Sunny Vishwakarma

Software Engineer | Applied AI Engineer
Mumbai, India | Open to remote roles | +91-93729-78703 | Sunny.work70@gmail.com
sunny-v-ebon.vercel.app | github.com/sunnyaiml | linkedin.com/in/sk-tech | dev.to/sunnyaiml | contra.com/sunny_vishwakarma

Professional Summary

Software engineer building production LLM features and multi-tenant backends in Python, TypeScript, Node.js, Laravel, and AWS. Builds tool-calling assistants, structured PDF extraction, evaluation harnesses, and fail-closed guardrails from client requirements through deployment.

Technical Skills

Languages: Python, TypeScript, JavaScript, PHP, SQL
AI / LLM: OpenAI API, Anthropic Claude API, Google Gemini, Ollama, LLM tool calling (function calling), AI agents, structured outputs, prompt engineering, prompt caching, LLM evaluation, guardrails, RAG, LangChain, FAISS vector search, voice AI
Backend & Frontend: Node.js, NestJS, FastAPI, Laravel, React, Next.js, REST APIs, WebSockets, Zod
System Design: multi-tenant isolation, PostgreSQL RLS, idempotency, concurrency control, transactional outbox, serverless
Data: PostgreSQL, MySQL, Redis, DynamoDB, Prisma, Supabase
Cloud & Testing: AWS (Lambda, SAM, S3, SES, EventBridge), Docker, Git, GitHub Actions, CI/CD, PHPUnit, Jest, Playwright

Professional Experience

ElevateCode IT Services Remote
Generative AI Developer *Mar 2026 - Present*

- Built a role-scoped AI assistant (Laravel, React, TypeScript, Claude and OpenAI tool calling, MySQL, Redis) for a B2B telecom platform (116 customers, 4,354 contracts); the LLM sees only summaries and dataset handles, never raw rows.
- Improved the assistant pass rate from 60/87 (69%) to 79/87 (91%) on a hand-graded evaluation; traced 16 incorrect answers to 9 root causes; all 21/21 confidentiality and adversarial cases passed.
- Added fail-closed guardrails: allow-listed tool outputs, a forbidden-field scanner, and reply-figure checks against tool results; backed by 210 PHPUnit tests covering tenant isolation and cross-principal access.
- Extended a serverless invoice engine (16 AWS Lambdas, Node.js, DynamoDB, S3, SES, EventBridge) with OpenAI structured PDF extraction verified against printed text and a duplicate-charge gate (96 Jest files).
- Translated client review feedback into assistant rules and built a monthly unit-cost model (about 1,740 invoice emails/month; AI parsing $55 to $70/month with prompt caching).

Solution Planets Remote
AI Developer & Systems Owner *Oct 2025 - Feb 2026*

- Built CoachAI, a local voice AI tutor (Python, FastAPI, Next.js, TypeScript, Docker Compose) chaining Moonshine STT, Ollama (DeepSeek-R1 14B), and Kokoro streaming TTS with in-memory PyAV audio decoding.
- Built customer-engagement and marketing SaaS apps (FastAPI, Next.js, Google Gemini, OpenAI, Supabase) with contextual reply generation, auth, Stripe and Razorpay payments, and verified webhooks.

Sellergize Web Technology Kalyan, Maharashtra
Service Engineer *Jan 2025 - Sep 2025*

- Built automated web data collection and SQL/JSON validation scripts (Python, Selenium) to verify dataset integrity.
- Built internal REST APIs and resolved daily service tickets, API bugs, and data-rendering issues across web modules.

Projects

Power Dialer - AI Voice Sales Platform | *Laravel, React, TypeScript, Twilio, OpenAI Realtime, Node.js* | *In development*

- Built multi-agent campaign execution with deterministic lead allocation, atomic claims, and leases (1,644 PHPUnit tests).
- Implemented three voice lanes (Twilio Gather, ConversationRelay, OpenAI Realtime) behind one engine with per-span latency tracking, a provider circuit breaker, per-tenant spend caps, and server-authorized call transfer.

CareWell / VitaSetu - Healthcare Platform | *NestJS, Next.js, TypeScript, PostgreSQL, Prisma* | *In development*

- Enforced tenant isolation and consent via 191 PostgreSQL RLS policies and a non-owner DB role boot check; isolation tests cover access before consent, after consent, and after revocation.
- Prevented double-booking with PostgreSQL advisory locks (concurrent race-tested) and built a SKIP LOCKED transactional outbox with retries and dead-letter handling (1,144 Vitest and Playwright CI checks).

Document Q&A with RAG | *Python, LangChain, OpenAI API, FAISS*

- Built a retrieval-augmented document Q&A pipeline using LangChain chunking and a FAISS vector store, grounding model answers in retrieved passages.

Education & Certifications

Bachelor of Engineering, Computer Science (AI & ML) 2020 - 2024
Smt. Indira Gandhi College of Engineering, University of Mumbai | GPA: 7.5/10

IBM Certified Data Science Professional, Nettech India (2025) **Accenture Data Analytics Job Simulation**, Forage (2024)