Nirav Joshi

Full-stack & AI Engineer

India. Open to any city, Pune preferred | niravjoshi3000@gmail.com | +91 94082 68734 | www.niravjoshi.dev

github.com/NiravJoshi33 | linkedin.com/in/niravjoshi3

Summary

Full-stack and AI engineer, and a mechanical design engineer for five years before that. Autonomous mobile robots, AGVs and special-purpose machines at three automation companies.

Skills

Languages: TypeScript/JavaScript, Python, SQL, Rust

Frontend: React, Next.js, Svelte, browser extensions (WXT)

Backend: Node.js, Express, Hono, FastAPI

Data: PostgreSQL, SQLite, Drizzle, Supabase, Redis/BullMQ, Meilisearch, object storage (S3 API on Cloudflare R2, Supabase Storage)

Infrastructure: Docker, AWS (Lambda, ECR, IAM), GitHub Actions CI/CD, Azure (Microsoft Foundry)

Testing: Unit and integration (Vitest, pytest), end-to-end (Playwright)

AI: LLM APIs (OpenAI, Anthropic, Azure OpenAI via Microsoft Foundry), AI agents, tool calling and structured outputs, multi-agent orchestration, MCP servers, RAG, embeddings, vector search (Weaviate)

Evals and LLMOps: Benchmarks on hand-labelled sets (precision, recall), LLM-as-judge, schema-validated outputs, guardrails, human approval before AI output ships, prompt versioning, per-call cost tracking, Prometheus metrics

Web3: Solana, Anchor

Mechanical: CAD (SOLIDWORKS), FEA/CAE (ANSYS), PLM (Teamcenter), GD&T, DFM, MATLAB/Simulink

Experience

Product Lead Engineer, CodeCrunch Techlabs (Ahmedabad) Dec 2023 to Oct 2026

- **BugBuzzer, SaaS website security scanner (Apr 2026 to Oct 2026).** Built and owned the scan engine: 121 security checks across 10 collector types, plus 76 provider modules that verify a leaked credential is live and say what it can reach. Turns "a key is sitting in the JS bundle" into "this key is active and can read your S3 bucket". Run as an external black-box scan across **52,000 sites**, surfacing confirmed vulnerabilities on **900** of them.
- Built the verification layer for an AI pentesting agent. An LLM judge rescores each finding, a sandboxed replay re-runs the exploit and a second agent re-checks it before anything is reported, measured with a hand-labelled precision eval.
- Shipped its subscriptions program end to end: plan entitlements, atomic credit gating, teams and seats, recurring and deploy-triggered scans, alerting, uptime monitoring, hosted status pages.
- **Lead Engineer, BlockchainHQ (Sep 2025 to Apr 2026).** Built and owned the backend of a web3 developer platform, and reviewed the team's code before merge.
- Owned its SEO/GEO programme end to end: JSON-LD structured data, dynamic sitemaps, IndexNow, programmatic landing pages and `llms.txt` for AI-answer retrieval. Planned, implemented, then written up as a developer guide so it wouldn't decay.
- **Developer tooling and protocol work.** *x402 Pocket Nodes*, n8n community nodes that let no-code workflows pay x402 APIs in USDC on Solana: built solo, 22 releases on npm, **3,319 downloads**. *Testship*, a testing CLI for Solana programs. Lead author; published on npm across 16 releases, **2,202 downloads**. *x402test*, a testing framework for x402 payment flows on Solana, built solo and published on npm, **675 downloads**. *SeedPay*, a payment-channel protocol built solo: specification, Anchor program, TypeScript SDK, ~20 documented security invariants, deployed to devnet.
- **SMBmarket (Sep 2024 to Aug 2025).** Solo build of the ETL platform behind a business marketplace: 7 source sites into a 6-stage async worker pipeline, Docker on a VPS, dev and prod CI/CD. **Over 500,000 listings** in production, still serving the live marketplace a year after I handed it over.
- Joined as Software Development Intern (Dec 2023 to Feb 2024), then Software Developer, Product R&D from Mar 2024. As an intern, shipped a domain-expert RAG chat backend (Haystack, Weaviate) and a global event-scraping pipeline.

