

Navin Bhaskar

Firmware, backend and AI engineer: From the chip up to the LLM.

Bengaluru, India navin.bhaskar.5@gmail.com navin-portfolio-4g1.pages.dev
linkedin.com/in/navin-bhaskar-5b281219 github.com/navin-bhaskar

SUMMARY

13+ years across firmware, backend and AI systems — from NFC stacks and NAND firmware at NXP and Western Digital, to microservices at Intel, to production LLM agents today.

EXPERIENCE

AI Engineer, Freelance

Jan 2026 – Present

- Built n8n workflows that OCR user-uploaded Google Drive screenshots and index the text for a RAG system.
- Architected and deployed a voice AI agent on ElevenLabs, backed by a custom RAG pipeline.
- Fine-tuned and trained coding AI agents to improve development productivity and accuracy.
- Fine-tuned Llama 3.2 3B with QLoRA on a 100K-example synthetic dataset to extract medical keywords; deployed on Modal with a Gradio UI.

Full Stack Staff Engineer, Incubyte

Jan 2025 – Nov 2025

- Designed a CrewAI multi-agent system that turns medical provider expertise into human-readable summaries.
- Built an agentic chat tool that lets customer support run complex, multi-step operations from natural-language instructions.
- Led a feature team, owning system architecture and the integration of AI components.

Full Stack Engineer, Intel Corporation

Feb 2020 – Nov 2024

- Built a Python interpreter for a domain-specific language with Lark, replacing a legacy .NET solution — +25% validation accuracy and -50% processing time.
- Developed API and executor microservices in Node.js that cut validation time by 30%.
- Built Socket.IO real-time communication and a Python CLI for the client software.

Firmware Engineer, Poynt

Apr 2018 – Dec 2019

- Developed firmware features for a Cortex-M3 payment system that communicates with an Android-based device.
- Built Python tools for log analysis and debugging, speeding up third-party firmware integration.

Firmware Engineer, NXP Semiconductors

Sep 2015 – Apr 2018

- Added context switching and user emulation to a dual-card architecture.
- Automated DESFire test-case generation in Python, cutting time to market by about 20%.

Firmware Engineer, Western Digital (SanDisk)

Jun 2014 – Jul 2015

- Implemented a "Design for Debug" feature for log collection on SanDisk USB drives.
- Developed peak-power programming to reduce power draw while programming 3D NAND.

Design Engineer, NXP Semiconductors

Oct 2012 – Jun 2014

- Implemented ISO 18092 (NFC target) emulation on the LPC1768 MCU.
- Engineered the NFC discovery loop for both initiator and target configurations.

Engineering Intern, NXP Semiconductors

Jun 2012 – Sep 2012

- Added PC/SC reader and logging features to NXP's .NET "RFID Discover" app, improving debugging efficiency by 30%.

SELECTED PROJECTS

MyPersona Forge: An AI digital twin for job seekers. Visitors chat with your profile, and you hear about the questions it can't answer. (React, TypeScript, Cloudflare Workers, Cloudflare D1, Clerk)

RSSJobHunter + ESP32 pager: An agent that scores new job postings against my resume, writes tailored resumes for strong matches, and lights up a physical pager. (Python, LLM agents, PyQt6, RSS, Pushover)

Noterary: Ask plain-English questions about your own videos and recordings. Fully local, with every answer linked to the exact moment it was said. (Python, faster-whisper, Qdrant, Ollama, LM Studio)

UnSilence Studio: A browser-native video toolkit. Remove silence, extract audio, compress, trim, make GIFs and transcribe, all without uploading anything. (FFmpeg (WebAssembly), Whisper, Client-side processing)

MyResumeVault: A free, no-login resume builder. Your data and PDF generation stay entirely in your browser, with on-device AI. (SmolLM2-360M-Instruct, all-MiniLM-L6-v2, In-browser PDF generation)

Medical keyword extraction LLM: Llama 3.2 3B fine-tuned with QLoRA on 100K synthetic examples to pull medical keywords from symptom descriptions. (Llama 3.2 3B, QLoRA, Synthetic data, Modal, Gradio)

SKILLS

Languages	Python, JavaScript, TypeScript, C, Java
Backend & web	Node.js, FastAPI, Django, Flask, React, Redis, Socket.IO
Data	MongoDB, MySQL, Cloudflare D1, Qdrant
Cloud & ops	AWS S3, AWS Lambda, AWS Secrets Manager, Cloudflare Workers, Cloudflare Pages, Docker, Git
AI & LLM	CrewAI, n8n, RAG, OpenAI Responses API, Ollama, LM Studio, QLoRA, Whisper, ElevenLabs, MCP
Embedded	ARM Cortex-M, LPC1768, ESP32, NFC / ISO 18092, DESFire, NAND flash, MQTT

EDUCATION

Master of Science, Manipal Institute of Technology, 2012 · Scaler: Software Development & Problem Solving, 2023