

Anna Fariha

Kahlert School of Computing, University of Utah
Warnock Engineering Building, #2851
72 Central Campus Drive, Salt Lake City, UT, 84112

afariha@cs.utah.edu
afariha.github.io
+1 801-581-3037

Research Interests

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, discovery, wrangling, evolution.
- *Metadata management and its applications*: data constraints, data drift, data cleaning.
- *Systems for machine learning and AI*: Large language models for data management, trusted machine learning, fair machine learning, explainable AI, personalized text summarization.
- *Causal reasoning and explanation frameworks*: software debugging, automatic program repair, data debugging, explaining data non-conformance, causal insights.

Education

Doctor of Philosophy in Computer Science 2021
Manning College of Information and Computer Sciences *University of Massachusetts, Amherst, MA*

- Dissertation: Enhancing Usability and Explainability of Data Systems
- Advisor: Alexandra Meliou
- Committee members: Emery Berger, Peter Haas, Suman Nath
- CGPA: 4.00 out of 4.00

Master of Science in Computer Science 2020
Manning College of Information and Computer Sciences *University of Massachusetts, Amherst, MA*

- CGPA: 4.00 out of 4.00

Master of Science in Computer Science and Engineering 2011
Department of Computer Science and Engineering *University of Dhaka, Bangladesh*

- Thesis: A New Approach for Frequent Human Interaction Pattern Mining in Meeting Databases
- Advisor: Suraiya Pervin
- CGPA: 4.00 out of 4.00

Bachelor of Science in Computer Science and Engineering 2010
Department of Computer Science and Engineering *University of Dhaka, Bangladesh*

- Thesis: An Algorithm for Graph Mining using Efficient Graph Indexing for Evolving Database
- Advisor: Chowdhury Farhan Ahmed
- CGPA: 3.83 out of 4.00

Employment History

University of Utah 7/2023 – present
Assistant Professor
Kahlert School of Computing
Salt Lake City, UT, USA

Microsoft 6/2021 – 5/2023
Researcher
PROSE research and engineering team for AI-assisted programming
Managers: Gustavo Soares, Sumit Gulwani
Redmond, WA, USA

Microsoft <i>Research Intern</i> PROSE research and engineering team for AI-assisted programming Mentor: Ashish Tiwari Project: Conformance constraints and trusted machine learning	5/2019 – 11/2019 Bellevue, WA, USA
Microsoft Research <i>Research Intern</i> Data Management, Exploration and Mining (DMX) Mentor: Suman Nath Project: Adaptive interventional debugging for finding root causes of nondeterministic software bugs	5/2018 – 8/2018 Redmond, WA, USA
University of Massachusetts <i>Research Assistant</i> Data Systems Research for Exploration, Analytics, and Modeling (DREAM) Lab Manning College of Information and Computer Sciences Supervisor: Alexandra Meliou	9/2016 – 5/2021 Amherst, MA, USA
Lalmatia Women's College <i>Guest Lecturer</i> Department of Business Administration	1/2015 – 4/2015 Dhaka, Bangladesh
University of Dhaka <i>Lecturer</i> Department of Computer Science and Engineering	5/2014 – 8/2016 Dhaka, Bangladesh
United International University <i>Lecturer</i> Department of Computer Science and Engineering	5/2012 – 4/2014 Dhaka, Bangladesh
University of Liberal Arts <i>Competitive Programming Trainer</i> Coach for competitive programming in preparation for ACM ICPC Programming Contests	10/2011 – 1/2012 Dhaka, Bangladesh
Structured Data Systems Limited <i>Software Engineer</i> Android and Blackberry mobile applications	9/2011 – 5/2012 Dhaka, Bangladesh
University of Asia Pacific <i>Competitive Programming Trainer</i> Coach for competitive programming in preparation for ACM ICPC Programming Contests	10/2010 – 10/2010 Dhaka, Bangladesh

