JACOB CHAFFIN

FULLSTACK ENGINEER | VOICE AI　♦ SF BAY AREA, UNITED STATES　📞 6503803288

○ DETAILS ○

94530
SF Bay Area
United States
6503803288
jchaffin57@gmail.com

○ LINKS ○

Github

LinkedIn

Website

○ SKILLS ○

Node.js

Typescript

Python

LLMs

Conversational AI

GraphQL

NLP

Rust

👤 PROFILE

Full-Stack Engineer with 5+ years of experience in Voice AI and real-time conversational systems. Focused on building low-latency infrastructure, modular applications, and orchestration pipelines that connect ASR, NLP, and TTS components. Proficient in Node.js, TypeScript, and Python. Skilled in full-stack development, complex API integration, and system optimization. Delivers production-ready tools that enhance user experience and automate intelligent workflows.

💼 EMPLOYMENT HISTORY

Infra + Product at prosody.ai
2024 — Present

- Built a real-time voice AI infrastructure platform that injects phonetic tone and emotional context into LLMs via a multimodal pipeline.
- Designed and deployed latency-optimized microservices for acoustic feature extraction, tone modeling, and token conditioning.
- Integrated WebSocket + REST APIs with Whisper, GPT-4, and LangChain agents for live sentiment augmentation.
- Created SDK and open API spec; deployed dashboard in Next.js + WebRTC with real-time charts using Recharts.
- Led full-stack architecture including Dockerized backend, CI/CD workflows, LLM fine-tuning, and synthetic data generation.

Software Engineer at Uniphore, Hybrid
June 2021 — October 2024

- Developed modular UI components and dynamic pages for agent-assist tools using **Next**.js, **React**, and **GraphQL**, integrating live ASR and voice analytics.
- Engineered real-time sentiment visualization layers with phonetic and lexical embeddings for enhanced client-side rendering.
- Streamlined SSR/ISR logic for data-heavy voice sessions, improving initial load times by 35%.
- Modernized analytics pipelines to support LLM-driven conversational interfaces, implementing deep tone-aware inference.
- Built internal analytics tools with **NumPy**, **Pandas**, **Jupyter**, and Python for behavioral insights.

AI Associate at Wave Computing, Campbell, CA
June 2018 — September 2020

- Built custom TensorFlow ops to simulate neural network execution on the DPU Compiler, enabling early validation of operator behavior and alignment with hardware constraints.
- Developed and optimized neural networks for audio and vision classification using PyTorch.
- Designed and maintained an interactive visualization tool for error tracing on Wave's proprietary DPU-based compiler.
- Developed full-stack admin tools with Node.js, React, and GraphQL for secure login and data transfer to AWS S3, deployed via Docker containers. Reduced API payload size by 40% and streamlined file management workflows.

EDUCATION

BS, University of California, Los Angeles

September 2017 — December 2020

Linguistics and Computer Science

PROJECTS

ImageCraftAI

January 2025 — July 2025

Batch image generation platform for ecommerce. Takes a product catalogue and generates images for each product using Stable Diffusion. Results can be fine-tuned using sample images, text prompt or speech to text.

VoiceSentiment

April 2025 — July 2025

Built a real-time pipeline using Pipecat and Cartesia Voice API to analyze sentiment during live support calls. Orchestrated ASR, tone detection, and LLM-based intent tagging with low latency. Improved agent responsiveness and post-call insights. Stack: Node.js, Python, WebSockets.

InterviewAssistant

July 2025 — July 2025

Stack: OpenAI GPT-4 streaming, Whisper, ElevenLabs, WebSockets, Node.js.