

Amir Esmaeili

AI/ML ENGINEER | LLM • RAG • BACKEND SYSTEMS
esmaeili.amir@gmail.com [LinkedIn](#) [GitHub](#) [Telephone](#)
Location: Tabriz, Iran (Open to relocation)



Professional Summary

AI/ML Engineer specializing in production LLM applications, Retrieval-Augmented Generation (RAG), and scalable back-end systems. Experienced in designing AI services, vector search, and distributed ML infrastructure from development to deployment.

Work Experience

ML Engineer/ Data Scientist *Desmer Guvenlik* [🔗](#) **02/2024 – 03/2026**

- Engineered and deployed AI optimization models for logistics operations, reducing service delays by 30% and cutting operational costs across a network of 100+ service points.
- Applied causal inference techniques to evaluate business process changes, directly contributing to a 20% reduction in service delivery delays.
- Integrated ML models into live production systems, enabling real-time decision-making and reducing manual oversight by 40% across daily operations.
- Conducted demand forecasting on ATM cash withdrawals across 50+ locations during seasonal periods, optimizing resource allocation and reducing cash shortages by 25%

Backend Developer *Aral Studio* **02/2023 – 01/2024**

- Designed and optimized backend systems capable of handling high user traffic in real time, ensuring excellent uptime and system reliability under peak loads.
- Improved API response time by 35% through codebase refactoring and strategic database indexing across 10+ high-traffic endpoints.
- Architected microservices infrastructure serving multiple product lines, reducing deployment time by 40% and improving team scalability.

Backend Developer *NOAY* **09/2021 – 01/2023**

- Built secure and scalable backend infrastructure for a cryptocurrency platform, supporting high-frequency transactions with near-zero failure rates.
- Developed robust APIs and microservices that reduced server load by 25% and improved overall system responsiveness.
- Partnered with frontend teams to ensure seamless integration and intuitive user experience, increasing user satisfaction scores.

Qualifications

Programming	Python, C++, JavaScript, PHP
LLMs & AI	PyTorch, Transformers, RAG, Embeddings, OpenAI API, Prompt Engineering
Backend	FastAPI, Django, Flask, SQLAlchemy, REST APIs, Microservices
Databases	PostgreSQL, Redis, MongoDB, pgvector
Search	Qdrant, OpenSearch
Infrastructure	Docker, Celery, Linux, Git
Languages	English – IELTS: 7, Persian, Turkish, Azerbaijani
Soft Skills	Problem-solving, Research methodology, Teamwork, Project management

Projects

Enterprise Multi-Tenant RAG Platform *Personal Project* **06/2026 – Present**

- Built a multi-tenant Retrieval-Augmented Generation (RAG) platform for secure document search and AI-powered chat across workspaces and knowledge bases.
- Developed a FastAPI backend with PostgreSQL, SQLAlchemy, RBAC, JWT authentication, and tenant isolation for production-ready deployment.
- Designed an asynchronous document ingestion pipeline using Celery and Redis for OCR, parsing, chunking, embedding, and indexing.
- Implemented hybrid retrieval using Qdrant, OpenSearch BM25, Reciprocal Rank Fusion (RRF), reranking, and citation-based answer generation.
- Added production observability with Prometheus, OpenTelemetry, Grafana, and retrieval quality evaluation using Recall@K, MRR, and nDCG.

Question Generating App *Freelance Project*

07/2025 – 08/2025

- Built a Flask-based AI service that generates multiple-choice questions from PDFs using OpenAI-compatible LLMs and configurable prompt templates.
- Implemented a document processing pipeline including PDF extraction, semantic chunking, MongoDB storage, duplicate detection, and context retrieval.
- Added production-grade generation safeguards including answer validation, retry logic, explanation refinement, response repair, and LaTeX normalization.
- Developed REST APIs, bilingual (Persian/English) quiz generation, interactive HTML output, and comprehensive unit tests for core system components.

Resume Matching System *Personal Project*

07/2025 – 09/2025

- Built an AI-powered resume matching platform using semantic embeddings to rank candidates against job descriptions.
- Developed a Django backend with PostgreSQL and pgvector for scalable vector search and modular REST APIs.
- Integrated Hugging Face embedding models and OpenAI-powered job description enhancement through an AI chatbot.
- Implemented asynchronous processing with Celery, Redis, and Docker for scalable deployment and background task execution.

AI Fitness Video Studio *Personal Project*

06/2026 – Present

- Built a full-stack AI video generation platform for creating consistent fitness videos using reusable characters, environments, and prompt templates.
- Developed a FastAPI backend for prompt generation, job orchestration, provider routing, asset management, and reproducible generation history.
- Designed a provider-agnostic generation layer supporting local ComfyUI workflows and cloud-based video generation models through a unified interface.
- Automated AI workflows for image generation, video synthesis, upscaling, and asynchronous job execution with configurable pipelines.
- Built a Next.js dashboard for asset management, prompt editing, provider selection, job tracking, and generated media preview.

Education

M.Sc. Soft Computing and Artificial Intelligence *University of Tabriz*

09/2023 – present

- Selected as a "Brilliant Talent", granted direct admission to M.Sc. without National University Entrance Exam
- Awarded a full scholarship for M.Sc. studies, covering all educational expenses
- Selected Specialized Course Grades: Artificial Intelligence: A, Data Mining: A+, Machine Learning: B, Deep Learning: B+

B.Sc. Computer Science *University of Tabriz*

09/2019 – 06/2023

- Full Scholarship Recipient
- Ranked 4th in Overall B.Sc. GPA
- Honorable Mention in The International Collegiate Programming Contest - 2019, 2020, 2021
- Top 3% in National University Entrance Exam
- Selected Specialized Course Grades: Advanced Programming: A+, Data Structures: A+, Probability: A+, Discrete mathematics: A+, Linear algebra: A, Algorithms: A+, Theory of Computation: A+, Compiler: A+

High School Diploma *National Organization for Development of Exceptional Talents (Sampad)*

2014 – 2018