

Amir Mehrpanah

amirmehrpanah@gmail.com | +46 76 707 9243 | github.io | LinkedIn | Github

Research Interests

I enjoy problem solving and integrating engineering with AI research to deliver a better product.

My PhD concerns with formal approaches towards a deeper understanding of AI. This can be a result of defining explainability or interpretability in mathematical forms and building theoretical models that can predict the defined quantities of a trained AI model. My work has been focused on neural networks trained for vision tasks.

Education

- **KTH Royal Institute of Technology** Jun '22–Jan '27
Sweden, WASP Affiliated Ph.D. student
(Expected)
Thesis: Explainability and Uncertainty of Deep Neural Networks
- **Shahid Beheshti University of Tehran** Sep '19–Oct '21
Iran, Ms.c. in Data Science
Thesis: The Problem of Oversmoothing in Graph Neural Networks
Ranked 7th in Iran's national entrance exam
Ranked 1st in the batch
GPA: 18/20
- **Ferdowsi University of Mashhad** Sep '15–Jun '19
Iran, Bs.c. in Applied Mathematics
Ranked 2nd in the batch
GPA: 18/20

Selected Publications/Talks

- **Improving Adversarial Robustness of Attribution via Implicit Regularization** 2026
International Conference on Machine Learning (ICML),
Amir Mehrpanah, Matteo Gamba, Hossein Azizpour
We analyze attribution robustness in CNNs and ViTs, showing that SGD can improve robustness of gradient-based attribution, while attention-based attribution requires kernelized attention for such gains.
- **Toward a Principled Theory of XAI via Spectral Analysis** 2025
Oral presentation of accepted paper at the inaugural EurIPS Conference
Theory of Explainable AI Workshop,
Amir Mehrpanah, Matteo Gamba, Hossein Azizpour
- **Toward a Principled Theory of XAI via Spectral Analysis** 2025
eXCV Workshop (ICCV), Non-Proceedings Track
Amir Mehrpanah, Matteo Gamba, Hossein Azizpour
Position paper proposing spectral analysis as a principled framework for XAI, analyzing ante-hoc vs. post-hoc approaches, and outlining future research directions.

- On the Complexity-Faithfulness Trade-off of Gradient-Based Explanations** 2025
 International Conference on Computer Vision (ICCV)
Amir Mehrpanah, Matteo Gamba, Kevin Smith, Hossein Azizpour
 Proposed a spectral framework to quantify the smoothness–faithfulness trade-off in gradient-based attribution, formally defined the ‘explanation gap,’ and validated on smooth parameterization of ReLU networks across datasets.
- On Spectral Properties of Gradient-based Explanation Methods** 2024
 European Conference on Computer Vision (ECCV)
Amir Mehrpanah, Erik Englesson, Hossein Azizpour
 This work proposes a unified probabilistic view to gradient based explanation methods, and shows formally and experimentally that these methods are biased towards amplification of high frequency information in classification. The proposed mathematical model is then used to define a quantifiable measure of faithfulness, and tune a perturbation scale that maximizes it.
- Logistic-Normal Likelihoods for Heteroscedastic Label Noise** 2023
 Transactions on Machine Learning Research (TMLR)
 Erik Englesson, **Amir Mehrpanah**, Hossein Azizpour
 Reformulating classification tasks as regression tasks and using techniques developed for robust regression to train robust classifiers. This results in better performance in the presence of label noise.

Work Experience

- Remote Research Internship**, Tehran, Shomara Jul '20–Jun '21
 A literature study and classification of methods around approximate nearest neighbor search algorithms, e.g. (FAISS). Writing technical reports, comparing different methods, and quick-start guides for future solutions.
- Open Source Project for Training Classifiers on Persian License Plates** Aug '19–Jul '20
 A long-term open-source project aimed at making training object detection or segmentation models easier for Persian fonts by generating synthetic images of Persian license plate with data augmentation.
- C# Game Developer with Unity**, Mashhad, Soda Fanavar Rayaneh Mandehgar Jun '18–Sep '19
 Gamification of finding and renting parking spots under the supervision of Rasool Ramezani. Implementing game logic and optimization for Android devices.

Academic Experience

- Teaching Assistant for a Graduate Course in Advanced Deep Learning** Sep '22–Jan '27 (Expected)
 KTH, Course given by Hossein Azizpour.
- Teaching Assistant for a Graduate Course in Probabilistic Graphical Models** Mar '23–Sep '23
 KTH, Course given by John Folkesson.

Skills

- Programming Languages:** Python, Java, C#, C + , MATLAB
- Frameworks:** PyTorch, JAX, TensorFlow