Amirabbas Afzali

Curriculum Vitae

+98 911 317 2840
amir8afzali@gmail.com
amirabbas-afzali.github.io
socials: - - -

Research Interests

- Preference Optimization and AI Alignment
- Trustworthy AI & Adversarial ML
- Theory of Optimization and ML
- Bandits Algorithms
- Reinforcement Learning, with a focus on Offline RL and Robust RL

Education

2021–2025 **Sharif University of Technology | SUT**, *Bachelor's degree*, Tehran, Iran,
Major in Electrical Engineering, Communication Systems sub-major, Minor in Mathematics.
GPA – 19.40/20; Ranked 9th out of 184 students.

2018–2021 **Shahid Beheshti High School (NODET)**, *High School Diploma*, Sari, Iran,
GPA – 19.40/20.

Publications

2025 **LORE: Lagrangian-Optimized Robust Embeddings for Visual Encoders**, *Borna Khodabandeh*, Amirabbas Afzali*, Amirhossein Afsharrad, Shahabeddin Mousavi, Sanjay Lall, Sajjad Amini, Seyed-Mohsen Moosavi-Dezfooli*, Under review for NeurIPS 2025, Arxiv

2025 **Towards Fair Retrieval: Controlling Bias through a Backpack-Inspired Architecture**, *Amirabbas Afzali*, Amirreza Velae*, Iman Ahmadi, Amirhossein Afsharrad, Sanjay Lall, Mohammad Aliannejadi*, Under review for EMNLP 2025

2024 **One Goal, Many Challenges: Robust Preference Optimization Amid Content-Aware and Multi-Source Noise**, *Amirabbas Afzali, Amirhossein Afsharrad, Shahabeddin Mousavi, Sanjay Lall*, Published at RLC 2025, OpenReview, Arxiv
- Shorter version accepted at the ICLR workshop on Spurious Correlation and Shortcut Learning, 2025.

2024 **Aligning Visual Contrastive learning models via Preference Optimization**, *Amirabbas Afzali*, Borna Khodabandeh*, Ali Rasekh, Mahyar JafariNodeh, Sepehr Kazemi, Simon Gottschalk*, Published at ICLR 2025 , OpenReview, Arxiv

2024 **ULTra: Unveiling Latent Token Interpretability in Transformer-Based Understanding and Segmentation**, *Hesam Hosseini, Ghazal Hosseini, Amirabbas Afzali, Sajjad Amini, Amir Houmansadr*, Under review for ICCV 2025, Arxiv
- Shorter version accepted at the ICLR workshop on XAI4Science, 2025.

2024 **Clustering Time Series Data with Gaussian Mixture Embeddings in a Graph Autoencoder Framework**, *Amirabbas Afzali*, Hesam Hosseini*, Mohmmadamin Mirzai, Arash Amini*, Published at AAAI 2025 AI4Research Workshop, OpenReview, Arxiv

2024 **Toward efficient decoding of index modulated FMCW Signals**, *Amirreza Zameni*, Hesam Hosseini*, Amirabbas Afzali*, Mohammad Mahdi Mojahedian, Rouhollah Amiri*, Under review for IEEE Journal on Selected Areas in Communications

In-Process

2025–Present **MIXALIGN: Preference Data Augmentation for Easy-to-Hard Alignment**, *Advised by Prof. Maria Brbic*, Investigated easy-to-hard generalization in LLM alignment and developed MixAlign, a data augmentation method that aligns models on complex tasks using only easy, reliably annotated preferences.

Course projects

2024 **GAN-BERT**, *Deep Learning project*, SUT, **source**
Implemented the GAN-BERT architecture, which adversarially trains a BERT-based generator against a discriminator to detect and classify LLM-generated texts to the specific model used for generation.

2024 **Hidden Markov Map Matching Through Noise**, *AI project*, SUT, **source**
Implement a map-matching algorithm that uses a Hidden Markov Model (HMM) to find the most likely road route represented by a time-stamped sequence of latitude/longitude pairs.

2023 **Regression problem on CarPrice-assignment dataset**, *Applied Statistics project* , SUT, **source**
A project with the purpose of getting familiar with Regression and some classical ML methods.

2022 **Social Media App**, *OOP project* , SUT, **source**
Developed an application in Java with a GUI using JavaFX and CSS, integrated with a MySQL database for efficient data management.

Experience

Research Experience

2025 **ICLR 2025 Attendance**, Invited to present three accepted papers at the Thirteenth International Conference on Learning Representations (ICLR 2025): *Aligning Visual Contrastive Learning Models via Preference Optimization*, along with two additional workshop papers.

2024–Present **Volunteer Researcher**, Conducting research and contributing to the ongoing projects at Aktus AI, focusing on leveraging artificial intelligence for innovative applications.

2024–2025 **Volunteer Researcher**, Conducting remote research at the L3S Research Center, focusing on controlling biases in vision-language models (VLMs), which resulted in this paper.

2023–2024 **Research Project**, Conducting research on early Alzheimer's disease risk assessment using pre-trained language and audio models. Analyzing speech patterns of healthy individuals and Alzheimer's patients, and investigating neural mechanisms underlying decision-making and cognitive dementia recognition, Ambient Intelligence Research Lab | SUT.

Voluntary Teaching Assistance

- Machine Learning (Head TA)
- Deep Generative Models
- Artificial Intelligence
- Engineering Probability and Statistics
- Reinforcement Learning
- Deep Learning
- Signal processing
- Computer Structure

Relevant Coursework

The symbol "+" denotes graduate coursework.

- High Dimensional Probability +
- Stochastic Process
- Signals and Systems
- Artificial Intelligence
- Machine Learning with Graphs + (Auditing)
- Reinforcement Learning +
- Deep Learning +
- Convex Optimization 1
- Machine Learning
- Online Learning & Bandits + (Auditing)

Skills

Technical Skills

Programming: C, C++, Java, Python, MATLAB, Julia, R, Verilog, Assembly, SQL

Machine Learning Libraries: PyTorch, Tensorflow, Pytorch-Geometric, OpenCV, Scikit-learn, NumPy, Pandas, Matplotlib

Soft Skills

Languages: Persian (native), English (advanced)

Misc: Problem-Solving, Collaboration, Communication, Teaching