

HENRY EBOMAH

AI Engineer | Fullstack Developer | Software Engineer

US Citizen (Secret) | Remote, GA | (619) 408-9375 | henryebomah@gmail.com | [Henry's LinkedIn](#)

GitHub: <https://github.com/awesomething>

SUMMARY

Stanford-trained Full-stack and AI Engineer with 5+ years of building secure, high-impact LLM systems, RAG pipelines, and complex web applications. Proven ability to own features end-to-end, from architecture to deployment, across cloud and on-prem environments. Deep experience with OpenAI Agents SDK, vector search, embeddings, multimodal interfaces and performance optimization. Passionate about building mission critical AI systems for the public sector.

TECHNICAL SKILLS

Advanced:

React, Python, Vector Search, Azure AI Services, Agents SDK, Next.js, HTML5

Intermediate:

Langchain, AWS, Kubernetes, Terraform, TypeScript, PyTorch, Docker, observability, streaming and tool calling

CAREER HIGHLIGHTS

Project Tech Lead | VideoBook.ai

Tech Stack: Next.js, Google ADK, Python, TypeScript, PyTorch, vector DBs, OpenAI API, Voice model training stack

- Led development of AI career agents - a voice-enabled calendar assistant and 6-in-1 AI recruiter agents that reduced sourcing time by 35%.
- Designed high-level system architecture supporting an AI career agent using Next.js and TypeScript for the frontend and Python for backend integration.
- Trained a foundation voice model from scratch and deployed, production-quality synthetic voices for the app.
- Built an end-to-end voice call feature enabling the AI agent to call users upon resume upload and manage data flow into the platform's database that reduced job-search time for applicants.

PROFESSIONAL EXPERIENCE

Senior Software Engineer | Accenture Federal Services | Remote, CA

Tech Stack: React, Python, Azure, Docker, Kubernetes (exposure), Terraform, OpenAI Agents SDK, Splunk, Weave, Azure

- Co-led the redesign of a major LLM web chat, integrating Azure vector and hybrid search; increasing user engagement by 40%
- Integrated OpenAI Agents SDK including streaming, handoffs, guardrails, and tool calling, enabling more reliable and secure conversational workflows.
- Developed Splunk and Weave observability dashboards to monitor success, latency, and cost tracking by segment; collaborated with stakeholders to define and track SLOs for inference performance.
- Supported R&D of MCP server infrastructure implementing asynchronous event processing, secure tool-calling integration and data flow reliability improvements.

- Worked closely with product, infra, and security teams to ensure compatibility with very strict enterprise and quasi-on-prem environments, gaining exposure to kubernetes, terraform, network constraints and compliance workflows.
- Frequently partnered with users and Federal stakeholders to refine features based on data and feedback.

Cofounder & CTO | ObserveYourFuture.com | Atlanta, GA

Tech Stack: Next.js, Python, TypeScript, AWS, Mongo DB, OpenAI API, WebRTC, Cloudinary, Azure

- Architected and co-developed a high quality WebRTC application using Next.js, TypeScript, AWS, and MongoDB.
- Integrated OpenAI for repeated automation and Cloudinary for secure media processing and distribution.
- Developed a RAG chatbot system enabling course-based Q&A using PDF ingestion and Azure AI search.
- Led technical architecture decisions, working closely with dev team and managed end-to-end feature execution.

Lead Coding Instructor | www.thrivedx.com | San Diego, CA

- Held office hours and assisted junior developers in debugging their MERN app, as they built their milestone projects, until it was launched.
- Led daily stand ups and performed rigorous code reviews for quality performance and attention to detail.
- Participated in the model development lifecycle architecture to create web applications.

EDUCATION & TRAINING

Artificial Intelligence Program | Stanford University | San Francisco, CA | 2024

- Built and tested advanced video synthesis models using deep learning frameworks.
- Implemented AI safety modules to prevent harmful or unethical content generation in text-to-video tasks.
- Conducted model comparisons, optimization experiments and dataset diversity analysis.
- Authored a research paper contributing insights to text-to-video model generation, safety and alignment.

Software Engineering | Chegg's Thinkful | San Diego, CA | 2020

- Fullstack software development, industry-relevant curriculum.
- Created and deployed mobile-first apps with industry relevant languages and pair programming.
- Technical skills included JS and ES6, advanced frontend development, server-side programming like RESTful APIs with No SQL and Computer Science foundations.

CERTIFICATIONS & COURSEWORK

Chegg Skills

<https://www.chegg.com/skills/thinkful-is-now-chegg-skills/>

Simplilearn

<https://success.simplilearn.com/07d640e5-46df-4ef5-a9e9-3b06b73e792d#acc.Khvx69cr>

DeepLearning.AI (Coursera)

<https://coursera.org/share/c42284a982cc2013f86df2cdd7f99558>

Course Topics Completed Include:

- Supervised Machine Learning: Regression and Classification
- Unsupervised Learning
- Recommenders
- Reinforcement Learning