab, Microsoft Research, Redmond (Mentor: Suman Nath)
- **Teaching Faculty (Lecturer)** (2012 – 2016)
    - Computer Science and Engineering, University of Dhaka, Bangladesh (2014 – 2016)
    - Computer Science and Engineering, United International University, Bangladesh (2012 – 2014)

HONORS, AWARDS, & GRANTS

- VLDB Best Demonstration Runner-up award, 2020
- Microsoft Research Dissertation Grant, 2020 ($25,000 USD for academic year 2020 – 2021)
- Travel and accommodation award in CRA-W Grad Cohort Workshop, Washington DC, 2017
- Jim Gray scholarship, CICS, UMass Amherst, 2016
- Special Achievement Award – Bangladesh ICT Journalist Forum & Bangladesh Women in IT, 2012
- Special Achievement Award for outstanding performances in programming contests, U Dhaka, 2011
- Dean's Honor Award and Dean's Award – Faculty of Engineering, University of Dhaka (top 1%), 2010
- Fatema-Iqbal trust scholarship, University of Dhaka, 2008
- Best Student of the School (Kushtia Govt. Girls' High School), 2003
- National Education Board Scholarship: 12th grade, 10th grade, 8th grade, & 5th grade

PUBLICATIONS

1. A Fariha, L Cousins, N Mahyar, A Meliou: (under revision) <u>ACM Trans. on Database Systems</u> 2022
2. R Bavishi, J Cambronero, A Fariha, S Gulwani, V Le, I Radiček, A Tiwari: (under review) <u>PLDI</u> 2022
3. E Rezig, A Bhandari, A Fariha, B Price, A Vanterpool, A Bowne, L McEvoy, V Gadepally: Examples are All You Need: Iterative Data Discovery by Example in Data Lakes. (abstract) <u>CIDR</u> 2022.
4. S Galhotra, A Fariha, R Lourenço, J Freire, A Meliou, D Srivastava: DataPrism: Exposing Disconnect between Data and Systems. <u>SIGMOD</u> 2022.
5. MT Islam, A Fariha, A Meliou, B Salimi: Through the Data Management Lens: Experimental Analysis and Evaluation of Fair Classification. <u>SIGMOD</u> 2022.
6. N Yadav, M Brucato, A Fariha, O Youngquist, J Killingback, A Meliou, PJ Haas: SubSumE: A Dataset for Subjective Summary Extraction from Wikipedia Documents. <u>NewSum@EMNLP</u> 2021.
7. E Rezig, A Bhandari, A Fariha, B Price, A Vanterpool, V Gadepally, M Stonebraker: DICE: Data Discovery by Example. (demo paper) <u>VLDB</u> 2021.
8. A Fariha, A Tiwari, A Meliou, A Radhakrishna, S Gulwani. CoCo: Interactive Exploration of Conformance Constraints for Data Understanding and Data Cleaning. (demo paper) <u>SIGMOD</u> 2021.
9. A Fariha, A Tiwari, A Radhakrishna, S Gulwani, A Meliou. Conformance Constraint Discovery: Measuring Trust in Data-Driven Systems. <u>SIGMOD</u> 2021.
10. A Fariha, S Nath, A Meliou. Causality-guided Adaptive Interventional Debugging. <u>SIGMOD</u> 2020
11. A Fariha, M Brucato, PJ Haas, A Meliou. SuDocu: Summarizing Documents by Example. (demo paper) <u>VLDB</u> 2020 (Best demonstration runner-up award).
12. A Fariha, A Tiwari, A Radhakrishna, S Gulwani. ExTuNe: Explaining Tuple Non-conformance. (demo paper) <u>SIGMOD</u> 2020.
13. A Fariha, A Meliou. Example-Driven Query Intent Discovery: Abductive Reasoning using Semantic Similarity. <u>VLDB</u> 2019.
14. A Fariha, SM Sarwar, A Meliou. SQuID: Semantic Similarity-Aware Query Intent Discovery. (demo paper) <u>SIGMOD</u> 2018.
15. A Fariha, CF Ahmed, CK Leung, M Samiullah, S Pervin, L Cao. A New Framework for Mining Frequent Interaction Patterns from Meeting Databases. <u>Engineering Applications of Artificial Intelligence</u> 2015.
16. A Fariha, CF Ahmed, CK Leung, SM Abdullah, L Cao: Mining Frequent Patterns from Human Interactions in Meetings Using Directed Acyclic Graphs. <u>PAKDD</u> 2013.

SERVICE & OUTREACH

- **Round Table Chair:** Interactive Querying and Visualization for Large Data, VLDB 2021
- **Session Chair:** Learning and Explainability, EDBT 2021
- **Program committee member:**
  - SIGMOD 2022 & 2021 (research track)
  - VLDB 2022 (research track) & 2021 (demonstration track)
  - EDBT 2022 & 2021 (demonstration track)
- **External reviewer**: SIGMOD 2018, 2019, and 2020; CHI 2021; UIST 2021; The VLDB Journal
- **Mentor**: Research Experience for Undergraduate Students (REU), UMass, Summer 2017
- **Volunteer**: Girls Inc. Eureka! Workshop, Summer 2017; WECode, Harvard University, 2017

INVITED TALKS

- **Univ. of Washington, Northwest Database Society, Microsoft Database Lecture Series:** Nov, 2021
- **MIT CSAIL, Semantic data management reading group:** Oct, 2021
- **CIS, University of Pennsylvania:** Feb, 2021
- **DATA Lab, Northeastern University:** Jan, 2021

COMPETITIVE PROGRAMMING & PROBLEM SOLVING

Event	Position	Team	Year
ACM ICPC (Dhaka Regional, Bangladesh)	5	DU Resonance	2011
IUT ICT Fest Programming Contest (Dhaka, Bangladesh)	1	DU Resonance	2011
DU NCPC (Dhaka, Bangladesh)	1	DU Resonance	2011
DU NCPC – Women (Dhaka, Bangladesh)	1	DU Felicity	2011
ACM ICPC (Amritapuri Regional, India)	6	DU Resonance	2010
ACM ICPC (Dhaka Regional, Bangladesh)	4	DU Resonance	2010
ACM ICPC (Dhaka Regional, Bangladesh)	3	DU Ouranos	2009

- **Problem Setter & Judge**
  - SAARC Programming Contest, 2014 & 2013, Bangladesh University
  - Bangladesh Informatics Olympiad, 2013, National Round
  - ACM ICPC – Dhaka Site, 2012, Daffodil International University