

MANUJ RAI

Full-Stack & AI Engineer

manuj.services@gmail.com | +91 87999 33509 | Ahmedabad, India (open to Pune / Bengaluru)
manuj.online | linkedin.com/in/manujrai | github.com/manuj-rai

PROFESSIONAL SUMMARY

AI engineer who has shipped two internal AI systems into production use by an editorial team: a human-in-the-loop news discovery and drafting agent, and a document-to-article pipeline. Work spans Python and FastAPI services, multi-provider LLM orchestration, and integration inside an existing ASP.NET portal, with the token budgets, security controls, and metrics production requires. Two years at a global B2B textile and fashion platform, 18 months full-time after a 7-month engineering internship, on a foundation of ASP.NET Core, C#, and SQL Server.

TECHNICAL SKILLS

AI and LLM: OpenAI, Azure OpenAI, Claude, Gemini, generative AI, AI agents, agentic workflows, LangChain, RAG, embeddings, vector search, structured JSON output, prompt engineering, text-to-SQL, token budgeting and cost control

Backend: Python, FastAPI, ASP.NET Core, C#, .NET 6, ASP.NET Web Forms, REST API design, Server-Sent Events, SQLAlchemy, Pydantic, async services

Frontend: React, Next.js, TypeScript, JavaScript, Angular, HTML, CSS, Tailwind

Databases: SQL Server, T-SQL, stored procedures, query tuning and indexing, PostgreSQL, Supabase, pgvector, vector databases, Pandas, ETL

Automation and Infrastructure: Playwright, Selenium, IIS (hosting Python services via httpPlatformHandler), Docker, Git, GitHub, Prometheus, pytest, unit and integration testing, Vercel

PROFESSIONAL EXPERIENCE

Software Engineer, Fibre2Fashion

Feb 2025 - Present | Ahmedabad, India

Global B2B textile and fashion platform. Work spans AI systems, backend services, and SQL Server performance across the news, article, and subscription modules.

- Designed and shipped **two internal production AI systems** end to end, from Python and FastAPI agent services through integration inside the company's existing ASP.NET admin portal, covering LLM orchestration, document processing, security controls, and observability. Both are detailed under Projects.
- Removed a manual editorial step from a high-traffic news pipeline with an Azure OpenAI integration in .NET 6 that detects and masks sensitive financial figures in paid submissions before publication.
- Delivered Corporate Team Management for seat-based B2B subscriptions: REST APIs, PostgreSQL and SQL Server schemas, invite and activation flows, and payment workflows.
- Author and tune SQL Server stored procedures for large-scale reads and multi-result-set reports, refactoring legacy procedures to table-valued parameters.
- Build Python automation: Pandas cleaning pipelines, Excel-to-SQL Server import tooling, and Playwright and Selenium scrapers handling authentication, pagination, and JS-rendered content.

Freelance Full-Stack Developer, Independent

Jul 2024 - Present | Remote, alongside full-time role

- Ship production web products on Next.js, React, TypeScript, and Supabase with Razorpay and Stripe payments, deployed on Vercel, Netlify, and Render.

- **Karlooloo:** 20+ browser-first tools on a registry-driven architecture, where one entry makes a tool discoverable across search, sitemap, and structured data; AI features sit behind a validated, rate-limited route.
- **Trendverse, Hair Crezz, Kat-Katha:** a D2C storefront with row-level security and server-verified Razorpay webhooks, a salon booking product, and an NGO giving platform with an idempotent, HMAC-verified donation pipeline.

IT Intern, Fibre2Fashion

Jul 2024 - Jan 2025 | Ahmedabad, India

- Completed structured training in ASP.NET, C#, and advanced SQL Server, covering stored procedures, indexing, and query plans.
- Automated exhibitor and supplier data collection with Python scrapers (Playwright, Requests, Selenium) and Pandas pipelines into SQL Server, and shipped an internal Inquiry Management System in ASP.NET and C# that internal teams still use.

PROJECTS

AI Editorial Platform: News Desk and AI Article Gen | Fibre2Fashion, in production

Stack: Python, FastAPI, OpenAI (pluggable Gemini and Claude), Playwright, SQL Server, SSE, IIS, Prometheus

- **Story discovery agent:** replaced the editorial team's manual morning story hunt at a desk publishing 20-30 textile articles daily, delivering a ranked, de-duplicated shortlist each morning while story selection and review stayed with the editors. Harvests a portal-managed SQL source master (~65 primary and ~75 secondary pages, 17 countries) with Playwright, using RSS and Google News fallbacks, then LLM relevance scoring, source-authority weighting, and primary-source confirmation. Covered by 375 tests across 43 files.
- **Document-to-article pipeline:** cut article preparation from retyping eight derived fields by hand to upload, review, save. A raw Word document becomes a publish-ready package (headline, key takeaways, table of contents, body, SEO metadata, FAQs) plus the DOCX, PDF, and HTM artefacts the public site serves, through a six-phase pipeline with three JSON-mode LLM calls per short article. Editorial rules live as fail-open SQL Server rows, so the desk's guidelines change without a code deploy.
- **Shared LLM and reliability layer:** three providers behind one interface with JSON-mode output, retries, and per-job token budgets, running a cheap model for high-volume discovery and a stronger one for drafting; SSRF-guarded fetching, rate limiting, structured JSON logging, and Prometheus metrics.
- **Enterprise integration:** both services surfaced as pages inside the existing ASP.NET Web Forms portal under its access gate, with FastAPI hosted under IIS via httpPlatformHandler and the token-spending endpoints secured behind a shared secret.

Developer Portfolio with RAG-Grounded AI Assistant | Personal, manuj.online

Stack: Next.js, TypeScript, Supabase (pgvector), OpenAI, Vercel

- Answers only from an embedded corpus of the site's own content, matched with an HNSW cosine index so every answer links back to a real page. Google OAuth with an admin gate enforced in both the application layer and Postgres row-level security.

EDUCATION

Bachelor of Business Administration, Computer Applications

Savitribai Phule Pune University | 2024

CERTIFICATIONS AND TRAINING

Company-Sponsored Python and AI Engineering Program | Fibre2Fashion | 2025

Production Python, LangChain, RAG pipelines, AI agents, and text-to-SQL.