Arshia Soltani Moakhar

arshia.soltani2@gmail.com Google Scholar Personal Website

Education

University of Maryland **June 2025 – Present**
Ph.D. in Computer Science *Maryland, USA*
Sharif University of Technology **September 2019 – September 2024**
B.S. in Computer Engineering, GPA: 18.41/20.00 *Tehran, Iran*

- Overall GPA: 18.41/20.00
- Last three years: 19.15/20.00

Publications and Preprints

- **A. Soltani Moakhar***, E. Iofinova*, Elias Frantar, D. Alistarh, "SPADE: Sparsity-Guided Debugging for Deep Neural Networks," *ICML Conference*, 2024, (ICML 2024).
- **A. Soltani Moakhar*** ··· (44 authors) ··· Enzo Ferrante, Sara Hooker, Marzieh Fadaee, "Kaleidoscope: In-language Exams for Massively Multilingual Vision Evaluation," *ICLR*, 2025, (ICLR).
- **A. Soltani Moakhar**, T. Laoaron, F. Ghahremani, K. Banihashem, MT. Hajiaghayi , "Active Learning for Decision Trees with Provable Guarantees," *ICLR*, 2025, (ICLR).
- M. Azizmalayeri, **A. Soltani Moakhar**, A. Zarei, R. Zohrabi, M.T. Manzuri, M.H. Rohban, "Your Out-of-Distribution Detection Method is Not Robust!," *Advances in Neural Information Processing Systems 36*, 2022, (NeurIPS 2022).
- Maral Jabbarishiviari, **A. Soltani Moakhar**, "Software 1.0 Strengths for Interpretability and Data Efficiency," *The Second Tiny Papers Track at ICLR*, 2024, (ICLR 2024 Tinypapers).
- **A. Soltani Moakhar**, M. Azizmalayeri, H. Mirzaei, M.T. Manzuri, M.H. Rohban, "Seeking Next Layer Neurons' Attention for Error-Backpropagation-Like Training in a Multi-Agent Network Framework," *arXiv*, 2023, (arXiv).
- Y. Cheng and **A. Soltani Moakhar**, C. Fan, K. Faghih, P. Hosseini, W. Wang and S. Feizi, "Temporal Blindness in Multi-Turn LLM Agents: Misaligned Tool Use vs. Human Time Perception," *Priprint*, 2025, (arXiv).
- H. Mirzaei, M. Jafari, H.R. Dehbashi, A. Ansari, S. Ghobadi, M. Hadi, **A. Soltani Moakhar**, Mohammad Azizmalayeri, M. Soleymani Baghshah, M.H. Rohban, "RODEO: Robust Out-of-Distribution Detection Via Exposing Adaptive Outliers," *ICML Conference*, 2024, (ICML 2024).
- (53 authors) ··· **A. Soltani Moakhar**, ···, Sara Hooker, Antoine Bosselut, "INCLUDE: Evaluating Multilingual Language Understanding with Regional Knowledge," *ICLR* ***Spotlight***, 2024, (ICLR Spotlight 2024).
- A.Y. Parast, P. Hosseini, H. Asadollahzadeh, **A. Soltani Moakhar**, B. Azam, S. Feizi, N. Akhtar, "GHOST: Hallucination-Inducing Image Generation for Multimodal LLMs," *ICLR*, 2025, (ICLR).

Research Interests

• Multi-Agent Systems • LLM Reasoning • LLM Efficiency • ML Interpretability

Research Experience

Researcher in Agentic LLMs and Theoretical Computer Science **Sep 2024- Present**
University of Maryland, Supervised by: Prof. MohammadTaghi Hajiaghayi and Prof. Soheil Feizi

- **Agentic systems**

  From June 2025, I worked in Soheil's lab on tool-using agents, focusing on mitigating time blindness. This refers to the agents' inability to perceive or account for the time elapsed between user interactions.

- **Mathematical Analysis of Active Learning in Decision Trees**

  In MohammadTaghi's lab, I investigated the theoretical aspects of active learning in decision trees, focusing on sample and time complexity. In one project, I derived sample complexity bounds for decision trees handling continuous features. In another, I established time complexity upper bounds for decision trees with non-uniform binary features.

Independent Research

- **Multilingual Evaluation (With Cohere team) (Spotlight ICLR 2025)**

  We introduced a dataset consisting of multilingual multiple-choice questions sourced from official exams across various countries, ensuring cultural authenticity beyond simple translations of English benchmarks. In a subsequent project, where I am a joint first author, we expanded this dataset to include multimodal questions.

Internship in Interpretability and Sparsity in Deep Neural Networks **Feb 2023 - Sep 2023**

IST Austria, Supervised by: Prof. Dan Alistarh

- **Sparsity-Guided Debugging for Deep Neural Networks (ICML 2024)**

  I improved the performance of several interpretability techniques by sparsifying the network on a selected subset of samples before applying the interpretability method. As the first author on an ICML paper, I developed the core theoretical ideas and implemented most of the code. After publication, we extended this work to explore similar sparsification techniques for large language models.

Research Assistant in Robust and Interpretable Machine Learning Lab **Aug 2021 - Feb 2024**

Sharif University, Supervised by: Prof. Mohammad Hossein Rohban

- **Robust Out-of-Distribution (OOD) Detection Using GAN Architecture (NeurIPS 2022)**

  I first demonstrated critical vulnerabilities in existing robust OOD detection methods against end-to-end adversarial attacks. Subsequently, we proposed an OOD detection algorithm inspired by Generative Adversarial Network (GAN) architecture and adversarial training.

Honors and Awards

2022 International Collegiate Programming Contest **(ICPC)** World Final participation

2019 **Silver Medal** in International Olympiad of Informatics (IOI)

2018 **First Place** in Iranian National Olympiad of Informatics

2017 **Silver Medal** in Iranian National Olympiad of Informatics

Academic services

Question Designer and Lecturer **Aug 2020 – Feb 2024**

Iranian National Olympiad in Informatics

Created and reviewed combinatorial problems for the National Olympiad in Informatics.

Taught and mentored Iranian gold medalists and IOI team members in competitive programming.

Delivered lectures on graph theory at the National Olympiad in Informatics summer camp.

Designed and selected algorithmic problems for the National Olympiad in Informatics, including summer camp exams and International Olympiad in Informatics (IOI) team selection tests.

Volunteer Scientific Committee Member **Feb 2021 - Sep 2021**

Rastaiha (Student Association)

Designed online workshops in Game Theory for high school students.

Conference Reviewer

NeurIPS 2024, ICML 2025, NeurIPS 2025, AAAI 2026

Teaching Assistant **Spring 2023**

Sharif University of Technology

Teaching Assistant for Machine Learning course lectured by Prof. Abolfazl Motahari.

Related Coursework

Sharif University

Artificial Intelligence	20.0/20.0
Machine Learning	20.0/20.0
Adv Information Retrieval(NLP)	19.9/20.0
Medical Image Processing	18.9/20.0

Online Courses

Deep Learning Specialization	DeepLearning.AI
Practical Reinforcement Learning	HSE university
Generative Adversarial Networks	DeepLearning.AI
Game Theory I, II	Stanford University