Honors, Awards, & Achievements

SIGMOD Distinguished PC member	2024
SIGMOD Distinguished PC member	2023
SIGMOD Distinguished PC member	2022
SIGMOD Comprehensive Reproducibility Award	2022
VLDB Best Demonstration Runner-up Award	2020
Jim Gray Scholarship, CICS, UMass Amherst	2016
Special Achievement Award: Bangladesh ICT Journalist Forum	2012
Special Achievement Award: Bangladesh Bangladesh Women in IT	2012
University Scholarship for result in Bachelors of Science, U of Dhaka	2012
Special Achievement Award for outstanding performances in programming contests, U of Dhaka	2011
Dean's Honor Award (top 1%), Faculty of Engineering, U of Dhaka	2011
Dean's Award, Faculty of Engineering, U of Dhaka	2010
Fatema-Iqbal Trust Scholarship, U of Dhaka	2008

Travel and Accommodation Scholarships

Invited Participant, CRA Career Mentoring Workshop, Washington, DC	2024
Invited Participant with Accommodation Support, Logic and Algebra for Query Evaluation – Simons Institute for the Theory of Computing, Berkeley, California	2023
Invited Participant with Travel & Accommodation Support, CRA-WP Career Mentoring Workshop – Chicago, Illinois	2023
CRA-W Grad Cohort Workshop Travel and Accommodation Award, Washington DC	2017

Programming Contests

Rank 5, ACM International Collegiate Programming Contest, Dhaka (Team: DU Resonance)	2011
Rank 1, Islamic University of Technology ICT Fest Programming Contest (Team: DU Resonance)	2011
Rank 1, Dhaka University National Collegiate Programming Contest (Team: DU Resonance)	2011
Rank 1, Dhaka University National Collegiate Programming Contest [Women] (Team: DU Felicity)	2011
Rank 2, SUST National Collegiate Programming Contest (Team: DU Resonance)	2010
Rank 6, ACM International Collegiate Programming Contest, Amritapuri (Team: DU Resonance)	2010
Rank 4, ACM International Collegiate Programming Contest, Dhaka (Team: DU Resonance)	2010
Rank 3, ACM International Collegiate Programming Contest, Dhaka (Team: DU Ouranos)	2009
Rank 515, Google Code Jam, (Individual, id: chorui12)	2009
Rank 4, AUST National Collegiate Programming Contest (Team: DU Resonance)	2009

Grants

1. **One Utah Data Science Hub Seed Grant Award:** *Explaining Data Evolution* [PI]. Total award amount: \$50,000. Award duration: Mar 2024–Feb 2025.
2. **CIRC: ENS/Grand. NSF CNS 2346555:** *POWDER-ENS - Enhancing and Sustaining the POWDER Platform* [Co-PI, PI: Jacobus Van Der Merwe, other co-PIs: Aditya Bhaskara, Eric Eide]. Total award amount: \$4,999,980. Award duration: Oct 2024–Sep 2028.
3. **CRA-WP Distributed Research Experiences for Undergraduates (DREU):** *Summer support for two students.* Total award amount: \$14,000. Award duration: Summer 2024.
4. **USHE: Computer Science Targeted Workforce Application:** *Training the Next Generation of Database Engineers: Responsible, User-Centric, and Efficient Data Management* [led by Jeff Phillips]. Teaching grant for developing a new course: *Human-centered Data Management*. Academic year: 2023–2024.
5. **Microsoft Research Dissertation Grant.** Total award amount: \$25,000. Academic year 2020–2021.

Publications

Journal and Conference Publications

- [1] Zifan Liu, Shaleen Deep, Anna Fariha, Fotis Psallidas, Ashish Tiwari, and Avrilia Floratou. Rapidash: Efficient Detection of Constraint Violations. *Proc. VLDB Endow.*, 17(8):2009–2021, 2024
- [2] Anjali Singh, Anna Fariha, Christopher Brooks, Gustavo Soares, Austin Z. Henley, Ashish Tiwari, Chethan M, Heeryung Choi, and Sumit Gulwani. Investigating Student Mistakes in Introductory Data Science Programming. In *SIGCSE*, pages 1258–1264. ACM, 2024
- [3] Rohan Bavishi, Harshit Joshi, José Cambronero, Anna Fariha, Sumit Gulwani, Vu Le, Ivan Radicek, and Ashish Tiwari. Neurosymbolic Repair for Low-code Formula Languages. *Proc. ACM Program. Lang.*, 6(OOPSLA2):1093–1122, 2022
- [4] Sainyam Galhotra, Anna Fariha, Raoni Lourenço, Juliana Freire, Alexandra Meliou, and Divesh Srivastava. DataPrism: Exposing Disconnect between Data and Systems. In *SIGMOD*, pages 217–231. ACM, 2022
- [5] Maliha Tashfia Islam, Anna Fariha, Alexandra Meliou, and Babak Salimi. Through the Data Management Lens: Experimental Analysis and Evaluation of Fair Classification. In *SIGMOD*, pages 232–246. ACM, 2022

