

Diego J. Robles Costa

Tampa, FL • (603) 974-3262 • djrobles444@gmail.com • diegojrobles.com • github.com/diegojrobles • linkedin.com/in/diegojrobles44

SUMMARY

Computer science undergraduate who ships production-grade AI software: a cross-platform LLM healthcare app presented at Google DevFest, a multi-provider AI API server with async streaming inference, and end-to-end web tools built without frameworks. Hands-on with Python, C/C++, LLM application architecture, and API design; seeking a software engineering internship building AI and accelerated-computing products at NVIDIA.

PROJECTS

Ad+ — E-Commerce Advertising Platform

JavaScript, HTML/CSS, LocalStorage

- Architected a local-first advertising dashboard as a single self-contained HTML file in vanilla JavaScript — zero frameworks, zero backend, zero API dependencies — with full client-side data persistence via browser localStorage.
- Built four integrated modules: weekly trending-product tracker, drag-and-drop kanban prospect pipeline, automated generation of personalized outreach and offer letters, and a unified client-communication organizer.

AI Doctor — Cross-Platform AI Healthcare App

React Native, FastAPI, Llama 3.3, Docker

- Built an LLM-powered mobile app delivering symptom assessment, medical recommendations, emergency detection, and provider suggestions; presented architecture and safety design at Google DevFest to developers and industry mentors.
- Dockerized the FastAPI backend and implemented layered LLM safety controls (guardrails, JSON-schema output validation) within a secure API architecture.

AI-Server — Multi-Provider LLM API Server

Python, FastAPI, Pydantic

- Engineered a production-ready inference gateway unifying OpenAI, Anthropic, and Google LLM providers behind one REST interface with async request handling and token-streaming responses.
- Designed endpoints with Pydantic validation, per-client rate limiting, CORS configuration, and comprehensive error handling and logging.

Email Database

C, Data Structures

- Developed a lightweight in-memory database managing user records, strengthening low-level fundamentals in pointers, manual memory management, and data-structure design.

EXPERIENCE

WS Development — IT Intern

Boston, MA | Jul 2026 — Aug 2026

- Administer Windows Server, on-premises Active Directory, and Azure/Entra ID hybrid-identity environments — user accounts, group policies, and access control — across a portfolio of 100+ retail and mixed-use properties.
- Support security operations through system hardening, access reviews, and vulnerability remediation spanning corporate and property-site infrastructure.
- Troubleshoot server, network, and authentication issues across distributed site networks while maintaining infrastructure documentation.

Google Developer Group — Mentorship Program

Tampa, FL | Aug 2025 — Nov 2025

- Selected for a competitive mentorship program in AI/ML, cybersecurity, and CS career development; collaborated with mentors and peers on coding workshops, agile projects, and production AI deployment in Python and C.

SKILLS

Languages	Python, C, C++, C#, JavaScript, TypeScript, HTML/CSS
AI / ML	LLM integration (OpenAI, Anthropic, Google, Llama), prompt engineering & evaluation, AI guardrails & JSON-schema validation, vector databases
Backend	FastAPI, REST API design, async & streaming request handling, rate limiting, error handling & logging, Docker, React Native
Tools & Systems	Git, Linux/CLI, Azure/Entra ID, Windows Server, Active Directory, API deployment

EDUCATION

University of South Florida — B.S. Computer Science

Tampa, FL | Expected Dec 2027

Relevant coursework: Program Design, Discrete Structures, Computer Organization, Linear Systems

Certifications: Google AI Essentials — Google