

ALIREZA RAZINEJAD

Senior Full-Stack Web & AI Systems Engineer

Yerevan, Armenia • +374 55 475950 • a.razinejad@gmail.com • linkedin.com/in/alirazinejad • github.com/USSDlover • alrazi.netlify.app

PROFESSIONAL SUMMARY

Innovative **Senior Full-Stack Web Engineer** and **AI Systems Architect** with 11+ years of experience designing scalable microservices, high-concurrency web applications, and production-grade Machine Learning/Deep Learning systems. Specialist in bridging modern web frameworks (**React, Next.js App Router, Angular, Node.js, NestJS, Python**) with advanced AI architectures, including Vision-Language Models (VLMs), Retrieval-Augmented Generation (RAG) pipelines, and vector databases (pgvector, Pinecone). Accomplished in end-to-end system design, high-throughput token streaming via SSE/WebSockets, and autonomous agent workflows. Proven track record of leading cross-functional engineering teams, optimizing system latency, and deploying deep learning pipelines into production.

CORE TECHNICAL COMPETENCIES

Full-Stack Web & System Architecture: Next.js (App Router, Server Actions), React, Angular (v2–v19, NgRx), Vue.js, TypeScript, Node.js, NestJS, Express.js, Python, C# .NET Core, Java, Microservices Architecture, RESTful APIs, GraphQL.

AI, Machine Learning & Deep Learning: Deep Learning Architectures, Vision-Language Models (Qwen2.5-VL), Multimodal Models (CLIP), Image Processing/CUDA Kernels, LLM Tool/Function-Calling Agents, Dynamic Prompt Caching, Model Fine-Tuning & Ingestion Pipelines.

AI Infrastructure & Vector DBs: RAG Pipelines, pgvector, Pinecone, Ollama Local Runtimes, Real-time SSE/WebSocket Interceptors, Stream Sanitization, Cursor IDE, Claude Code, OpenAI & Anthropic APIs.

Databases, Cloud & DevOps: PostgreSQL, MySQL, MongoDB, Redis, Schema Design & Query Tuning, Docker, GitHub Actions CI/CD, AWS (S3, Lambda, CloudWatch Metrics), Linux, System Performance Optimization.

PROFESSIONAL EXPERIENCE

Senior Lead Full Stack & AI Systems Engineer — Next Top Tech (Dubai, UAE / Remote) March 2024 – July 2026

- **AI Pipeline Architecture:** Spearheaded the end-to-end system design and integration of generative AI pipelines and LLM streaming responses into web applications via WebSockets and Server-Sent Events (SSE) with sub-second response times.
- **AI-Assisted Developer Workflows:** Pioneered AI-native development workflows using Cursor IDE and custom AI code assistants across an 8-engineer squad, increasing feature throughput by 25% while maintaining strict testing quality.
- **Full-Stack System Design:** Architected modular Node.js microservices and high-performance React/Angular frontends, supporting high-concurrency client demands and system uptime.
- **DevOps & Quality Infrastructure:** Containerized microservices using Docker and implemented automated CI/CD deployment pipelines via GitHub Actions, establishing rigorous quality gates across distributed teams.

Senior Full Stack Software Engineer (Contractor) — Vyne (United States / Remote) March 2023 – February 2024

- **Server Data Processing & Dashboards:** Built scalable server-side telemetry processing pipelines on AWS using Python and C# .NET Core to process and aggregate high-volume metric logs for real-time visualization on interactive React dashboards.
- **System Integration:** Standardized REST and GraphQL integration specifications across distributed backends, increasing system data transport reliability and cross-database synchronization.

Senior Full Stack Software Engineer — Techstep (Stockholm, Sweden / Remote) June 2021 – February 2023

- **Monolith Decoupling & Distributed Systems:** Decoupled legacy enterprise monoliths into high-availability Java and Node.js microservices, reducing backend latency by 35% and improving platform scalability.
- **Cross-Platform Infrastructure:** Created standardized Docker environments across staging and production systems, ensuring full environmental consistency for web and mobile (React Native) applications.

Senior Full Stack Lead — IDP Education Ltd (Delhi, India / Remote) March 2019 – May 2021

- **Distributed System Engineering:** Led an engineering team of 10+ developers in refactoring legacy codebases into event-driven Node.js microservices and modular React frontend applications, accelerating system performance by 25%.

Senior Software Engineer (Contractor) — Itonomy E-Commerce Professionals September 2017 – February 2019
(Utrecht, Netherlands / Remote)

- **Full-Stack E-Commerce Architecture:** Developed full-stack web applications featuring optimized payment systems, database query tuning, and optimized Core Web Vitals.

Full Stack Engineer — Trade Me (New Zealand / Remote) January 2015 – August 2017

- **High-Traffic Web Systems:** Engineered database schemas, server endpoints, and user interfaces for high-traffic web applications, lowering average backend latency across user workflows.

KEY AI & FULL-STACK SYSTEM PROJECTS

MedEd AI: Real-Time Local Clinical Vision & Educational Platform

Role: Creator, AI Architect & Lead Systems Engineer | Stack: Next.js 14 App Router, Ollama (qwen2.5vl), TypeScript, Tailwind CSS, Prisma, PostgreSQL, SSE

- **Deep Learning Vision Engine:** Engineered an on-device clinical vision platform using a local Vision-Language Model (qwen2.5vl:latest) via Ollama to process image/text data with zero external data exposure.
- **Streaming Interceptor Architecture:** Designed a custom Server & Client Stream Interceptor to parse raw model reasoning streams and dynamically route structured output states into UI drawers in real time.

Hostel AI: Autonomous B2B Property Management SaaS

Role: Founder & Lead Systems Architect | Stack: Next.js App Router, PostgreSQL (pgvector), Node.js/NestJS, Docker

- **Autonomous ML Workflows:** Architected an autonomous B2B SaaS platform utilizing function-calling agent workflows, vector embeddings (pgvector), structured JSON output parsing, and prompt caching to achieve 99.9% reliability in agent executions.

EDUCATION

Bachelor of Science in Computer Science — University of Tehran

2011 – 2014