Farshad Sangari Abiz

Generative AI Researcher · AI Software Engineer

(+98) 99-0066-8659 farshads7778@gmail.com Farshadsangari
Farshad-Sangari Google Scholar farshadsangari.github.io
School of ECE, University of Tehran, Tehran, Iran

Summary

AI Engineer & Team Lead for Computer Vision and Generative AI at ModAI, building production LLM/VLM systems for multimodal retrieval and recommendation—embedding pipelines, hybrid search, re-ranking, and evaluation. MSc in Artificial Intelligence focused on interpretability in generative models and representation learning. I combine research-driven rigor with scalable deployment to deliver reliable, robust vision–language systems in real products.

Education

- **College of ECE, University of Tehran** — Tehran, Iran
  MSc: Computer Engineering – Artificial Intelligence; GPA: 18.73/20 (3.89/4) — Sep. 2021 – Sep. 2024
  - *Thesis:* Interpretable Representation Learning via Deep Generative Models; Advisers: Prof. Babak N. Araabi, Dr. Reshad Hosseini
- **K.N.Toosi University of Technology** — Tehran, Iran
  BSc: Electrical Engineering – Telecommunication; GPA: 17.00/20 (3.64/4) — Sep. 2017 – Sep. 2021

Research & Industry Experience

- **ModAI – AI-Powered Fashion Discovery Platform** — Tehran, Iran
  AI Engineer & Team Lead – Computer Vision & Generative AI — *Oct. 2023 – Present*
  - Lead CV/GenAI R&D for a fashion marketplace: multimodal retrieval & recommendation (text/image/hybrid); raised ranking quality via preference fine-tuning (NDCG@10 +15%, Precision@10 +10%).
  - Ship production LLM/VLM services for routing, retrieval, and re-ranking on Docker/Kubernetes with blue/green releases, model versioning, and continuous evaluation (99.5% uptime).
  - Design scalable embedding pipelines (batch + real-time) with autoscaling; harden for peak load via load and chaos testing.
  - Own search & data stack (Qdrant, Elasticsearch, PostgreSQL, MongoDB); tune HNSW/ANN and hybrid BM25+vector retrieval for latency/quality trade-offs.
  - Build low-latency MLOps data plane (MinIO, Redis, Kafka/Redpanda, NATS JetStream) with idempotent consumers, backpressure, and effectively-once processing.
  - Drive observability and reliability (Prometheus/Grafana, OpenTelemetry, SigNoz/Sentry, SLO alerting); improve robustness to noisy/cropped images with metric/contrastive learning.
  - Deliver product AI features: caption refinement, image editing, virtual try-on, and SEO product copy with schema.org markup.
- **RamonAI – Agentic AI Sales Assistant** — Tehran, Iran
  AI Software Engineer (Part-time) — *Feb. 2025 – Present*
  - Build a robust, scalable agentic AI sales assistant (LangGraph/LangChain) that acts as a 24/7 admin for online shops—handling product discovery, Q&A, and sales workflows.
  - Ship multimodal shopping features including image search, descriptive (text) search, and virtual try-on to help merchants convert and support customers end-to-end.
- **Machine Learning and Computational Modeling Lab (University of Tehran)** — Tehran, Iran
  Interpretable Representation Learning with Generative Models — *Feb. 2022 – Sep. 2024*
  - Researched interpretable/disentangled latent representations in VAEs, diffusion models, and GANs; explored contrastive objectives (angular/circle loss) to improve representation quality.
  - *Research Advisers: Prof. Babak N. Araabi – Dr. Reshad Hosseini*
- **NBIC, School of ECE (University of Tehran)** — Tehran, Iran
  Emotion Detection — *Dec. 2022 – Sep. 2023*
  - Built CLIP-based emotion detection; collected a 120-subject dataset and trained/evaluated aligned image-text representations.
  - *Research Adviser: Dr. A. H. Vahabie*
- **Freelance** — Tehran, Iran
  Artificial Intelligence Engineer — *Nov. 2019 – Sep. 2023*
  - Delivered freelance AI/ML projects spanning computer vision, deep learning, and data analysis; selected open-source work available on GitHub.

