

Reza Enayati

Los Angeles, CA | r3zsoft@gmail.com | rezaenayati.co | [GitHub](https://GitHub.com/rezaenayati) | [LinkedIn](https://LinkedIn.com/in/reza-enayati)

SKILLS

Languages: C#, Python, TypeScript, Java, Go, SQL

Frameworks: .NET, Next.js, React, PyTorch, Semantic Kernel, Django, Spring Boot

AI/ML: OpenAI API, Anthropic SDK, TensorFlow, RAG, LoRA

Cloud & Data: Azure (Blob, Functions, Service Bus, Foundry), AWS (S3, Bedrock), PostgreSQL, Cosmos DB, Redis

EXPERIENCE

koderAI

Backend Engineer

Irvine, CA

Sept 2025 – Present

- Reduced LLM integration time from days to minutes by architecting unified gateway in **C#/.NET** supporting 37+ models across 7 providers with automatic failover and cost tracking
- Cut API response latency by 40% through **Redis** caching of model catalogs and optimized token estimation, saving \$2K+/month in redundant API calls
- Enabled AI agents to autonomously manage 10K+ project files by exposing file-system operations as **REST APIs** with **Azure Blob Storage** backend

Software Engineer Intern

Jun 2025 – Sept 2025

- Prevented \$5K+/month in cost overruns by implementing pre-request token estimation using **tiktoken** and credit limit enforcement across all LLM endpoints
- Improved prompt reliability by 35% by building structured pipelines in **C#** for multi-agent workflows with automated validation and **xUnit** regression testing

nexlife

Contract Software Engineer

Irvine, CA

Aug 2025 – Present

- Reduced patient intake processing time by 60% by building async workflow orchestration using **Inngest** and **Next.js** with zero dropped transactions
- Achieved 99.9% webhook reliability by implementing cryptographic signature verification and idempotent retry logic in **TypeScript** across 7 API integrations
- Eliminated runtime type errors by engineering automated schema generation using **Zod** and **GraphQL** codegen, cutting integration bugs by 80%

DJAI

Seattle, WA

Software Engineer Intern

Jun 2024 – Sept 2024

- Reduced model inference latency by 50% by optimizing **Python** ML pipelines for real-time audio feature extraction from live DJ sets
- Deployed end-to-end ML infrastructure on **AWS** using **Spring Boot** microservices, processing 1000+ audio samples daily for real-time mixing models

EDUCATION

California State University, Northridge

Los Angeles, CA

Bachelor of Science in Computer Science; GPA: 3.70

Expected May 2026

PROJECTS

koderAI Console | C#, .NET, Spectre.Console

- Built agentic coding CLI to test LLM capabilities across 40+ models; evolved into daily driver for engineering team after implementing 20+ **AI-callable tools** for Git, file ops, and code generation
- Designed Claude Code-style streaming UI with **Spectre.Console**; enabled rapid prototyping of tool integrations, reducing feature validation time from days to hours

RezAi – Personal AI Twin | PyTorch, LoRA, Hugging Face | [Live Demo](#)

- Fine-tuned **LLaMA 3.1-8B** using **LoRA** and **DPO** on 140+ custom Q&A pairs; deployed to Hugging Face Spaces with 500+ inference requests served