A N N A F A R I H A

Homepage: afariha.github.io
Email: AnnaFariha@microsoft.com

RESEARCH INTEREST

The aim of my research is to democratize data and data systems towards boosting productivity and agility. 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, SQL program synthesis, data summarization.
- **Metadata management and its applications:** data constraints, data drift, trusted machine learning.
- **Debugging, causal reasoning, and explanation:** 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
    o   Advisor: Alexandra Meliou
    o   Committee members: Emery Berger (UMass), Peter Haas (UMass), Suman Nath (MSR)
- Master of Science in Computer Science (2020)
  College of Information and Computer Sciences, University of Massachusetts, Amherst
    o   CGPA: 4.00 (in the scale of 4.00)
- Master of Science in Computer Science and Engineering (2011)
  University of Dhaka, Bangladesh
    o   CGPA : 4.00 (in the scale of 4.00)
- Bachelor of Science in Computer Science and Engineering (2010)
  University of Dhaka, Bangladesh
    o   CGPA: 3.83 (in the scale of 4.00)

EMPLOYMENT HISTORY

- **Researcher** (June 2021 – present)
  PROSE, Microsoft
- **Research Assistant** (September 2016 – May 2021)
  Data Systems Research for Exploration, Analytics, and Modeling (DREAM) Lab, CICS, UMass
- **Research Intern** (Summer 2019 & 2018)
  -   PROSE Research & Engineering, Microsoft (Mentor: Ashish Tiwari)
  -   DMX Lab, Microsoft Research, Redmond (Mentor: Suman Nath)
- **Lecturer** (2012 – 2016)
  -   Computer Science and Engineering, University of Dhaka, Bangladesh (2014 – 2016)
  -   Computer Science and Engineering, United International University, Bangladesh (2012 – 2014)
- **Software Engineer** (2011 – 2012)
  Structured Data Systems Limited, Dhaka

HONORS, AWARDS, & GRANTS

- VLDB Best Demonstration Runner-up award, 2020
- Microsoft Research Dissertation Grant, 2020
- 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, 2012
- Special Achievement Award – Bangladesh Women in IT, 2012
- Special Achievement Award for outstanding performances in programming contests, 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
- National Education Board Scholarship: 12th grade (2006) & 10th grade (2004)
- Best Student of the School (Kushtia Govt. Girls' High School), 2003
- Talent Pool Scholarship in National Scholarship Examination: 8th grade (2001) & 5th grade (1998)

PUBLICATIONS

1. E Rezig, A Bhandari, A Fariha, B Price, A Vanterpool, V Gadepally, M Stonebraker: DICE: Data Discovery by Example. (demonstration paper) <u>VLDB</u> 2021.
2. A Fariha, A Tiwari, A Meliou, A Radhakrishna, S Gulwani. CoCo: Interactive Exploration of Conformance Constraints for Data Understanding and Data Cleaning. (demonstration paper) <u>SIGMOD</u> 2021.
3. A Fariha, A Tiwari, A Radhakrishna, S Gulwani, A Meliou. Conformance Constraint Discovery: Measuring Trust in Data-Driven Systems. <u>SIGMOD</u> 2021.
4. A Fariha, S Nath, A Meliou. Causality-guided Adaptive Interventional Debugging. <u>SIGMOD</u> 2020.
5. A Fariha, M Brucato, PJ Haas, A Meliou. SuDocu: Summarizing Documents by Example. (demonstration paper) <u>VLDB</u> 2020 (Best demonstration runner-up award).
6. A Fariha, A Tiwari, A Radhakrishna, S Gulwani. ExTuNe: Explaining Tuple Non-conformance. (demonstration paper) <u>SIGMOD</u> 2020.
7. A. Fariha, A. Meliou. Example-Driven Query Intent Discovery: Abductive Reasoning using Semantic Similarity. <u>VLDB</u> 2019.
8. A Fariha, SM Sarwar, A Meliou. SQuID: Semantic Similarity-Aware Query Intent Discovery. (demonstration paper) <u>SIGMOD</u> 2018.
9. 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.
10. 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

- **Program committee member:**
    - VLDB 2022 (research track)
    - SIGMOD 2021 (research track)
    - EDBT 2021 (demonstration track)
    - VLDB 2021 (demonstration track)
- **External reviewer**: SIGMOD 2018, 2019, and 2020; CHI 2021; UIST 2021; VLDB Journal;
- **Mentor**: Research Experience for Undergraduate Students (REU), UMass, Summer 2017
- **Volunteer**: Girls Inc. Eureka! Workshop, Summer 2017

INVITED TALKS

- **CIS, University of Pennsylvania:** February, 2021
- **DATA Lab, Northeastern University:** January, 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
Google Code Jam	515	Individual	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

TECHNICAL SKILLS

- Python, LaTeX, Java, SQL, C, C++, PHP, Javascript, R