Product Design Engineer, R&D Mechanical, Addverb Technologies (Noida) Sept 2022 to Dec 2023

- Designed Dynamo-series autonomous mobile robots and their accessories, **500 kg to 2,500 kg payload**, from written specification through prototype assembly and testing.
- Owned the complete design cycle for each robot, including PLM in Teamcenter: check-in/out, BOMs, change orders and releases.

Mechanical Engineer, Beauto Systems (Pune) Aug 2020 to Aug 2022

- Mechanical design of special-purpose machines and robotic automation: a graphite roll slicing machine holding **±0.2 mm**, a wire-mesh strip cutter, a SCARA arm, and a full AGV product line (**25 kg, 50 kg, 1,000 kg and 2,000 kg**), taken from client requirement to delivery.
- SOLIDWORKS and ANSYS; engineering drawings and BOMs; technical support to production and purchase; managed multiple concurrent projects and changed how the team ran its design process.

Mechatronics Engineer, Engineering Services International (Ahmedabad) July 2018 to Aug 2020

- Design and structural simulation of the robots for **India's first robotic museum**. SOLIDWORKS modelling, ANSYS simulation, detailed part and assembly drawings, fabrication-vendor support, assembly, and prototype failure testing and redesign.

Projects

Fineprint Oct 2026 to present

RAG engine over Indian health insurance policy wordings. Answers cite the clause, or say the policy is silent.

- Hybrid retrieval (BM25 and dense embeddings, reciprocal rank fusion) with a re-ranker; a graded eval set showed needed clauses retrieved rising from 19% to 53%. Deployed on AWS Lambda from GitHub Actions through OIDC, no stored AWS keys (container image in ECR, IAM-authenticated URL).

Kairos June 2026 to Aug 2026

AI opportunity intelligence platform. Solo, built in six weeks alongside a full-time job.

- Next.js web app, public REST API and a Python scraper (13 sources, 14 queue-backed workers on arq/Redis) in one monorepo; Meilisearch, Docker on a VPS.

Agent Adapter Apr 2026

Self-hosted runtime that turns any API or MCP server into a paid AI agent. My design; built for the Open Wallet Foundation hackathon, continuing as open source.

- Discovers capabilities from OpenAPI specs and MCP servers and exposes them to an embedded LLM agent as tools. Python runtime with a FastAPI management API and dashboard.

vid.txt May 2026 to Aug 2026

In-browser transcription. Nothing leaves the device.

- Whisper on WebGPU with a WASM fallback, audio extracted with `ffmpeg.wasm`, inference in a Web Worker. Static site, no backend.

Education

U.V. Patel College of Engineering, BTech in Mechanical Engineering (Gujarat) Sept 2014 to June 2018

- Robocon India 2018: led algorithm development for the manual and autonomous robots in MATLAB and deployed to embedded hardware. Team won **Best Usage of MATLAB and Simulink**, MathWorks India.

Certifications

Microsoft Applied Skills: Develop a Generative AI Chat App Using the Microsoft Foundry SDK (Azure), Nov 2026

Also

State of Technology: technology news deduplicated by clustering RSS stories on embeddings and ranked by how many sources corroborate them. FastAPI, Next.js, self-hosted.

AGICitizens agents: three autonomous agents (task requester, Supabase security auditor, PR agent) that bid on and deliver work in an AI-agent marketplace, paid in USDC through on-chain escrow on Solana devnet. Shared Python core with an orchestrator loop, LLM planner and tools.

Sutra: local-first prompt manager with prompt version history, a web app and browser extension on one schema.

Web3 (Solana, Anchor, payment protocols) and application security. Real, but not what I'm applying for.