- [6] Anna Fariha, Ashish Tiwari, Arjun Radhakrishna, Sumit Gulwani, and Alexandra Meliou. Conformance Constraint Discovery: Measuring Trust in Data-Driven Systems. In *SIGMOD*, pages 499–512. ACM, 2021 ([Comprehensive Reproducibility Award](#))
- [7] Anna Fariha, Suman Nath, and Alexandra Meliou. Causality-Guided Adaptive Interventional Debugging. In *SIGMOD*, pages 431–446. ACM, 2020
- [8] Anna Fariha and Alexandra Meliou. Example-Driven Query Intent Discovery: Abductive Reasoning using Semantic Similarity. *Proc. VLDB Endow.*, 12(11):1262–1275, 2019
- [9] Anna Fariha, Chowdhury Farhan Ahmed, Carson K. Leung, Md. Samiullah, Suraiya Pervin, and Longbing Cao. A New Framework for Mining Frequent Interaction Patterns from Meeting Databases. *Eng. Appl. Artif. Intell.*, 45:103–118, 2015
- [10] Amit Mandal, Mehedi Hasan, Anna Fariha, and Chowdhury Farhan Ahmed. GSCS - Graph Stream Classification with Side Information. In *APWeb*, pages 389–400, 2015
- [11] Md. Samiullah, Chowdhury Farhan Ahmed, Anna Fariha, Md. Rafiqul Islam, and Nicolas Lachiche. Mining frequent correlated graphs with a new measure. *Expert Syst. Appl.*, 41(4):1847–1863, 2014
- [12] Shafaet Ashraf, Sheikh Muhammad Sarwar, Md. Abeed Hassan, Saifuddin Md. Tareeq, and Anna Fariha. An Efficient Method for Extracting Subtrees against Forest Query. In *IMCOM*, pages 98:1–98:7. ACM, 2015
- [13] Anna Fariha, Chowdhury Farhan Ahmed, Carson Kai-Sang Leung, S. M. Abdullah, and Longbing Cao. Mining Frequent Patterns from Human Interactions in Meetings Using Directed Acyclic Graphs. In *PAKDD*, pages 38–49, 2013
- [14] Md. Samiullah, Chowdhury Farhan Ahmed, Manziba Akanda Nishi, Anna Fariha, S. M. Abdullah, and Md. Rafiqul Islam. Correlation Mining in Graph Databases with a New Measure. In *APWeb*, pages 88–95, 2013
- [15] Shariful Islam, Anna Fariha, Chowdhury Farhan Ahmed, and Byeong-Soo Jeong. EGDIM: Evolving Graph Database Indexing Method. In *ICUIMC*, pages 56:1–56:10, 2012

