

YAHYA OZAY

Houston, TX

yahyaozay1905@gmail.com

linkedin.com/in/yahyaozay

github.com/yaozay

yahyaozay.com

Education

University of Houston - Honor's College

Bachelor of Science in Computer Science, Minor in Data and Society

Expected Graduation: May 2026

Houston, TX

Technical Skills

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

AI/ML: PyTorch, TensorFlow, pandas, LoRA/QLoRA fine-tuning, RAG pipelines, Pinecone, ChromaDB

Frameworks/Tools: React, AWS, Docker, Git (flow & CI/CD), Postman, Linux, FastAPI, Supabase, Tailwind CSS, Figma

Experience

Software Engineer Intern

June 2025 – Present

Pearson Labs (YC F24)

Remote

- Developed 4 AI-powered dataroom features using **TypeScript**, **FastAPI**, **GPT-4o**, **Pinecone**, and **Supabase**, helping streamline document embedding, querying, and retrieval for early-stage users
- Collaborated on multiple backend and frontend tickets through **Linear**, contributing to core product improvements with **Python**, **FastAPI**, and **TypeScript**
- Integrated **Supabase** and **Vercel** for full-stack development, contributing to the **CI/CD flow** with GitHub Actions and feature-based PRs

AI Research Assistant

May 2025 – Present

University of Houston - System and Simulation Lab

Houston, TX

- Fine-tuned a 7B-parameter Mistral LLM on 1M tokens with **HuggingFace** and **LoRA**, reduced average query latency by 92% (850 to 65 ms) for the research team
- Built a **ChromaDB**-backed **RAG** service with HNSW indexing and streaming **FastAPI** endpoints, delivering answers at 99.6% accuracy on a 500-question benchmark
- Shipped a **TypeScript/React** web app that visualizes retrieval context and model confidence

Software Engineer Resident

Nov 2024 – Apr 2025

Headstarter AI

Remote

- Built 14+ Machine Learning, AI-engineering and Full-Stack projects in fast-paced software team environments
- Developed 5+ **neural networks** in **Python**, 11 apps in **Typescript** on **AWS/Vercel** with dev environments
- Implemented **LLM-chaining**, **hyperparameter tuning**, **fine tuning** on 10+ LLM models controlling for latency

Projects

AI Document QA Platform | *Python, FastAPI, Next.js, TypeScript, Supabase, Pinecone*

May 2025

- Ingested PDFs, extracted text and bounding boxes with pdfplumber, embedded 50K+ vectors via **OpenAI**, and stored in **Pinecone** for instant semantic search with page-linked citations
- Launched a **Render**-hosted **FastAPI** service that auto-indexes new **Supabase** Storage docs and returns GPT-4o answers in under 1 second

Real-Time Air Traffic Safety Dashboard | *Node.js, TypeScript, WebSocket, MongoDB, Next.js*

April 2025

- Streamed live flight data from AviationStackAPI via **WebSockets** into a modular **Node** backend, flagging hazards with 80 ms update rate
- Designed a responsive **Next.js** map UI and **MongoDB** alert log, winning **3rd place at Create & Innovate Hackathon** sponsored by HP

PopReel: Tiktok Clone | *TypeScript, Next.js, AWS S3, Prisma, Tailwind CSS, Clerk*

Feb 2025

- Engineered a social video app with **Clerk** auth, real-time feeds, and **AWS S3** storage
- Implemented **Prisma/PostgreSQL** backend and Tailwind-responsive UI

Leadership / Awards

Designer, CougarCS — Designed digital products in Figma

Nov 2024 – Present

Lead Instructor, Coder Kids — Taught K–8 coding basics

Mar 2025 – May 2025

3rd Place, Create & Innovate Hackathon — Real-time flight-safety dashboard

Apr 2025

IT Intern, HSA Lorain — Troubleshoot for a K-12 school

Jul 2021 – Jul 2022