Arshia Soltani Moakhar

arshia.soltani2@gmail.com Google Scholar Personal 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).
- **A. Soltani Moakhar***(joint first author) ··· (44 authors) ··· Enzo Ferrante, Sara Hooker, Marzieh Fadaee, "Kaleidoscope: In-language Exams for Massively Multilingual Vision Evaluation," *arXiv*, 2025, (arXiv).
- **A. Soltani Moakhar**, T. Laoaron, F. Ghahremani, K. Banihashem, MT. Hajiaghayi , "Active Learning for Decision Trees with Provable Guarantees," *under review in NeurIPS*, 2025, .
- 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," *ICLR **Spotlight***, 2024, (ICLR Spotlight 2024).
- **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.

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 Evaluation (With Cohere team) (Spotlight ICLR 2025)+ (arXiv)**

  In the first project, we introduced a dataset consisting of multilingual multiple-choice questions sourced from official exams across various countries. This approach captures the cultural nuances, providing a more authentic representation than translations of English questions. In the second project, where I am a joint first author, we extended this to incorporate multimodal questions. As a result, we proposed the largest multilingual multimodal dataset designed for evaluating VLMs.

Internship in Interpretability and Sparsity in Deep Neural Networks **Feb 2023 - Mar 2025**

IST Austria, Supervised by: Prof. Dan Alistarh

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

  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.

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, infO(1)CUP

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

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

Academic services

ICML 2025 reviewer **March 2025**

NeurIPS 2024 reviewer **Aug 2024**

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

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

Skills

Programming	Python | C++ | SQL | Bash | LaTeX| CUDA
Frameworks	PyTorch | NumPy | Pandas | Scikit-Learn | Matplotlib | Jupyter | Django
Languages	Persian: Native | English: TOEFL iBT Score 106