Short, Demonstration, and Workshop Papers

- [16] [Whanhee Cho](#) and Anna Fariha. UTOPIA: Automatic Pivot Table Assistant. *Proc. VLDB Endow.*, 2024
- [17] Bhavya Chopra, Anna Fariha, Sumit Gulwani, Austin Z. Henley, Daniel Perelman, Mohammad Raza, Sherry Shi, Danny Simmons, and Ashish Tiwari. CoWrangler: Recommender System for Data-Wrangling Scripts. In *SIGMOD*. ACM, 2023
- [18] El Kindi Rezig, Anshul Bhandari, Anna Fariha, Benjamin Price, Allan Vanterpool, Andrew Bowne, Lindsey McEvoy, and Vijay Gadepally. Examples are All You Need: Iterative Data Discovery by Example in Data Lakes. In *CIDR*, 2022
- [19] El Kindi Rezig, Anshul Bhandari, Anna Fariha, Benjamin Price, Allan Vanterpool, Vijay Gadepally, and Michael Stonebraker. DICE: Data Discovery by Example. *Proc. VLDB Endow.*, 14(12):2819–2822, 2021
- [20] Nishant Yadav, Matteo Brucato, Anna Fariha, Oscar Youngquist, Julian Killingback, Alexandra Meliou, and Peter Haas. SUBSUME: A dataset for subjective summary extraction from Wikipedia documents. pages 131–141, November 2021
- [21] Anna Fariha, Ashish Tiwari, Alexandra Meliou, Arjun Radhakrishna, and Sumit Gulwani. CoCo: Interactive Exploration of Conformance Constraints for Data Understanding and Data Cleaning. In *SIGMOD*, pages 2706–2710. ACM, 2021
- [22] Anna Fariha, Ashish Tiwari, Arjun Radhakrishna, and Sumit Gulwani. ExTuNe: Explaining Tuple Non-conformance. In *SIGMOD*, pages 2741–2744. ACM, 2020
- [23] Anna Fariha, Matteo Brucato, Peter J. Haas, and Alexandra Meliou. SuDocu: Summarizing Documents by Example. *Proc. VLDB Endow.*, 13(12):2861–2864, 2020 ([Best Demonstration Runner Up Award](#))

- [24] Anna Fariha, Sheikh Muhammad Sarwar, and Alexandra Meliou. SQuID: Semantic Similarity-Aware Query Intent Discovery. In *SIGMOD*, pages 1745–1748. ACM, 2018

Technical Reports

- [25] Yuqing Wang and Anna Fariha. Development of Data Evaluation Benchmark for Data Wrangling Recommendation System (Project report for CRA DREU), 2024
- [26] Rania Saber and Anna Fariha. Formative Study for AI-assisted Data Visualization (Project report for CRA DREU), 2024

PhD Thesis

- [27] Anna Fariha. *Enhancing Usability and Explainability of Data Systems*. PhD thesis, University of Massachusetts Amherst. https://scholarworks.umass.edu/dissertations_2/2311/

Software and Educational Content

- UTOPIA: Automatic Pivot Table Assistant. [Link](#)
- SQuID: Semantic similarity-aware Query Intent Discovery. [Link](#)
- Conformance Constraints Discovery: Measuring Trust in Data-Driven Systems (Winner of the ACM SIGMOD 2022 Most Reproducible Paper Award). [Link](#)
- SubSumE Dataset for Subjective Document Summarization. [Link](#)
- A Data Management Perspective on Fair Classification: An Experimental Analysis and Evaluation. [Link](#)
- Educational video on “Research in Computer Science (Databases/Data Management Track)”. [Link](#)

Panels and Invited Talks

- Panelist – Northwest Database Society Annual Meeting: Panel on AI and Data Management February, 2024
- Panelist – NSF Information and Intelligent Systems division (IIS) 2023
- Panelist – NSF’s CSGrad4US Panel: “What I wish I knew before I started graduate school” October, 2023
- Panelist – SIGMOD New Researcher Symposium June, 2022
- Round Table Chair – VLDB: Interactive Querying and Visualization for Large Data August, 2021
- *Explaining Data Evolution*
 - DELPHI Data Science Symposium, University of Utah November, 2024 (Upcoming)
- *Recent Trends in Scalable and Human-centric Data Systems*
 - Research Challenges in Computing Seminar Series, University of Utah October, 2023
- *Through the Data Management Lens: Experimental Analysis and Evaluation of Fair Classification*
 - Bias, Ethics and Trustworthiness in Data Science and AI event, University of Utah September, 2023
- *Blame the data, not the system: how data constraints can help in trustworthy machine learning and explain causes of data-system malfunction*
 - Data Science Lecture Series, University of Utah September, 2023
 - University of Washington, Northwest Database Society November, 2021
- *Blame the data, not the system: How data profiles can help explain causes of data-system malfunction*
 - Northwest Database Society (NWDS) Annual Meeting May, 2023
- *Enhancing usability and explainability of data systems*
 - Database Seminar, University of California San Diego (virtual) April, 2023
 - Cornell University April, 2022
 - Georgia University of Technology April, 2022
 - University of California, Santa Barbara March, 2022
 - University of Minnesota (virtual) March, 2022
 - University of Waterloo (virtual) March, 2022
 - University of Utah February, 2022
 - University of Toronto (virtual) February, 2022
 - Pennsylvania State University (virtual) February, 2022
 - EPIC meeting, University of California, Berkeley (virtual) February, 2022
 - Ohio State University (virtual) January, 2022

