amadrezafrh.github.io
linkedin.com/in/amadrezafrh
amadreza.farahani@outlook.com
+39 351 641 1678
Turin, Italy (**open to relocation**)

Amadreza Farahani

AI/ML Engineer

Skills

Software Development	Python, C/C++, SQL, Bash, Linux, Git, CI/CD, FastAPI, Docker, Kubernetes, Terraform, Prometheus, Grafana
Machine Learning	NumPy, Pandas, scikit-learn, XGBoost, LightGBM, PyTorch, TensorFlow, MLflow, Airflow, GANs, U-Net, YOLO
Large Language Models	Transformers, RAG, LangChain, LangGraph, vLLM, TorchServe, Milvus, HuggingFace
Cloud Technologies	AWS (Bedrock, SageMaker, ECS, EKS, EC2)

Professional Experience

Machine Learning Engineer

Turin

Zirak

11/2024 – Current

- Led a team of two to build a cross-platform LLM meeting assistant (Electron/React + **FastAPI**) with RAG (**LangChain**/**LangGraph**), integrating speaker verification, and real-time transcription on **AWS Bedrock** for meeting summaries and Q&A.
- Built a rail-safety computer-vision pipeline (**Python**/**YOLOv8**/**PyTorch**) to detect sign defects and segment/track rails from nadir and frontal views.
- Defined user stories, architecture designs, and detailed design requirements for safety and diagnostics modules based on customer needs, ensuring alignment with project specifications and driving an efficient architecture.
- Designed and implemented a custom **U-Net** for multispectral imagery, automated data ingestion, training and evaluation with **Airflow**, outperforming internal baselines (Random Forest, **XGBoost**, K-Means) in F1 by ~10%.
- Refined and optimized an aquatic-vegetation detection pipeline on AWS using multispectral satellite data, dockerized and trained on **SageMaker**, and deployed on **Kubernetes** (**EKS** on **EC2**) with **Prometheus** metrics, keeping accuracy drift <3% and reducing manual image review by ~80%.
- Deployed an LLM assistant for consultancy and HR workflows, serving **HuggingFace** models via **vLLM**/**TorchServe** with **MLflow** tracking and **Grafana** dashboards, cutting manual document processing by ~50 hours/month and enabling the product for other clients.

Visiting Researcher – Master Thesis

Darmstadt

Technische Universität Darmstadt | CROSSING

08/2023 – 11/2024

- Designed a Diffusion **Transformer** for voice conversion (Python, PyTorch) on **HPC** GPUs, improving speaker similarity by ~9% over open-source baselines with fewer sampling steps, enabling real-time conversion.
- Deployed low-latency inference on AWS **ECS** with EC2 GPUs, using MLflow to track experiments and monitor production performance.

AI Engineer

Tehran

Part AI Research Center

06/2019 – 03/2022

- Refined and fine-tuned an on-device keyword-spotting Android app in **Java**, using hard negative samples to reduce false positives by ~30%.
- Optimized a speaker verification system on **TensorFlow**, fine-tuned on new domains that reduced the Equal Error Rate (EER) by ~2%.
- Engineered **Test-Driven Development TDD** for Persian **ASR** and **NLP** services; wrote unit/integration tests to reach 100% coverage and set up **Jenkins CI/CD** to automate builds, tests, and releases.
- Mentored new team members on clean, production-ready Python (structure, testing, logging), improving code quality and easing model iterations.

Internship Experience

Associate Machine Learning Developer

Turin

AROL S.p.a

01/2023 – 09/2023

- Developed a synthetic machinery data generator (TensorFlow + **Flask** + **MongoDB**) and a customer testing dashboard, packaged with **Docker Compose** (client/server/DB) and REST API docs to avoid exposing proprietary data.
- Designed KPIs and Python tests to validate synthetic vs. real machinery signals and API responses, integrating these checks into the team's **CI pipeline** and reducing manual QA effort for customer demos.
- Reviewed over 5,000 lines of production **Python** and ML pipeline code (training, inference, data preprocessing), adding tests and reducing compute cost while making models easier to debug and extend.

Computer Vision

Tehran

Megamouj Co. | University of Science and Technology (IUST)

11/2018 – 06/2019

- Automated a traffic-flow monitoring system (**C++**/Python, **OpenCV**, YOLOv3) for vehicle detection, tracking, and flow estimation, reducing manual video review in a national-scale project.
- Labeled and augmented traffic video data and trained/evaluated detection and tracking models in **Python**, tuning YOLOv3 hyperparameters and metrics to keep counting errors within acceptable limits across varying lighting and weather conditions.

Education

M.Sc. Computer Engineering - Artificial Intelligence & Data Analytics

Turin

Politecnico di Torino

2022 - 2025

B.Sc. Electrical and Computer Engineering - Telecommunications

Tehran

University of Science and Technology (IUST)

2016 – 2021