

Tanuj Mehta

Forward Deployed Engineer | Backend + Applied AI

tanujmehta2901@gmail.com · +91 9034346026 · Gurugram, India
tanujmehta.mehtasystems.dev · linkedin.com/in/mehta-tanuj · github.com/mehtatanuj

SUMMARY

Engineer with **4+ years** delivering production systems **inside client environments** across healthcare, fintech, and enterprise IT. I take an ambiguous business problem, integrate with whatever the customer already runs — **EPIC FHIR R4**, payment gateways, **SQL Server**, **vSphere**, Jira — and ship something live. Now doing the same with **LLMs in production**: agentic tool calling, **RAG** pipelines, voice AI, and code-execution sandboxes. Shipped **9 live products end to end, solo**, from schema to deploy.

WHAT I BRING TO A FORWARD DEPLOYED ROLE

- **Client delivery is my background, not a stretch.** Four years across two services organisations, building for named customers against real deadlines and real constraints.
- **Integration into systems I do not own.** EPIC FHIR R4, HL7, multiple payment gateways, SQL Server to PostgreSQL, VMware vSphere, Jira and Bitbucket REST APIs.
- **Solo end-to-end delivery.** Nine products live in production, built without a team: backend, frontend, infra, deploy, monitoring.
- **Applied LLM work that runs in front of users**, not in a notebook — tool-calling agents, RAG, Model Context Protocol, offline inference, voice.
- **Constraint-driven engineering.** Offline-first when cloud cost was the blocker. Local LLM inference when data could not leave. Judge0 sandboxing when untrusted code had to run.

SKILLS

Languages: Python, Rust, TypeScript, JavaScript, Java, C++

Backend: FastAPI, Actix-web, Django, Flask, Spring Boot, Node.js, Microservices, System Design

Applied AI: LLM Integration, RAG, AI Agents, Tool Calling, Model Context Protocol (MCP), Embeddings & Semantic Search, Vector Databases (pgvector), Local LLM Inference (Ollama), Voice AI (STT/TTS)

Enterprise integration: HL7 FHIR (EPIC), Payment Gateways (Razorpay), REST API automation, ETL, Code Execution Sandboxing (Judge0)

Data: PostgreSQL, MySQL, MongoDB, Redis, Supabase, Kafka, RabbitMQ

Cloud & delivery: AWS (Lambda, S3, RDS, VPC, IAM), Docker, Kubernetes, GitHub Actions, CI/CD, Playwright

Frontend: React, Next.js

PROFESSIONAL EXPERIENCE

Senior Software Engineer · AI Engineer

Jul 2025 – Present

Kellton

Gurugram

- Delivered a production **ETL** system integrating **SQL Server** and **PostgreSQL** across **API-to-API** and **database-to-database** boundaries, processing thousands of HR records per day with automated scheduling, encryption, and error recovery.
- Built an **AI voice assistant (Gemini 2.0 + offline LLM)** giving hands-free dashboard control at **90%+** accuracy. Re-architected offline-first once cloud cost became the customer constraint, cutting it **60%**.
- Shipped a **natural-language search** layer translating voice commands into database queries across **50+** query types, so non-technical users could reach the data without SQL.
- Built a real-time monitoring dashboard with **25+** custom error types, alerting, and self-serve recovery tools — cutting troubleshooting time **70%** and holding **99.9%** uptime.

Software Engineer

Jun 2023 – Jun 2025

Webcubator Technologies

Remote, India

- **Patient Health Platform (Healthcare).** Integrated **EPIC FHIR R4** APIs into a **FastAPI** platform, built an **NLP** chatbot at **95%+** intent accuracy, and designed **AWS Lambda** pipelines handling **10K+** monthly events — enterprise healthcare interop against a system I did not control.

- **Rocco Lending & Payments (Fintech).** Integrated multiple **payment gateways**, developed loan-qualification APIs in **Rust** and **Python**, and launched subscription features increasing revenue **30%**.
- **Health Data Aggregation (Healthcare).** Unified fitness data streams at **50K+** records/day, built dashboards improving engagement **40%**, and secured user documents in encrypted **S3** with automated PDF delivery.

Software Engineer

Crest Data Systems

Jan 2022 – May 2023

Pune, India

- Automated end-to-end **Jira** and **Bitbucket** ticket workflows using **REST APIs**, reducing manual effort **60%** and accelerating change tracking across engineering teams.
- Built a **Python FastAPI** service orchestrating the full VM lifecycle for **50+ vSphere** instances, cutting provisioning time **80%** and enabling audit-compliant infrastructure operations.
- Designed reusable workflows for code repository setup and Jira dashboard generation, slashing developer onboarding time **30%**.

PRODUCTS SHIPPED END TO END

Nine live products, built solo — the same loop a forward deployed engineer runs: scope it, build it, deploy it, keep it up.

DevInterviewMaster

devinterviewmaster.com

Rust, Actix-web, SQLx, PostgreSQL, Redis, JWT, React

Rust/Actix-web API serving **10,000+** questions at **5K+ RPS** with sub-10ms p99 latency (async **SQLx**, **JWT**, **Redis**).
LLM-driven content generation pipeline.

ArchitectIQ

architectiq.com

Next.js (API routes), Supabase, LLM API, Judge0, Razorpay

Next.js API over **Supabase** with a **Judge0** code-execution sandbox, **Razorpay** webhook billing, and an **LLM** voice tutor.

VidyaArena

vidyaarena.org

Spring Boot (Java), PostgreSQL, React (Vite), JWT, Docker

Spring Boot API for a **K-12 EdTech** platform: gamified quiz arenas, **JWT** role-based auth, and a **React (Vite)** client in a **git-submodule** monorepo.

ResumeBanao

resumebanao.com

Next.js, Prisma, PostgreSQL, NextAuth, Puppeteer, LLM API

Resume-builder backend on **Prisma/Postgres** with **NextAuth** OAuth, a **Puppeteer** PDF service, and an **LLM** ATS-scoring engine returning section-level feedback.

JobPortal (Auto-Apply)

jobs.mehtasystems.dev

Node.js, Playwright, LLM API, Queue workers, Fly.io

Fly.io automation service: **Playwright** workers auto-apply across job boards, an **LLM** layer ranks roles, with queue orchestration and status tracking.

EDUCATION

Bachelor of Engineering, D.Y. Patil College of Engineering

Jul 2018 – Jun 2022

ACHIEVEMENTS

- **CAP Distinction – Crest Data Systems (2022):** recognised among the top 5% of interns for technical excellence and delivery in backend automation.
- **Runner-Up – GitHub Hackathon, The Code Geek Fest (2022):** top 2 of 150+ teams; built a real-time code collaboration platform in under 48 hours.
- **Global Rank 2253 – Google Hash Code 2021:** among 100K+ participants in Google's global algorithmic coding competition.
- **Tech Community Leader – D.Y. Patil College (2019–2022):** organised 5+ coding contests and technical workshops, growing club participation 60%.