

EBRAHIM MOUSAVI

Generative AI Researcher and Engineer

+374 931 841 22

<https://ebimsv.github.io/>

ebimsv0501@gmail.com

[linkedin.com/in/ebimsv](https://www.linkedin.com/in/ebimsv)

Yerevan, Armenia

Google Scholar

github.com/Ebimsv

medium.com/@ebimsv

WORK EXPERIENCE

AI Team Lead – LLMs & Agents

Makan System / Research Collaborations

Aug 2024 – Now

- Led development of domain-specific LLMs for code and protein analysis, improving benchmark performance by 22% via custom tokenizers and LoRA/QLoRA fine-tuning.
- Engineered sophisticated RAG architectures and autonomous agents using LangGraph and ReAct patterns, reducing hallucination rates by 35% for complex multi-step reasoning.
- Implemented advanced alignment strategies, including DPO and GRPO, to optimize model instruction-following, resulting in a 1.4x increase in human-eval preference scores.
- Scaled production inference using vLLM and TensorRT-LLM, implementing 4-bit AWQ quantization to reduce latency by 60% and memory footprint by 4x.

AI Team Lead

Resana (Joint venture with Rutelia Co., Japan)

Feb 2024 – Aug 2024

Tehran

- Led the development of Image GenAI systems and Intelligent Document Processing (IDP) agents, achieving a 94% accuracy rate in automated data extraction.

Computer Vision Developer

AIHomeDesign | aihomedesign.com

Sep 2023 – Feb 2024

Remote

- Designed and deployed production-grade Stable Diffusion pipelines for real-estate image enhancement, including virtual staging, item removal, and scene transformation.
- Reduced manual editing workload by 75% through automated computer vision pipelines, improving turnaround time and operational scalability.
- Optimized diffusion model inference for high-resolution image generation, maintaining strong visual consistency across 10k+ daily Day-to-Dusk transformations.
- Improved model quality, latency, and reliability through prompt engineering, parameter tuning, and performance optimization.

AI Instructor & Technical Writer

Freelance | medium.com/@ebimsv

July 2023 – Present

Remote

- Authored 100+ technical articles on Machine Learning, Deep Learning, Generative AI, LLMs, and PyTorch, achieving 120,000+ total views and strong global engagement.
- Wrote practical tutorials covering model training, evaluation, neural network design, and Python programming, simplifying complex concepts with hands-on code examples.

SUMMARY

AI Team Lead and Researcher with a proven track record in architecting LLM agents, RAG pipelines, and generative vision systems. Published Q1 researcher with expertise in multi-GPU training and alignment (DPO/GRPO). Specialized in bridging the gap between cutting-edge deep learning research and scalable production deployment.

EDUCATION

M.Sc. in IT (AI Focus)

Tarbiat Modares University

Sept 2018 – Sept 2021 Tehran, Iran

- Thesis:** Designing a Fashion Recommendation System using Deep Learning.
- Key Contribution:** Developed a multi-modal feature extraction pipeline (Vision + Metadata) to improve personalized recommendation.

B.Sc. in Software Engineering

Islamic Azad University

Feb 2009 – Feb 2015 Isfahan, Iran

- Thesis:** Implementation of a Pyramid Security System with Firewall.

PUBLICATIONS

Q1 Journal Articles

- ATTENDEE:** an Affective Tutoring system based on facial Emotion recognition and head pose Estimation to personalize e-learning environment *Journal: Education and Information Technologies (Springer)*, 2023
- FELA-DNN framework** *Journal: Soil Dynamics and Earthquake Engineering (Elsevier)*, 2023

arXiv Preprint

- How Self-Supervised Learning Can be Used for Fine-Grained Head Pose Estimation? *arXiv Preprint*, 2021

- Mentored 50+ engineers and students through structured programs, focusing on real-world ML, DL, and Generative AI applications.
- Developed educational content and learning roadmaps on fine-tuning, retrieval-augmented generation (RAG), prompt engineering, and modern AI deployment workflows.

Deep Learning Researcher and Developer

Baharan Company & Freelance

Jan 2021 – Aug 2023 Tehran, Iran

- Developed end-to-end computer vision solutions for commercial clients, including facial recognition, identity verification, and intelligent image analysis systems.
- Designed and implemented deep learning pipelines for model training, evaluation, and deployment using modern frameworks such as PyTorch and TensorFlow.

Network & Systems Engineer

Freelance / Contract Roles

Feb 2017 – Mar 2020 Iran

- Designed, configured, and maintained secure network infrastructures using Cisco (CCNA-level) and Microsoft systems (Active Directory, DNS, DHCP, Group Policy).

TECHNICAL SKILLS

- **Generative AI & LLMs:** Transformers, LoRA/QLoRA, DPO/GRPO, RAG, LangChain, LangGraph, AutoGen, CrewAI.
- **Deep Learning & CV:** PyTorch, Computer Vision, Diffusion Models, Stable Diffusion, Self-Supervised Learning.
- **Engineering & Ops:** Multi-GPU Training (FSDP/DeepSpeed), vLLM, TensorRT-LLM, Quantization (AWQ/GGUF), Docker, Linux.
- **Languages & Core:** Python, SQL, Git, REST APIs.

LANGUAGES

Persian

English

(IELTS 6.0 – Duolingo 110)

Turkish

SELECTED PROJECTS

LLMs and AI Agents Course

- A comprehensive tutorial series for developers interested in GenAI, LLMs, and deploying AI agents in real-world scenarios.

Large Language Model with PyTorch

- Implemented a custom LLM from scratch using PyTorch, focusing on training efficiency and text generation, showcasing a strong foundational understanding of LLM architectures.

Facial Age Estimation

- Prediction of human age from facial images with Transformers.

REFEREES

Prof. of IT Gholam Ali Montazer

@ Tarbiat Modares University, Tehran, Iran

Montazer@modares.ac.ir

Google Scholar

Prof. of Bioinformatics Ali Masoudi-Nejad

@ University of Tehran

Amasoudin@ut.ac.ir

Google Scholar

LinkedIn