Reza Enayati

rezaenayati.co

Email : r3zsoft@gmail.com

Mobile : +1-347-405-3316

SKILLS

Languages: C#, Python, Java, Swift, C
Frameworks & Libraries: .NET, Semantic Kernel, LangChain, PyTorch, FastAPI
Tools & Ops: Azure Storage, Git, Azure DevOps, Azure AI Foundry, xUnit, OpenTelemetry, Cosmos DB, Redis

EXPERIENCE

koderAi Irvine, CA
Member of Technical Staff *Sept 2025 - Present*

- **koderAI.Chat.Gateway**: Architected and published the **NuGet package**, building on **Microsoft.Extensions.AI's** `IChatClient` to unify access to 37+ LLMs across 7 major providers (OpenAI, XAI, Gemini, Anthropic, Microsoft, Meta, DeepSeek). Added support for tool calling, telemetry, token/cost tracking, Redis-cached model catalogs, and a custom Azure Content Safety layer; adopted as the core dependency across all koderAI applications.
- **LLM Tool Invocation**: Developed internal APIs for file-system operations (clone, create, update, delete) integrated with Azure Blob Storage, and exposed them as `AIFunctions` for LLMs enabling AI agents to directly read, write, and manage project assets in the file repository.
- **Sketch-to-Wireframe API**: Built a user-facing `/v1/design/sketchToWireframe` API where an LLM interprets sketches and generates HTML/CSS, executes the code, captures rendered screenshots, and stores outputs in **Azure Blob**; enabling users to instantly transform hand-drawn ideas into high-fidelity design prototypes.

koderAi Irvine, CA
Software Engineer Intern *Jun 2025 - Sept 2025*

- **Token Estimation & Credit Management**: Implemented token usage estimation with **Microsoft.ML.Tokenizers** and **tiktoken** to enforce credit limits before LLM requests, preventing cost overruns in production; integrated logging of token counts and estimated costs into Azure Table Storage for full visibility.
- **Cosmetic Bug Detection**: Engineered the `/v1/testing/findCosmeticBugs` API, where UI screenshots stored in **Azure Blob Storage** are analyzed by LLMs to identify visual errors and generate structured bug reports (title, description, severity). Used as a feedback loop for LLM-generated UI code, this system reduced recurring front-end issues by an estimated ~40%.
- **Prompt Workflows**: Developed structured prompt pipelines for multi-role AI agents (Product Manager, Web Developer, Designer, QA Tester), exposing them via REST endpoints and validating outputs with **xUnit** unit tests.

DJAI Seattle, WA
Software Engineer Intern *Jun 2024 – Sept 2024*

- **Backend Microservices**: Developed Java microservices with **Spring Boot** to power AI-enabled DJ automation features.
- **ML Pipelines**: Co-built end-to-end machine learning pipelines to extract audio features from live DJ sets and train inference models for real-time music mixing.
- **Data Ingestion & Deployment**: Implemented audio data ingestion, preprocessing, and model deployment workflows, contributing to a fast-paced **Agile** engineering team.

EDUCATION

California State University, Northridge Los Angeles, CA
Bachelor of Science in Computer Science; GPA: 3.70 *Expected May 2026*

PROJECTS

- **RezAi – Personal AI Interview Twin**: Fine-tuned a **LLaMA 3.1-8B** model with **LoRA** and **Direct Preference Optimization (DPO)** on 140+ handcrafted Q&A pairs to capture my tone, technical depth, and personality; deployed live via **Hugging Face Spaces (Gradio)** and integrated into my personal website. [Live Demo] [Repo]
- **RezuMe – AI Resume Tailoring Platform**: Co-developed a full-stack application that tailors resumes to job descriptions using **Gemini 1.5 Flash** and **GPT-4o**. Implemented **embeddings with cosine similarity** to match job description keywords against resumes, built the backend in **Flask** (resume parsing, AI integration), and contributed to the React frontend with **Firebase**-based authentication and resume management. [Repo]