– CSAIL, Massachusetts Institute of Technology (virtual)	October, 2021
– Microsoft (virtual)	January, 2021
– Megagon Labs (virtual)	February, 2021
– University of Pennsylvania (virtual)	February, 2021
– DATA Lab, Northeastern University (virtual)	January, 2021

Teaching

• University of Utah

– CS 3960/6959 - Human-Centered Data Management (undergraduate + graduate)	Spring 2025 (Upcoming)
– CS 5353/6353 - Deep Learning (undergraduate + graduate)	Fall 2024
– CS 3960/6959 - Human-Centered Data Management (undergraduate + graduate)	Spring 2024
– CS 5353/6353 - Deep Learning (undergraduate + graduate)	Fall 2023

• University of Dhaka, Bangladesh

– CSE 1201 - Fundamentals of Programming (undergraduate)	Spring 2014
– CSE 1211 - Fundamentals of Programming Lab (undergraduate)	Spring 2014
– CSE 1102 - Discrete Mathematics (undergraduate)	Fall 2014
– CSE 2101 - Data Structures and Algorithms (undergraduate)	Spring 2015
– CSE 2111 - Data Structures and Algorithms Lab (undergraduate)	Spring 2015
– CSE 3113 - Microprocessor and Assembly Language Lab (undergraduate)	Fall 2015
– CSE 1111 - Fundamentals of Computer and Computing Lab (undergraduate)	Fall 2015
– CSE 2112 - Object Oriented Programming Lab (undergraduate)	Spring 2016
– CSE 2212 - Design and Analysis of Algorithms Lab (undergraduate)	Spring 2016
– CSE 4111 - Artificial Intelligence Lab (undergraduate)	Spring 2016

• United International University, Bangladesh

- CSI 219 - Discrete Mathematics (undergraduate)
- CSI 121 - Structured Programming Language (undergraduate)
- CSI 122 - Structured Programming Language Laboratory (undergraduate)
- CSI 228 - Algorithms Laboratory (undergraduate)
- CSI 233 - Theory of Computing (undergraduate)
- CSI 341 - Artificial Intelligence (undergraduate)
- CSI 342 - Artificial Intelligence Laboratory (undergraduate)

Student Supervision

PhD Advisor

• Whanhee Cho	Fall 2023 – Present
• Chetan Bajaj	Spring 2025 – Present
• Shiyi He	Fall 2023 – Fall 2024

Masters & Undergraduate Advisor

• University of Utah

– Anirudh Kamath (MS)	Fall 2024 – Present
– Vijaysurya Vempati (MS)	Fall 2024 – Present
– Shreya Raj, UMass Amherst	Summer 2024 – Present
– Kuangfei Long, Boston University	Summer 2024 – Present
– Tal Blau, Technion - Israel Institute of Technology	Spring 2024 – Present
– William Erignac	Summer 2024 – Fall 2024
– Ziming Wang	Fall 2024
– Holden Ellsworth (MS)	Summer 2024

- Bhavya Chopra, University of California, Berkeley Summer 2024
- Yuqing Wang, University of Wisconsin-Madison (DREU by CRA, Project Report [25]) Summer 2024
- Rania Saber, University of California-Riverside (DREU by CRA, Project Report [26]) Summer 2024
- Qiaoyi Cai Fall 2023
- University of Dhaka
 - Quazi Marufur Rahman (MS) 2015
 - Amit Mandal 2015
 - Mehedi Hasan 2015
 - Nitish Ranjan Bhowmik 2015
 - Kazi Mazbah Uddin 2015

