Arefeh Qavami

Email: arefeh.qavami@gmail.com, Mobile: (+98)9115067027, Linkedin: Arefeh Qavami, Residency: Tehran, Iran

About me

AI Engineer with a strong background in **Machine Learning**, **Deep Learning**, **Computer Vision**, and **Natural Language Processing**, with experience applying modern architectures including **Transformers** to vision applications. Currently focused on **Large Language Models (LLMs)**, **Generative AI**, **Prompt Engineering**, **Agentic AI**, and **Multi-Agent Systems**, with hands-on experience in **AI agents, tool calling, LangGraph, LangChain, MCP, and Elasticsearch**. Experienced in designing AI workflows that integrate LLMs, tools, APIs, data retrieval, and specialized AI services.

Education

- **Alzahra University** — Tehran, Iran
  Master of Science in Artificial Intelligence; ranked 2nd — *September 2017 – October 2020*
- **Shahid Beheshti University** — Tehran, Iran
  Bachelor of Science in Software Engineering — *September 2012 – February 2017*

Work Experience

- **AI Engineer, RCDAT,** Tehran, Iran — August 2025 – Present
  - **LLM-Based Text Analysis and Summarization:**
    - Developed an end-to-end text analysis pipeline using **Elasticsearch**, **Jina embeddings**, **clustering**, and LLM-based **summarization**.
    - Applied clustering and topic modeling methods to organize large-scale textual data and select representative documents.
    - Developed **targeted prompting** and **structured LLM outputs** for summarization, topic analysis, and emotion analysis.
  - **Agentic AI and Multi-Agent Systems:**
    - Designed and prototyped **Supervisor-based Agentic AI** architectures using **LangGraph**, enabling LLM-driven **tool calling** and routing requests through specialized tools and subgraphs.
    - Built **AI workflows** integrating LLMs, APIs, and specialized tools, with **state management** for maintaining request, tool, and execution context across agent steps.
    - Developed **MCP-based AI services** that expose data-analysis and clustering capabilities as reusable **tools for AI agents** and enable tool-based LLM workflows.
  - **AI Engineering and System Development:**
    - Implemented **Elasticsearch-based retrieval, autocomplete, and search suggestions** to support user queries and large-scale text analysis workflows.
    - Implemented **asynchronous and concurrent processing** for AI workflows, using Python **multithreading and Semaphore-based concurrency control** to manage parallel LLM requests.
    - Developed and containerized AI backend services using **Python, FastAPI, Docker, and Elasticsearch**.
    - Developed frontend interfaces for AI applications using **React** and **Streamlit**.
    - Used **Cursor AI** as an AI-assisted development tool for coding and accelerating software development.
- **AI Engineer, Nojan Robotics,** Tehran, Iran — February 2025 – July 2025
  - **Pistachio Sorting System – Deep Learning Classification:**
    - Prepared and augmented pistachio image datasets for industrial classification tasks.
    - Developed and trained deep learning models using modern transformer-based architectures including **ConvNeXt**.
    - Converted trained models to **ONNX** for efficient deployment on real-time R&D sorting devices.
- **AI Engineer, City Development and Innovation Company,** Tehran, Iran — June 2021 – June 2024
  - **Face Verification System – Deep Learning:**
    - Developed a Siamese-network-based face verification system by fine-tuning pre-trained deep learning models, achieving approximately **98% accuracy**.

- **National Card Detection – Computer Vision:**
  - Developed national card detection using image processing and template matching with approximately **93% accuracy**.
  - Developed a YOLO-based detection solution achieving approximately **95% accuracy**.
- **CamQC – Database-Based Quality Control:**
  - Developed MongoDB queries and data-analysis workflows for quality control of ATM camera systems in collaboration with a three-person engineering team.

Technical Skills

- **LLM & Agentic AI:** Large Language Models, Generative AI, prompt engineering, AI agents, tool calling, AI workflows, multi-agent systems, structured outputs, LangGraph, LangChain, MCP
- **NLP & Information Retrieval:** text preprocessing, tokenization, semantic similarity, text embeddings, knowledge retrieval, topic modeling, text clustering, representative document selection
- **AI Engineering:** API integration, asynchronous programming, concurrent processing, multithreading, Semaphore-based concurrency control, AI service architecture
- **Machine Learning:** classification, regression, clustering, dimensionality reduction, DBSCAN, HDBSCAN, K-Means, Hierarchical, Spectral, BERTopic, Top2Vec, LDA, PCA, XGBoost, CatBoost
- **Deep Learning:** CNNs, Transformers, transfer learning, model evaluation, hyperparameter tuning, model deployment
- **Computer Vision:** object detection, image classification, face verification, image processing, template matching

Programming Skills

- **Languages:** Python, SQL
- **AI / ML Frameworks:** PyTorch, TensorFlow, Keras, Scikit-learn, Hugging Face Transformers, Timm, Jina AI
- **LLM / AI Frameworks:** LangChain, LangGraph, MCP
- **Backend:** FastAPI, Flask, Django
- **Frontend / UI:** React, Streamlit
- **Data & Search:** Elasticsearch, MongoDB, MySQL, NumPy, Pandas
- **Tools & Infrastructure:** Git, Docker, Linux, DVC, ONNX, Cursor AI

Languages

- **English:** Fluent, TOEFL iBT: 99 (2016)
- **Persian:** Native

Certificates

- Advanced Computer Vision with TensorFlow – DeepLearning.AI, Coursera