

MANUEL GOZZI

+1 831-238-5693 | mannygozzi@protonmail.com | manuelgozzi.com | github.com/MannyGozzi

SUMMARY

VP of Engineering with deep experience architecting multi-tenant AI platforms from zero to production for enterprise clients. Founding engineer behind a TypeScript-based agent platform spanning multi-agent orchestration, tool-calling infrastructure, RAG pipelines, real-time voice agents, and embeddable AI products. Lead engineering hiring, mentorship, and architectural direction across the stack.

TECHNICAL SKILLS

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

Frontend: React, Next.js, React Native, Expo, Tailwind CSS, shadcn/ui, Radix UI, Zustand, Three.js

Backend: Bun, Node.js, Hono, Express, PostgreSQL, Drizzle ORM, Redis, BullMQ, pgvector, MongoDB

AI/ML: Vercel AI SDK, OpenAI, Anthropic, Gemini, xAI, MCP, RAG pipelines, WebSockets, WebRTC, PyTorch

DevOps/Cloud: AWS, Azure, Docker, Kubernetes, Terraform, GitHub Actions, Turborepo, Nx, Linux, Git

EXPERIENCE

AgentPress | VP of Engineering | Remote | Feb 2026 - Present

- Architected a high-scale Bun monorepo and multi-tenant AI platform supporting text and real-time voice APIs using Vercel AI SDK, Hono, and WebSockets.
- Engineered a streamlined agent orchestration system for quick AI agent creation and multi-agent coordination with tool execution, reducing agent deployment time by 95%.
- Designed a chat-driven UI builder with a full in-browser code editor, enabling non-technical users to generate, preview, and deploy custom interface components without engineering involvement.
- Implemented embeddable IIFE widgets for client chat systems with extensive analytics functionality to better understand user engagement and drive iterative performance wins.
- Built a dynamic memory and persona system with cross-thread semantic search, enabling agents to recall prior conversations and adapt interactions to each user.
- Engineered real-time agent monitoring and async tool execution over WebSockets, paired with a performance evaluation engine for continuous quality measurement.
- Architected hybrid authentication (Clerk + in-house) and multi-tenant data isolation via Postgres Row Level Security; extended the platform with MCP servers, public SDKs, and webhooks for deep third-party automation.
- Hired, onboarded, and mentored engineers; built the technical interview process and codified engineering best practices that improved velocity and system stability.

AgentPress | Fullstack Generative AI Engineer | Jul 2024 - Jan 2026

- Founding engineer on AgentPress; built the initial Bun + Hono + Vite monorepo and shipped the first production version of the multi-tenant AI platform.
- Built the original RAG pipeline using PostgreSQL + pgvector with chunking, embedding, and retrieval workflows that became the foundation for all knowledge-base features.
- Designed and shipped the first agent orchestration loop with tool-calling, streaming responses, voice mode tool parity, and multi-turn state management on top of the Vercel AI SDK.
- Shipped the embeddable chat widget for enterprise clients, including shadow-DOM/iframe isolation, CSP workarounds, and a Discourse community forum integration.

Layer3 | Full Stack Developer | Hybrid | Jun 2023 - Jan 2024

- Revitalized gatekeeper.com's user interface using Tailwind CSS, React, TypeScript, and Styled Components, improving user engagement and page load performance.
- Integrated MongoDB, Express, and Redis backend to facilitate interactions with digital wallets and deploy verifiable contracts on blockchain infrastructure.

Monterey Bay Aquarium Research Institute | Machine Learning Intern | Hybrid | Jun 2022 - Aug 2022

- Optimized a PyTorch/YOLOv5 ML pipeline for benthic deep-sea imagery tracking on AWS SageMaker; co-authored research paper on hyperparameter evolution.

PROJECTS

- **NeetSeat** - Ticket Selling Platform | React Native (iOS/Android/Web), TypeScript, Kubernetes, Azure Kubernetes Service. Full stack cross-platform ticketing solution with auto-scaling Kubernetes clusters built and deployed onto Azure via GitHub Actions. MongoDB replica sets for data redundancy.
- **PrepHaven** - AI IELTS English Practice Platform | \$5k+ Revenue | Next.js, Firebase, OpenAI, Stripe Payments
- **SweatSink** - AI Fitness App | React Native (iOS/Android), TypeScript, MongoDB, OpenAI, Twilio Messaging

EDUCATION

California Polytechnic State University, San Luis Obispo | 2022 - 2024 | BS in Computer Science | GPA 3.87 | Magna Cum Laude

Monterey Peninsula College | 2020 - 2022 | AS in Computer Science; AS in Physics; AS in Mathematics | GPA 4.0