

# Vir Khanna

[vir.codes](https://vir.codes) | [vir@saidoen.com](mailto:vir@saidoen.com) | [linkedin.com/in/virkhanna](https://linkedin.com/in/virkhanna) | [github.com/v-khanna](https://github.com/v-khanna) | +1(650)-515-6435

## SUMMARY

---

Forward-deployed engineer who takes AI agents from prototype to production and keeps them reliable. I build them, then embed with the customer to test them against real usage, catch problems early, and ship fixes fast. Have shipped this way in healthcare and across fast-moving startups, using AI tools to move quickly.

## EDUCATION

---

**University of California, Davis**

*B.S. Computational Cognitive Science (Data & AI), Minor in Philosophy (AI Ethics)*

**September 2025**

*GPA: 3.8*

## EXPERIENCE

---

**Nudge AI - Healthcare AI** | [link](#)

*Agent Engineer*

**January 2026 – Present**

*San Francisco, California*

- I build the AI voice agents that answer phones for healthcare clinics, owning each from call-flow design through live deployment on real patient lines and onsite tuning with the customer until it holds in production
- Implemented 15+ HIPAA-aware AI receptionists across multiple states, owning intake and scheduling end to end
- Each books appointments, answers patient questions, verifies insurance, and routes urgent calls with no hold
- Built a fleet of internal agents that took over marketing and sales tasks, saving 6 engineering hours a week

**Blendid AI - Food Robotics** | [link](#)

*Technical Support Engineer*

**June 2025 – October 2025**

*Sunnyvale, California*

- Owned the end-to-end launch of automated refund flows, cutting ticket-resolution latency 72% across 15 sites
- Orchestrated a B2B inventory alerting platform that aligned cross-functional ops for 99% recipe availability

**Everise - BPO, Customer Experience** | [link](#)

*Software Engineer, Voice AI*

**June 2024 – December 2024**

*Plantation, Florida*

- Rolled out Retell AI voice agents across 4 global IT centers, automating 70% of internal IT call resolutions
- Authored prompt and tool-calling logic for IT support agents over internal knowledge bases for live retrieval

**CK Birla Group - Healthcare & IVF** | [link](#)

*Software Engineer, AI*

**June 2023 – December 2023**

*Gurugram, Haryana*

- Integrated LLM-powered chatbots via LangChain and Flask for internal Slack-based question answering workflows
- Minimized API latency by 25% for employee queries by optimizing Flask and Node backends with data extraction

## PROJECTS

---

**Bolo - neural TTS engine, from-scratch MLX-Swift port** | *Swift, MLX-Swift, SwiftUI, Groq* | [link](#)

- Ported the Chatterbox-Turbo neural TTS engine from Python/MLX into native MLX-Swift (T3, S3Gen, vocoder)
- Shipped a public MIT macOS menu-bar app: Groq read-aloud with system-wide hotkeys and Accessibility capture
- Validated the port with six numerical parity gates; tuned for Apple Silicon via 4-bit quantization, KV cache

**Brain - Persistent memory pool for stateless AI agents** | *Agent Skills, MCP, Obsidian, Telegram Bot* | [link](#)

- Engineered a 3-layer memory pool (capture, curated wiki, schema contract) so agents resume with full context
- Automated a Telegram-to-vault pipeline (VPS, git, a 5-skill triage, entity-resolved wiki) agents query live

**The Modelling Directory - full-stack web app** | *Django, PostgreSQL, Cloudflare R2, Render* | [link](#)

- Built and shipped a production-grade full-stack app end-to-end and launched it solo on a live public domain
- Designed the data model, the browse-and-search experience, and the image pipeline behind the public directory

**Echo Journal - AI voice journaling iOS app** | *SwiftUI, FastAPI, OpenAI Realtime API, PostgreSQL* | [link](#)

- Crafted an iOS voice-journaling app with full-duplex speech over the OpenAI Realtime API and FastAPI backend
- Modeled journaling sessions and transcripts in PostgreSQL behind a SwiftUI, voice-first capture-and-play UI

## TECHNICAL SKILLS

---

**Languages:** Python, TypeScript, Swift, SQL, Bash

**ML & AI:** LLM apps, AI agents, prompt engineering, RAG, evals, voice and conversational AI, NLP

**Technologies:** React, Next.js, Django, FastAPI, Node.js, PostgreSQL, Redis, WebSockets

**Tools & Infra:** Git, Docker, AWS, GCP, REST APIs, CI/CD

**Product:** A/B Testing, Product Roadmapping, Telemetry, UX, PRDs, Technical [Writing](#)

**Coursework:** Algorithms, Data Structures, Operating Systems, Machine Learning, AI, Databases, Computer Networks