Technical Skills & Tools

- **Languages**: Python, Go, C++, Bash, MATLAB, R; Persian (Native), English (Fluent)
- **Generative AI & CV**: PyTorch, PyTorch Lightning, TensorFlow, VLMs/LLMs, CLIP, diffusion models, representation learning, metric/contrastive learning
- **Agentic AI**: LangChain, LangGraph, CrewAI, Memori, Langfuse
- **Search & Data**: Qdrant, Elasticsearch, PostgreSQL, MongoDB, MySQL, QuestDB, MinIO, Redis
- **MLOps & Pipelines**: Docker, Kubernetes, Airflow, MLflow, Mage AI, GitLab CI, Argo CD, Terraform, Ansible
- **Systems & Observability**: Kafka/Redpanda, NATS JetStream, RabbitMQ, gRPC, Prometheus, Grafana, OpenTelemetry, SigNoz, Sentry, Whylogs

Publications

- **Enhancing Interpretability of Sparse Latent Representations with Class Information**
  Farshad Sangari, Reshad Hosseini, Babak Nadjar Araabi
  arXiv preprint Jan. 2026

Selected Courses

2022	Deep Generative Models (19.8/20)	2022	Trustworthy AI (20/20)
2021	Machine Learning (19.1/20)	2022	Deep Learning and Applications (18/20)
2022	Statistical Inference (18.4/20)	2021	Data Science (18.75/20)
2018	Basics of Optimization (19.5/20)	2018	Linear Algebra (18.5/20)

Teaching Experience

Sep. 2024	**(Lead TA)** Probabilistic Graphical Models & Deep Generative Models – Dr. M. Tavassolipour	University of Tehran
Jan. 2024	**(Lead TA)** Trustworthy AI – Dr. M. Tavassolipour	University of Tehran
Sep. 2023	**(Lead TA)** Probabilistic Graphical Models & Deep Generative Models – Dr. M. Tavassolipour, Dr. M.A. Sadeghi	University of Tehran
Jan. 2023	**(Lead TA)** Deep Learning with Applications – Dr. R. Hosseini	University of Tehran
Sep. 2023	Machine Learning – Dr. S. Haratizadeh	University of Tehran
Sep. 2023	Statistical Inference – Dr. M.-R. A. Dehaqani	University of Tehran
Apr. 2023	PyTorch Tutorial	YouTube course

Honors & Awards

- Ranked 3rd in the M.Sc. in Artificial Intelligence cohort, University of Tehran *(Year/Semester)*
- Ranked 30th among +25,000 in the National Master Entrance Exam *Sep. 2021*
- Among the top 1% of Iran Universities Entrance Exam *Aug. 2017*

Selected Msc Projects

- **Trustworthy A.I.** Tehran, Iran *Nov. 2023*
  - Robust Representation using Angular Loss
  - Enhancing Model Interpretability by leveraging LIME and SHAP for Comprehensive Explanations
  - Out-of-Distribution Detection: Enhancing Model Robustness and Generalization
- **Deep Generative Models** Tehran, Iran *Nov. 2023*
  - RealNVP Implementation
  - Score Matching Network Implementation
  - Diffusion Model Implementation
  - VAE and its Variations Implementation
- **Deep Learning and Applications** Tehran, Iran *Nov. 2023*
  - Deep Fully Connected Network Implementation (NumPy)
  - Saliency Map Prediction using Deep Neural Network
  - Inverting Visual Representations for Convolutional Networks
  - Medical Image Segmentation Using U-Net Architecture
- **Data Science** Tehran, Iran *Nov. 2023*
  - Analyse Coronavirus Infection
  - Amazon Scraper for Laptop
  - Analysis of Alibaba vs MrBilit Tickets' Price