Saaduddin Mahmud

LinkedIn | smahmud@umass.edu | saadmahmud.com

Education

University of Massachusetts Amherst — Massachusetts, USA
Ph.D. in Computer Science (3rd Year, CGPA: 4.00/4.00) — *September 2021 – Present*

University of Dhaka — Dhaka, Bangladesh
B.Sc. in Computer Science and Engineering (CGPA: 3.86/4.00) — *January 2016 – December 2019*

Experience

Research Internship — June 2023 – September 2023
Nissan Advanced Technology Center — *Silicon Valley, USA*
- Explanation generation for autonomous vehicle (AV) systems.

Teaching Assistant — February 2023 – May 2023
University of Massachusetts Amherst. — *Massachusetts, USA*
- COMPSCI 589: Introduction to Machine Learning

Research Assistant — September 2021 – Present
Resource-Bounded Reasoning Lab, MCICS, UMass Amherst. — *Massachusetts, USA*
- Advisor: Professor Shlomo Zilberstein
- Safe, complaint, and explainable sequential decision-making.

Lecturer — January 2021 – July 2021
CSE, Ahsanullah University of Science and Technology. — *Dhaka, Bangladesh*
- Lectured on Math for Computer Science and Compiler Design.

Research Assistant — September 2018 – December 2020
Cognitive Agents & Interaction Lab (CAIL), CSE, University of Dhaka — *Dhaka, Bangladesh*
- Multi-Agent decision-making.

Research Goals

My doctoral research is centered around enhancing the safety, compliance, and explainability of sequential decision-making systems. Over the course of my academic journey, I have developed a diverse skill set in relevant areas such as Reinforcement Learning (RL), Learning from Demonstrations (LfD), Explanation Generation, Shared Autonomy, and Continual Learning. Currently, my focus lies on two intriguing topics: learning discrete representations such as Automata, Binary Decision Diagrams, and Tree structures for transparent and incremental learning, and developing sequential perturbation-based explanations for RL agents.

PhD Research

1. Explaining the Behavior of POMDP-based Agents Through the Impact of Counterfactual Information.
Saaduddin Mahmud, Marcell VazquezChanlatte, Stefan Witwicki and Shlomo Zilberstein. (**Under Review**).

2. Learning Constraints on Autonomous Behavior from Proactive Feedback.
Connor Basich*, **Saaduddin Mahmud***, and Shlomo Zilberstein. (**IROS 2023**).

3. Explanation-Guided Reward Alignment.
Saaduddin Mahmud, Sandhya Saisubramanian, and Shlomo Zilberstein. (**IJCAI 2023**).

4. Semi-Autonomous Systems with Contextual Competence Awareness
Saaduddin Mahmud, Connor Basich, and Shlomo Zilberstein. (**AAMAS 2023**).

5. Causal Explanations for Sequential Decision Making Under Uncertainty: Foundations and Analysis.
Samer B. Nashed, **Saaduddin Mahmud**, Claudia V. Goldman, and Shlomo Zilberstein.(**Under Review JAIR, AAMAS 2023, Ext. Abs.**).

6. Estimating Causal Responsibility for Explaining Autonomous Behavior
   Saaduddin Mahmud*, Samer B. Nashed*, Claudia V. Goldman, and Shlomo Zilberstein. (**EXTRAAMAS workshop at AAMAS 2023**).

7. REVEALE: A Framework for Reward Verification and Learning.
   Saaduddin Mahmud, Sandhya Saisubramanian, and Shlomo Zilberstein. (**SafeAI sy AAAI 2023, Best Paper Award Nomination**).

Undergrad Research

1. Learning Optimal Temperature Region for Solving Mixed Integer Functional DCOPs.
   Saaduddin Mahmud, Md. Mosaddek Khan, Moumita Choudhury, Long Tran-Thanh, and Nicholas R. Jennings. **(IJCAI, 2020)**.

2. AED: An Anytime Evolutionary DCOP Algorithm.
   Saaduddin Mahmud, Moumita Choudhury, Md. Mosaddek Khan, Long Tran-Thanh, and Nicholas R. Jennings. **(AAMAS, 2020)**.

3. A Particle Swarm Based Algorithm for Functional Distributed Constraint Optimization Problems.
   Moumita Choudhury, **Saaduddin Mahmud**, and Md. Mosaddek Khan. **(AAAI, 2020)**.

4. Applying Population-Based Algorithms to Solve Large (F)DCOPs.
   Saaduddin Mahmud and Moumita Choudhury (Equal Contribution). **Undergrad Thesis, University Of Dhaka, 2020.**

5. A Simulation-Based Online Planning Algorithm for Multi-Agent Cooperative Environments.
   Rafid Amir Mahmud, Fahim Faisal, **Saaduddin Mahmud**, and Md. Mosaddek Khan. **(AAMAS, 2022, Ext. Abs.)**.

Honors & Awards

B.Sc. Scholarship by the University Grants Commission
Awarded for outstanding performance in B.Sc.

H.S.C. Scholarship by the Bangladesh Government
Awarded for outstanding performance in High-School.

Programming Skills

Languages: Python, Julia, C/C++, Java, JavaScript.
Libraries JAX, PyTorch, TensorFlow.

Reference

Professor Shlomo Zilberstein | *Professor, CICS, University of Massachusetts Amherst, MA, USA*

- Email: shlomo@cs.umass.edu
- Personal Website: https://groups.cs.umass.edu/shlomo/