

FRANCIS O'HARA AIDOO

franciskohara@gmail.com | [linkedin.com/in/francis-ohara](https://www.linkedin.com/in/francis-ohara) | github.com/francis-ohara | francisohara.com

EDUCATION

Colby College

Bachelor of Arts in Computer Science

Waterville, ME

Expected May 2027

- **Relevant Coursework:** Neural Networks, Bio-Inspired Machine Learning, Intro to Computer Vision, Software Engineering, Entrepreneurial Software Engineering, Computer Organization, Programming Languages, Data Structures & Algorithms

EXPERIENCE

Medidata Solutions

Data Science Intern, Platform AI | Python, Amazon Bedrock, LLMs, NLP

May 2026 – Present

New York, NY

- Engineered **LLM pipelines** to extract and resolve structured data from unstructured clinical trial protocol documents, enabling automated parsing of **Schedule of Assessments** tables, footnotes, and procedural metadata across **37,000+ clinical trials** processed by Medidata.

Google

Software Engineering Intern | Go, Java, Python, Terraform, Vertex AI, ADK

May 2025 – Aug 2025

Sunnyvale, CA

- Engineered an [officially published Terraform resource](#) in **Go** for deploying generative AI models at scale on Google Cloud **Vertex AI**, unblocking a **\$19.2MM ARR** opportunity for Google Cloud partners.
- Discovered and resolved a critical race condition in the **Vertex AI** backend in **Java** that prevented concurrent model deployment requests; validated fix with comprehensive **JUnit** and **Mockito** test suites.
- Engineered a chat-based **multi-agent system** with specialized sub-agents that autonomously discover open-source GenAI models, provision cloud infrastructure (endpoints), deploy models, and run inference on deployed models using **Python** and **Google Agent Development Kit (ADK)**.

Google

Software Engineering Intern | Java, Apache Beam, GCP Dataflow, Terraform

May 2024 – Aug 2024

Seattle, WA

- Engineered a high-performance, schema-aware **Apache Beam** transform in **Java** for processing CSV data, capable of achieving a throughput of **3 billion+ CSV records/hour** on **Google Cloud Dataflow** and enabling scalable migration of data from legacy systems into the cloud.
- Ensured data integrity and robust fault tolerance for distributed data processing pipelines by validating schema constraints ahead of execution and routing malformed records to a **Dead Letter Queue (DLQ)**.

StyleSyncs

Co-Founder & AI Lead | React, TypeScript, Supabase, IDM-VTON, Gemini API, GCP Cloud Run

Jan 2026 – Present

Waterville, ME

- Co-founded and launched an AI-powered fashion visualization platform, stylesyncs.com, enabling users to virtually try on clothes using generative AI; grew to **100+ registered users** and raised over **\$9,000** in funding through competitive pitch competitions.
- Architecting a **multi-model AI pipeline** across 4+ specialized models — **IDM-VTON** for diffusion-based virtual try-on, **Gemini 3 Pro/Flash Image** for digital avatar creation, and **Gemini 3.5 Flash** for multimodal clothing analysis — and developing a **real-time video try-on** feature for interactive garment overlay on live video feeds.

PROJECTS

Mule Mart | Python, Flask, PostgreSQL, Docker, AWS S3, PyTorch | [GitHub](#) | [Website](#)

Oct 2025 – Present

- Built a full-stack e-commerce marketplace for the Colby College community enabling verified users to securely buy, sell, donate items, and exchange messages in real time.
- Designed a scalable REST API backend with **Flask** and **PostgreSQL**, integrating **AWS S3** presigned URLs for media storage and optimizing **SQLAlchemy** queries to eliminate N+1 bottlenecks and improve response times.
- Implemented **AI-powered semantic search** feature using **SentenceTransformers** and **PyTorch**, which enabled meaning-based product discovery by calculating cosine similarity on vector embeddings.

MuseAnalytics | Python, FastAPI, React, PostgreSQL, Redis, Docker | [Website](#)

Dec 2025 – Present

- Building a full-stack data analytics platform for the Colby College Museum of Art to track and analyze visitor data using **FastAPI**, **React**, **TypeScript**, and **Tailwind CSS**.
- Engineered a high-performance data layer using **SQLAlchemy 2.0** and async **PostgreSQL**, significantly improving concurrency for complex reporting queries.
- Integrated **Redis** caching to serve pre-computed analytics, achieving **sub-100ms** dashboard response times for museum staff and enabling real-time visibility into visitor metrics.

TECHNICAL SKILLS

Languages: Python, Java, Go, TypeScript, JavaScript, C, C++, SQL

Frameworks & Libraries: FastAPI, Flask, React, Next.js, Express.js, Tailwind CSS, SQLAlchemy

Cloud & DevOps: Terraform, Docker, GitHub Actions, AWS (EC2, S3, Lambda, Bedrock), GCP (Cloud Run, Storage), Bash

Data & Testing: PostgreSQL, SQLite, Redis, Apache Beam, Dataflow, Pytest, JUnit, Mockito

AI & ML: TensorFlow, Keras, Scikit-Learn, PyTorch, NumPy, Pandas, Vertex AI, Google ADK, NLP, LLMops, MLOps