# Harisankar R N R

**Portfolio:** harisankar.me ↗

Email: harisankarrnr@gmail.com
Contact: +919384558750
Linked In
Leetcode

## About Me

Lead AI Engineer with a B.Tech in Computer Science (AI & Robotics) who designs, builds and runs AI products end to end, from intelligent agents and voice assistants to the backend, cloud infrastructure and deployment pipelines behind them. At Sentellent, leads engineering on an AI-powered learning platform used in production by colleges, a platform for orchestrating fleets of AI coding agents, and a proactive personal AI assistant; as a freelancer, built the AI planning and voice features of a US financial-planning app. Full-stack across Python, TypeScript, React and the cloud (AWS, Google Cloud), and moves fast with AI coding tools while keeping quality high through automated testing.

## Education

- **Vellore Institute of Technology – VIT University, Chennai** — Tamil Nadu, India
  Bachelor of Technology - Computer Science with spl. in AI and Robotics; CGPA: 8.34 — 2021 - 2025

## Skills

- **Languages** — Python, TypeScript, JavaScript, SQL, C++, C, Java, HTML, CSS
- **Generative AI** — LangGraph, LangChain, LlamaIndex, Pydantic AI, Google ADK, OpenAI SDK, Vercel AI SDK, Gemini Live API, MCP, A2A, RAG (pgvector), Claude Code (agent orchestration), n8n, CrewAI
- **ML / DL Libraries** — PyTorch, TensorFlow, Keras, Scikit-learn, NumPy, Pandas, Matplotlib, NLTK, OpenCV, Whisper, Docling, HtDemucs, Pyannote, CTC-forced Aligner
- **Backend** — FastAPI, Flask, Celery, SQLAlchemy, Alembic, Node.js, Express, Hono, Drizzle ORM, Supabase Edge Functions (Deno)
- **Frontend & Mobile** — React, Next.js, TypeScript, Vite, Tailwind CSS, shadcn/Radix, TanStack Query, React Native (Expo), Chrome Extensions (MV3)
- **Cloud Platforms** — AWS (ECS Fargate, EC2, Lambda, Step Functions, API Gateway, SQS, S3, CloudFront, Cognito, SES, RDS/Aurora, DynamoDB), Google Cloud (Cloud Run, Cloud SQL, Vertex AI, Pub/Sub, TensorFlow Serving), Supabase, RunPod
- **Databases** — PostgreSQL (pgvector, row-level security), DynamoDB, Aurora, MongoDB, Redis, SQL Server
- **DevOps & Testing** — Terraform, GitHub Actions, Docker, Kubernetes, RunPod GPU containerization, Pytest, Vitest, Playwright
- **Integrations** — Stripe, Plaid, Canvas LMS (REST API, LTI 1.3), Tableau Cloud, OpenSign, Jira, Google Workspace APIs
- **Design** — Figma (UI/UX), Canva, Mermaid

## Experience

### Lead AI Engineer – Full Time
**Sentellent** — **California, USA (Remote)** — **March 2025 – Present**

- **LXP – AI Learning Experience Platform:** Lead the engineering team (top contributor, 550+ merged PRs) for a multi-tenant LMS where faculty generate courses with AI, sync them with **Canvas** and run hands-on labs; live in production.
- Rebuilt AI course generation as a **LangGraph** planner feeding a **Celery/SQS** pipeline that writes, reviews and repairs each unit, cutting an 82-item course from ~146 to ~26 minutes (~$1.32 per course) and failed quality checks from 46 to 16.
- Built sandboxed student labs (browser **VS Code**, **MySQL/Postgres** GUIs and embedded **Tableau Cloud** on **ECS Fargate**) with terminal capture, checkpoints, instructor review and AI grading submitted back to **Canvas**; 750+ production lab sessions.
- Integrated **Canvas LMS** (two-way sync, **LTI 1.3**) and **Stripe** billing, unified 81 AI call sites across **OpenAI**, **Gemini** and **Anthropic** into one model registry, and cut CI from 135 to 35 billed minutes and a dashboard query from 4 min to 28 ms.
- **Sentellent Agent:** Architect and lead a team of 3 rebuilding an agent platform (**Pydantic AI**, **Gemini/OpenAI**) into a proactive personal agent: user-held encrypted vault, 8-layer memory, **SQS/EventBridge** workers.
- Own the memory and signal-inbox workstreams: a cross-channel people graph, context builder, triage of email, Slack and WhatsApp, drafting in the user's voice, approval-gated sends and **MCP** connectors secured with **OAuth/PKCE**.
- **Other Products:** Built document intelligence (**Docling** parsing, document maps, agentic retrieval) for Theuth, an AI writing tutor for colleges; hybrid **pgvector** + natural-language AI search and **OpenSign** contract flows for the Innovation Marketplace (live on **ECS Fargate**); and serverless budget and cloud-cost tracking tools.
- **Personal AI Agents:** Led the AI-Assistant, a **LangGraph** swarm of Gmail, Calendar, Jira and Drive agents on AWS.

