

ARSHIA SOLTANI MOAKHAR

✉ arshia.soltani2@gmail.com [in](#) [LinkedIn](#) [GH](#) [GitHub](#) [Website](#)

Education

Sharif University of Technology

September 2019 – September 2024

Bachelor of Science in Computer Engineering

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,” *NeurIPS ATTRIB Workshop, ICML Conference*, 2024, ([ICML 2024](#)).
- 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](#)).
- 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](#)).
- 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](#)).
- (53 authors) ... **A. Soltani Moakhar**, ..., Sara Hooker, Antoine Bosselut, “INCLUDE: Evaluating Multilingual Language Understanding with Regional Knowledge,” *Under review in ICLR*, scores: {8,8,8,5}, 2025, ([OpenReview](#)).
- **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](#)).

Research Interests

- Machine Learning Interpretability
- Sparse Neural Networks
- Deep Learning
- Robustness

Research Experience

Remote Researcher in Theoretical Computer Science

Sep 2024- Present

University of Maryland, Supervised by: [Prof. MohammadTaghi Hajiaghayi](#)

• Mathematical Analysis of Active Learning in Decision Trees

I am conducting research on the sample complexity of active learning for decision trees.

Internship in Interpretability and Sparsity in Deep Neural Networks

Feb 2023 - Present

IST Austria, Supervised by: [Prof. Dan Alistarh](#)

• Sparsity-Guided Debugging for Deep Neural Networks ([ICML](#))

I enhanced the performance of various interpretability methods by sparsifying the network on a selected sample, before applying the interpretability method. As the first author I wrote most of code and developed the theoretical insight.

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](#))

Initially, I identified vulnerabilities in existing Robust OOD detection methods to end-to-end adversarial attacks. Subsequently, we proposed an OOD detection algorithm inspired by Generative Adversarial Network (GAN) architecture and adversarial training.

Independent Research

• Integrating Neural Networks into Software for Enhanced Interpretability ([ICLR 2024 Tinypapers](#))

This project introduces a Python library that enables the use of neural networks to learn the conditions in if statements. This allows programs to leverage the power of neural networks while maintaining interpretability.

- **Multilingual Large Language Model Evaluation (With Cohere team) (Under review ICLR 2025)**
In this project, we proposed a dataset comprising multilingual multi-choice questions extracted from official exams of various countries. This approach accounts for the cultural biases of different languages, unlike translations of English questions. I employed OCR techniques to extract texts from Persian Olympiad exams and then used LLMs to identify issues, which were subsequently addressed manually.

Honors and Awards

- 2022 International Collegiate Programming Contest (**ICPC**) World Final participation
- 2019 **Silver Medal** in International Olympiad of Informatics (IOI)
- 2019 **Bronze Medal** in International junior competitive programming competition, infoO(1)CUP
- 2018 **First Place** in Iranian National Olympiad of Informatics
- 2017 **Silver Medal** in Iranian National Olympiad of Informatics

Selected Presentations

Leveraging Sparsity for Enhanced Interpretability in Deep Neural Networks

Sep 2023

Institute of Science and Technology Austria (IST Austria)

Delivered a lecture on various interpretability methods for Neural Networks, evaluation methods of interpretability faithfulness, and discussed how sparsity could significantly improve their performance.

Teaching Experience

NeurIPS 2024 reviewer

Aug 2024

Served as a reviewer for the Conference on Neural Information Processing Systems (NeurIPS), evaluating and providing feedback on 6 out-of-distribution detection research papers.

Teaching Assistant

Spring 2023

Sharif University of Technology

Teaching Assistant for Machine Learning course lectured by [Prof. Abolfazl Motahari](#).

Question Designer and lecturer

Aug 2020 - Feb 2024

Iranian National Olympiad in Informatics

Proposed and selected algorithmic problems for Iran’s National Olympiad in Informatics, specifically for summer camp exams and International Olympiad of Informatics (IOI) team selection exams.
Proposed and selected combinatorial problems for Iran’s National Olympiad in Informatics.
Instructed and consulted Iranian gold medalists and International Olympiad of Informatics (IOI) team members in competitive programming.
Graph theory lecturer in Olympiad of Informatics national summer camp.

Volunteer Scientific Committee Member

Feb 2021 - Sep 2021

Rastaiha (Student Association)

Designed online workshops in Game Theory for high school students.

Related Coursework

Sharif University		Online Courses	
Artificial Intelligence	20.0/20.0	Deep Learning Specialization	DeepLearning.AI
Machine Learning	20.0/20.0	Practical Reinforcement Learning	HSE university
Adv Information Retrieval(NLP)	19.9/20.0	Generative Adversarial Networks	DeepLearning.AI
Medical Image Processing	18.9/20.0	Game Theory I, II	Stanford University

Skills

Programming	Python C++ SQL Bash L ^A T _E X CUDA
Frameworks	PyTorch NumPy Pandas Scikit-Learn Matplotlib Jupyter Django
Languages	Persian: Native English: TOEFL iBT Score 106