Saaduddin Mahmud

smahmud@umass.edu — saadmahmud.com — LinkedIn

Education

University of Massachusetts Amherst — Amherst, MA
Ph.D. in Computer Science (5th Year) — *September 2021 – Present*

University of Massachusetts Amherst — Amherst, MA
M.S. in Computer Science (CGPA: 4.00/4.00) — *September 2021 – May 2024*

University of Dhaka — Dhaka, Bangladesh
B.S. in Computer Science and Engineering — *January 2016 – December 2019*

Experience

AI Research and Applications Researcher Intern — June 2024 – August 2024
Nissan Group North America, Inc. — *Nissan Advanced Technology Center, San Jose, CA*
- Natural language routing query processing on OpenStreetMap using LLMs.

AI Research and Applications Researcher Intern — June 2023 – September 2023
Nissan Group North America, Inc. — *Nissan Advanced Technology Center, San Jose, CA*
- Explanation generation for POMDPs.

Research Assistant — September 2021 – Present
Resource-Bounded Reasoning Lab, MCICS, UMass Amherst. — *Massachusetts, USA*
- Thesis: Toward Alignment of Agent-based Systems from Unstructured Instruction.
- Advisor: Professor Shlomo Zilberstein.

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

Research Assistant — January 2020 – December 2020
Cognitive Agents & Interaction Lab (CAIL), CSE, University of Dhaka — *Dhaka, Bangladesh*
- Topic: Multi-Agent Coordination.

Selected Peer-Reviewed Publications

1. Inference-Aware Prompt Optimization for Aligning Black-Box Large Language Models.
Saaduddin Mahmud, Mason Nakamura, Kyle Hollins Wray, Shlomo Zilberstein. (**AAAI 2026**)

2. CoLLAB: A Framework for Designing Scalable Benchmarks for Agentic LLMs
Saaduddin Mahmud, Eugene Bagdasarian, Shlomo Zilberstein. (**Scaling Environments for Agents Workshop at NeurIPS, 2025**)

3. MAPLE: A Framework for Active Preference Learning Guided by Large Language Models
Saaduddin Mahmud, Mason Nakamura, and Shlomo Zilberstein. (**AAAI 2025**).

4. Causal Explanations for Sequential Decision Making Under Uncertainty: Foundations and Analysis.
Samer B. Nashed, **Saaduddin Mahmud**, Claudia V. Goldman, and Shlomo Zilberstein. (**Journal of Artificial Intelligence Research, 2025.**)

5. Explaining the Behavior of POMDP-based Agents Through the Impact of Counterfactual Information.
Saaduddin Mahmud, Marcell VazquezChanlatte, Stefan Witwicki and Shlomo Zilberstein. (**AAMAS, 2024**).

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

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

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

9. REVEALE: A Framework for Reward Verification and Learning.
   Saaduddin Mahmud, Sandhya Saisubramanian, and Shlomo Zilberstein. (**SafeAI Workshop at AAAI 2023**).

10. Learning Optimal Temperature Region for Solving Mixed Integer Functional DCOPs
    Saaduddin Mahmud, Md. Mosaddek Khan, Moumita Choudhury, Long Tran-Thanh, Nicholas R. Jennings. (**IJCAI 2021**).

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

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

Patents

1. Dynamic Refinement of Custom Classes Using Zero-Shot Images *(Under review)*.

2. Vehicle Decision Making Using Sequential Information Probing. *US Patent App. 18/429,196*.

Preprints & Under Review

1. Terrarium: Revisiting the Blackboard for Studying Multi-agent Attacks.
   Mason Nakamura, Abhinav Kumar, **Saaduddin Mahmud**, Sahar Abdelnabi, Shlomo Zilberstein, Eugene Bagdasarian. (Under review, ICLR 2026)

2. Distributed Multi-agent Coordination Using Multimodal Foundation Models.
   Saaduddin Mahmud, Dorian Benhamou-Goldfajn, Shlomo Zilberstein. arXiv:2501.14189, 2025.

Honors & Awards

Distinguished PhD Candidate 2024
Recognition for outstanding performance in Ph.D. candidacy qualification.

B.Sc. Scholarship by the University Grants Commission 2021
Awarded for outstanding performance in B.Sc.

Professional Service

Conference Reviewer: **AAAI Conference on Artificial Intelligence (AAAI)**; **International Joint Conference on Artificial Intelligence (IJCAI)**.

Teaching & Mentoring

Teaching Assistant: **COMPSCI 589 — Machine Learning** (Fall, 2023).

Undergraduate Student Mentoring: **Commonwealth Honors College** for senior thesis, (2025 - 2026).

Programming Skills

Languages: Python, C++, C, Julia

ML/DL: PyTorch, JAX/Flax, PyTorch Lightning, vLLM, Triton

Systems: CUDA, Linux, Docker, SLURM, Git

Data/Maps: OpenStreetMap

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/