

Education

Stanford University
M.S. Computer Science. GPA

Stanford, CA
Mar 2028

University of California San Diego
B.S. Data Science. GPA 3.974/4.0
• Provost Honor

La Jolla, CA
Jun 2026

Walnut High School

Walnut, CA
Jun 2022

Experience

Samsung Electronics America
Network System Deployment Intern

Plano, TX
Jun 2026 – Aug 2026

- Built an autonomous agentic system for telecom alarm triage using **LangGraph**, automating alarm parsing, diagnostic planning/execution, and auditable root-cause reporting.
- Implemented a scoped ReAct executor over **FastMCP** read-only tools with hard safety limits and graceful degradation, guaranteeing failure handling
- Built a RAG pipeline (LanceDB + BGE-M3, Docling) turning a **500-page, 200+** alarm-type manual into structured, retrievable per-alarm records.
- Developed a **Django (MVT)** web application around the agents and a PostgreSQL-backed operational dashboard

Cadre AI (AI Consulting)
Part-Time AI Engineer

Del Mar, CA
Jul 2025 – Feb 2026

Project: Serverless AI Agent Automation Platform

- Developed **cloud-native, serverless AI agents** using Azure Functions, Python, and Blob Storage, enabling document automation with zero infrastructure overhead.
- Designed and deployed 5 modular **AI agents** (Sales Proposal, Preflight, SOP, Compliance, Q&A), reducing doc processing time from **2h** → **10m** for **100+** employees with **0%** formatting/compliance errors.

Project: MCP Proxy for Enterprise API Integrations

- Built a **FastAPI-based MCP proxy** enabling ChatGPT to securely call BigQuery/Zoom via **OAuth 2.0**, improving dataset exploration for 20+ internal users.
- Achieved **>99.9% reliability** for long-running ChatGPT sessions through proactive token refresh and request forwarding.
- Streamlined analytics workflows, reducing manual **BigQuery** console navigation by **~70%**.

Project: RAG Admin Portal & User Management System

- Designed and implemented a full-stack **admin portal** for a production RAG system, providing **user management, role-based access control (RBAC), and secure operational governance**.
- Built **end-to-end workflows** for user invitation, token-validated signup, role assignment, and account deactivation, integrating frontend state management with backend APIs.
- Identified and remediated security gaps in exposed authentication endpoints, and improved performance through **client-side caching and controlled refresh mechanisms**.

Project: Intelligent Email Automation Platform (Azure)

- Architected a cloud-based, event-driven email automation system using **FastAPI** and **Azure App Service**.
- Implemented **Microsoft Inbox webhook subscriptions** with automatic validation, renewal, and fault-tolerant startup handling.

- Integrated **OCR** and **LLM** pipelines to extract structured data from heterogeneous document formats and automate downstream validation and response drafting.

LLM Strategies (AI Startup)

Remote

Data Engineering Intern

Dec 2024 – Jun 2025

- Built distributed ingestion services and Flask microservices with **Celery**, **Redis**, and **Docker** to process **10M+ patent and trademark records** from APIs, scrapers, and XML feeds.
- Developed **Flask**-based microservices exposing REST APIs for data/task management and built a **high-accuracy document extraction pipeline** (USPTO API + OCR + Gemini), achieving **99.2%** accuracy on 500+ legal documents.
- Reduced redundant metadata by **85% (2.5M → 362k)** through large-scale **deduplication** and **normalization**.
- Improved pipeline reliability through **asynchronous task execution**, **retry handling**, and **containerized deployment workflows**.

The Wu lab @ Scripps Research

La Jolla, CA

Data Engineering Intern

Nov 2024 – Feb 2025

- Built **FastAPI** services and memory-efficient Python parsers to normalize large-scale genomic datasets (**PostgreSQL** + **S3**).

Research Experience

The Data Science Teaching and Learning Lab (DSTL)

La Jolla, CA

Research Assistant

Mar 2024 – Jun 2024

- Enhanced the [Pandas Tutor visualization tool](https://dl.acm.org/doi/10.1145/3641554.3701933) by improving DataFrame introspection for selective rendering of large datasets – reducing memory usage and improving in-browser interactivity.
- Conduct research on and publish this work as **first co-author** at **ACM SIGCSE** (paper: <https://dl.acm.org/doi/10.1145/3641554.3701933>).