Mentorship

- Microsoft
 - Bhavya Chopra (Predoctoral Research Fellow) [PhD student at UC Berkeley] 2022–2023
 - Yang Shi (Research Intern) [Assistant Professor at Utah State University] 2022
 - Chethan Mahadevaswamy (Predoctoral Research Fellow) [Data Scientist at Microsoft] 2021–2022
- University of Massachusetts Amherst
 - Maliha Islam (PhD) [Now at Microsoft] 2020–2021
 - Hasnain Heickal (PhD) 2021–2022
 - Oscar Youngquist (MS) 2021–2022
 - Julian Killingback (MS) 2021
 - Kanchi Masalia (MS) 2020
 - Lucy Cousins (MS) 2019–2021
 - Genglin Liu (Undergrad) 2020
 - Armand Asnani (Undergrad) 2019–2020
 - Nischal Dave (Undergrad) 2019
 - Zoey (Jingyi) Sun (Undergrad from Smith College) 2018
 - Lauren Beryl Larson (REU) (Undergrad from Wellesley College) 2017
 - Michael Satanovsky (High-school Intern from Hopkins School) 2017

Professional Services

Internal Service

- Member, Faculty Recruiting Committee, KSoC 2025
- Member, Research Support/Opportunities Committee, KSoC 2024
- Panelist, Graduate Student Mentoring Panel on “How to get a job in industry?”, KSoC 2024
- Founder and Advisor, Grad-CS Women, KSoC (gradcswomen-utah.github.io) 2024–present
- Graduate Admission Committee, KSoC 2024 & 2025
- Faculty Advisor, Women in Computing (wic.utahclubs.org) 2023–present
- Mentor of 1st Year PhD Student (Ishrat Jahan Eliza), KSoC 2023
- Co-founder of UtahDB Research Center (mod.cs.utah.edu) 2022–present

External Service

Program Committee and Workshop Organization

- Workshop Chair, Data-AI Systems (DAIS), co-located with ICDE (dais-workshop-icde.github.io) 2025
- IEEE International Conference on Data Engineering (ICDE) (research track) 2025
- International Conference on Very Large Databases (VLDB) (research track, rapid response reviewer) 2025
- ACM SIGMOD International Conference on Management of Data (demonstration track) 2025
- ACM SIGMOD International Conference on Management of Data (research track) 2025

- GUIDE-AI: ACM SIGMOD Workshop on Governance, Understanding, and Integration of Data for Effective and Responsible AI 2024
- IEEE International Conference on Data Engineering (ICDE) (demonstration track) 2024
- ACM SIGMOD International Conference on Management of Data (demonstration track) 2024
- ACM SIGMOD International Conference on Management of Data (research track) 2024
- IEEE International Conference on Data Engineering (ICDE) (Ph.D. Symposium) 2023
- ACM SIGMOD International Conference on Management of Data (research track) 2023
- International Conference on Very Large Databases (VLDB) (research track) 2023
- ACM SIGMOD International Conference on Management of Data (research track) 2022
- International Conference on Very Large Databases (VLDB) (research track) 2022
- International Conference on Extending Database Technology (EDBT) (demonstration track) 2022
- ACM SIGMOD International Conference on Management of Data (research track) 2021
- International Conference on Extending Database Technology (EDBT) (demonstration track) 2021
- International Conference on Very Large Databases (VLDB) (demonstration track) 2021

Other Reviewing and Community Service

- Mentor of Xuan Luo (Simon Fraser University) at New Researcher Symposium, SIGMOD 2024
- Reviewer of the TKDE Journal 2024
- Reviewer of the VLDB Journal 2023
- Reviewer of the VLDB Journal 2021
- Reviewer of User Interface Software and Technology (UIST) 2021
- Session chair at International Conference on Extending Database Technology (EDBT) 2021
- External reviewer of ACM CHI Conference on Human Factors in Computing Systems 2021
- External reviewer of ACM SIGMOD International Conference on Management of Data 2018–2020

Problem Setter & Judge of Competitive Programming Contests

- State University Bangladesh Inter-University Programming Contest, Bangladesh 2015
- SAARC Programming Contest, Bangladesh University, Bangladesh 2014
- Islamic University of Technology ICT Fest, Bangladesh 2014
- SAARC Programming Contest, Bangladesh University, Bangladesh 2013
- Bangladesh Informatics Olympiad, National Round, Bangladesh 2013
- North South University Inter-University Programming Contest, Bangladesh 2013
- ACM International Collegiate Programming Contest, Dhaka, Bangladesh 2012