

Anuj Balmiki

Full Stack Developer

Navi Mumbai, India | +91 77150 31900 | anujbalmiki14@gmail.com | <https://anujbalmiki.in> | <https://linkedin.com/in/anuj-balmiki> | <https://github.com/anujbalmiki>

PROFESSIONAL SUMMARY

Full-stack developer with 3+ years of building and running production systems end to end, from API and database design to deployment and monitoring. Currently lead a small development team. Skilled in Python, Django, FastAPI, React, Docker and Kubernetes, with applied AI work in RAG, MCP servers, text-to-SQL and self-hosted LLMs.

PROFESSIONAL EXPERIENCE

Software Developer, Segoma Imaging Technologies

July 2023 - Present

- Lead the in-house development team as the primary full-stack engineer: plan work, review code and own releases. Own about 80% of a Django/DRF ERP (HR, attendance, accounting, finance, diamond ops, QC, shipments, global reports) used by 500+ people across 11 branches.
- Own the supplier portal end to end, serving 2,000+ diamond suppliers worldwide (bookings, returns, pricing, shipments, real-time tracking): FastAPI + aiomysql on MariaDB, JWT + bcrypt, CSRF, Redis rate limiting, Sentry, Kubernetes auto-scaling. Also run an async FastAPI service on AWS EC2 for global integrators.
- Reduced ERP endpoint response times from tens of seconds to milliseconds (over 99% faster) by fixing severe ORM performance issues.
- Manage 5 production servers (ERP main + backup, DB server with MariaDB, MongoDB and Redis, portal, AWS EC2) behind Nginx, with AWS S3, Cloudinary, GitLab CI/CD and Docker. Maintain a Celery + MongoDB audit log of every API call and data change, and keep Pytest coverage above 85%.
- Build and maintain a text-to-SQL MCP server on a self-hosted LLM (Ollama) that any AI agent can use to query company data in plain English; used inside the ERP today. RAG over the 227-table schema sends the model only the relevant tables, and a SELECT-only guard blocks writes.
- Built invoice validation system for the supplier portal across ~1000 invoices a month from 300+ suppliers: read each invoice (OCR for scanned PDFs), check 12 key fields and every stone against the memo with deterministic rules, and send only mismatches to the team instead of every invoice.

UAT Intern, HDFC Life

December 2022 - March 2023

- Tested backend API responses against business requirements across critical insurance workflows; identified and documented 15+ defects.
- Tracked defect resolution with the engineering team and validated fixes before production sign-off.

Backend & API Security Trainee, CyberSmith Secure

September 2021 - March 2022

- Implemented JWT authentication, HSTS, Content-Security-Policy, and X-Frame-Options headers on REST API backends.
- Ran vulnerability assessments on web application backends; wrote remediation documentation used by the engineering team to patch identified issues.
- Applied secure coding practices (input validation, parameterised queries, rate limiting) that I have used in all production work since.

TECHNICAL SKILLS

- Backend: Python, Django, Django REST Framework, FastAPI, Pydantic, asyncio, Celery, Uvicorn, Gunicorn
- Databases: MySQL, MariaDB, MongoDB, Redis, query optimisation, indexing
- DevOps: Docker, Kubernetes, Nginx, Linux system administration, Git, GitLab CI, CI/CD, AWS EC2, AWS S3
- APIs & Auth: REST, JWT, OAuth-based auth flows, Swagger / OpenAPI, CSRF, rate limiting, secure headers
- Testing & Monitoring: Pytest, Sentry, Django Debug Toolbar, cProfile, structured logging
- AI / ML: LLM integration (Google Gemini, Ollama), RAG, MCP servers, text-to-SQL, tool/function calling, structured outputs, prompt engineering, OCR, AI output evaluation, Claude Code, Cursor AI
- Frontend: Next.js, React.js, TypeScript, JavaScript (ES6+), Jinja2, HTML, CSS

PROJECTS

AI Resume Builder: Buildit · buildit-lime.vercel.app

May 2025 - Present

Stack: Next.js · TypeScript · FastAPI · MongoDB · Google Gemini API · Pydantic · Docker

- Full-stack ATS resume builder with dynamic sections and PDF export. FastAPI backend with Pydantic validation, MongoDB, and JWT auth. Frontend in Next.js with TypeScript on Vercel, backend containerised with Docker.
- Integrated Google Gemini API for JD-based resume rewriting and smart autofill. Wrote the prompt logic from scratch to get consistent, structured output from the model.

Supervision Chart Generator

September 2022 - March 2023

Stack: Django · MySQL · Python

- Built for Jai Hind College to automate exam supervision scheduling across 50+ supervisors. Replaced a fully manual process and cut the workload by 95%. Still running in production.
- Wrote constraint logic in Python to enforce business rules (no double-booking, balanced staff distribution), exposed through a Django admin interface with CSV export.

Home Lab Server

Stack: Linux · Docker · Ollama · Self-hosted services · NAS

- Running local LLMs via Ollama on a personal Linux server. Experimenting with model behaviour, prompt engineering, and self-hosted inference pipelines. Keeps skills sharp on Docker and networking outside work hours.

EDUCATION

BSc in Information Technology, Jai Hind College, Mumbai

2020 - 2023

CGPA: 9.89 / 10