### Freelance Full-Stack AI Engineer
**EAL Plan** — **USA (Remote)** — **May 2025 – Present**

- Lead engineer on ealplan.com, a consumer financial-planning app for a US registered investment adviser covering retirement, goals, tax, insurance and real estate, with AI chat and voice assistants.
- Built a real-time voice assistant on **Gemini Live** via an edge **WebSocket** proxy (~0.5–1 s latency) and a 21-tool chat agent; raised topic-navigation accuracy from 0/10 to 70/71 (chat) and 1/10 to 67/72 (voice).
- Engineered the planning core: a regime-switching Monte Carlo engine with per-year tax, and a 2026 federal + 51-state tax engine (Social Security, IRMAA, RMDs, Roth rules) verified to the dollar against IRS rules on 8 test households.
- Added **Plaid** account linking and **Stripe** subscriptions (Free/Pro, trials, usage quotas), moved hosting to **AWS S3** + **CloudFront** with **Terraform**, and built **GitHub Actions** CI/CD for **Supabase** migrations, edge functions and the frontend.
- **Tech-Stack:** TypeScript, React, Supabase (Postgres, Deno Edge Functions), Gemini (Flash, Live), Vercel AI SDK, Plaid, Stripe, AWS (S3, CloudFront), Terraform, GitHub Actions, Vitest, Playwright

### Generative AI Applications Developer Intern
**Sentellent** — **California, USA (Remote)** — **Jan 2025 – March 2025**

- Led the core logic pipeline integration for jammin.ai, an innovative AI application for generating accurate, word-wise synchronized lyrics and chords for YouTube songs.
- Integrated and optimized open-source audio processing models (**HtDemucs**, **Pyannote VAD**, **CTC-forced aligner**, **Chordino**), leading to a 75% increase in overall accuracy and precision for lyric and timestamp generation.
- Orchestrated the MLOps pipeline on **AWS Step Functions** with parallel **Lambda** functions, deploying demanding models on **RunPod** GPU instances to accelerate output generation by 90%.
- Deployed the production **Python FastAPI** backend, integrating the AI core logic with the web application.

## Projects

### AI Code Manager – Control Plane for Claude Code Agent Fleets

- Took over an internal Sentellent prototype and became its owner (87% of commits, 56 of 57 PRs): a self-hosted web + mobile control plane for running many **Claude Code** agent sessions on shared cloud hosts, with live **tmux** terminals in the browser.
- Isolated each user's agents as a separate Unix user (own home, **tmux**, VNC display, **rootless Docker**) and built router → orchestrator → worker agent hierarchies coordinated over a custom **MCP** server, with Stop-hook reporting and tool-approval escalation.
- Shipped an **Expo** iOS/Android app with over-the-air updates, **Cognito** email-OTP MFA and a session-recorded break-glass admin console; **Terraform** + **GitHub Actions** deploys to **EC2** over **SSM** (median 3.5 min), backed by 1,400+ automated tests.
- **Tech-Stack:** TypeScript, Node.js, Express, WebSockets, node-pty, tmux, xterm.js, React 19, Tailwind CSS, Expo / React Native, Claude Code (hooks, MCP), AWS (EC2, Cognito, SES, S3, DynamoDB, SSM), Terraform, GitHub Actions, Vitest, Playwright

---

### LinkLoom – AI Second Brain for Saved Links

- Built a personal AI knowledge app: paste any link (site, YouTube, Instagram) and **Gemini** extracts, summarizes and auto-categorizes it; find it again by keyword, by voice, or by describing it in plain words.
- Added RAG chat over link groups and the whole library on **pgvector** embeddings (**Vertex AI**), a reader with highlights, smart collections, Pocket/Raindrop import and **Stripe** Free/Pro billing with usage quotas.
- One **Hono** REST API serves a **Next.js** web app, an MV3 browser extension and an **Expo** mobile app; packaged as one **Docker** image (API + worker) for **Cloud Run** and **Cloud SQL**, with **Terraform** and keyless **GitHub Actions** CI/CD.
- **Tech-Stack:** TypeScript, Hono, Node.js, Drizzle ORM, PostgreSQL + pgvector, Gemini (Vertex AI), Firebase Auth, Next.js, React, Expo / React Native, Chrome Extension (MV3), Stripe, Docker, Cloud Run, Cloud SQL, Cloud Tasks, Terraform, GitHub Actions

---

### Deep Learning for Medical and Agricultural Imaging

- **DermMultiNet (Skin Lesion Classification):** Designed a hybrid deep-learning model integrating Dilated Residual Blocks, Deformable Separable Depth Blocks, and Efficient Multi-Frequency Transformer Blocks, achieving 84.03% accuracy on the HAM10000 dataset for skin lesion classification. Published as IEEE Conference paper.
- **Leaf Disease Identification:** Developed a CNN-based leaf disease identifier, integrated into a web application with the **ImageGenerator API** for preprocessing and deployed on **Google Cloud Platform** using **TensorFlow Serving**.
- **Tech-Stack:** Python, PyTorch, TensorFlow, NumPy, OpenCV, React, FastAPI, GCP, TensorFlow Serving, ImageGenerator API, HAM10000 Dataset

---