

Pranjul Rathour

GenAI Engineer | LLM & RAG Systems | Fine-Tuning & Agentic AI | Python | AI Product Engineering

Kanpur, India | pranjulrathour41@gmail.com | +91 8467977141 | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

1st – IIT Roorkee Hackathon | 1st – CSJMU Kanpur | Top 3, IIT Kanpur Hackathon | Senior GenAI Engineer, SCULT INDIA

Summary

GenAI engineer and AI product builder with 2+ years shipping production-grade AI systems — Retrieval-Augmented Generation (RAG) pipelines, fine-tuned LLMs, hybrid retrieval, and multi-modal AI applications spanning vision, speech, and OCR — architected end-to-end from data ingestion through deployment. Led engineering for a 14-member team building AI-powered products, pairing deep GenAI/ML engineering with full-stack delivery, while driving measurable business revenue as a startup technical lead.

Technical Skills

- **AI/ML & GenAI:** LLMs (OpenAI, Claude, Gemini, Mistral, Groq/Llama), RAG Pipelines & Hybrid (Dense+Sparse) Retrieval, LLM Fine-Tuning (LoRA/QLoRA via Hugging Face Transformers/PEFT/TRL), Vector Search & Embeddings, Prompt Engineering, LangChain, MCP (Model Context Protocol), OpenRouter, ONNX Runtime, Computer Vision
- **Languages:** Python, JavaScript, TypeScript, HTML5, CSS3
- **Backend:** FastAPI, Node.js, Express.js, REST APIs, WebSockets
- **Databases & Vector Stores:** PostgreSQL, MongoDB, Supabase, Firebase, Redis, FAISS, Qdrant, pgvector
- **Frontend:** React.js, Next.js, Tailwind CSS, Shadcn UI, Redux, Zustand
- **Cloud & Tools:** AWS, Docker, Vercel, Railway, Git, GitHub, CI/CD

Experience

NEXT UPGRAD WEB SOLUTIONS

Feb 2026 – Jul 2026

GenAI Engineer

- Independently designed, built, and shipped 5 production-grade AI applications end-to-end — spanning computer vision, document intelligence, LLM fine-tuning, speech-to-text/OCR, and retrieval-augmented generation (full list under Projects).
- Built a production RAG platform with hybrid (dense + sparse) retrieval, re-ranking, and streaming citation-grounded answers, including a confidence gate that blocks the model from answering on weak evidence.
- Built an end-to-end LLM fine-tuning platform (QLoRA via Hugging Face Transformers/PEFT/TRL) with live training telemetry and side-by-side base-vs-fine-tuned evaluation.
- Deployed and operated all five services on Railway/Vercel with Docker, CI/CD, and observability (structured logging, metrics), achieving production-grade reliability across the portfolio.

SCULT INDIA

2024 – Present

Senior GenAI Engineer

- Built and shipped 5+ paid, production-grade web applications and AI-powered products for clients across industries.
- Led engineering, system architecture, UI/UX, deployment, and client delivery for a cross-functional 14-member team.
- Designed and developed **SCULT Studio CRM**, an internal platform for project, client, and operations management.
- Generated approximately INR 1,00,000 in revenue within the first six months of operation.

TechVerse Enclave

2023 – Present

Founder

- Built and grew a 500+ member developer community; mentored 200+ students in web development and career growth.

Projects

NEXTUPGRAD WEB SOLUTIONS — FEATURED AI PROJECTS

- **FaceVision** – [Demo](#) | [GitHub](#): Browser-only face detection, recognition & liveness-check app (ONNX Runtime Web); FastAPI/PostgreSQL backend stores only vector embeddings, never images.
- **DocuLens AI** – [Demo](#) | [GitHub](#): Groq-LLM document-intelligence platform — extracts, compares & checks compliance on invoices/contracts with schema-validated, deterministic scoring.
- **FineTune Studio** – [Demo](#) | [GitHub](#): End-to-end QLoRA fine-tuning platform (HF Transformers/PEFT/TRL) with live loss streaming and base-vs-fine-tuned evaluation.
- **OCR & Speech Workspace** – [Demo](#) | [GitHub](#): Mistral-AI workspace for concurrent page-batched PDF OCR, live speech-to-text, and document-scoped RAG chat with citations.
- **RAG.NextUpgrad** – [Demo](#) | [GitHub](#): Production RAG platform with hybrid dense + BM25 retrieval, a hallucination-blocking confidence gate, and automatic multi-provider fallback.

INDEPENDENT & CLIENT PROJECTS

- **SCULT INDIA** – [Demo](#) | [Website](#) | [Studio CRM](#): AI-first software agency platform with an internal CRM and workflow management system.
- **Super Digital Library** – [Demo](#) | [Live](#): Digital library platform built with a scalable architecture for organizing and accessing large content collections.
- **Vaidya AI** – [Demo](#) | [GitHub](#): LLM-powered healthcare chatbot delivering conversational patient assistance.
- **Annapurna** – [Demo](#) | [GitHub](#): NGO donation platform with a streamlined giving and outreach flow.
- **Nuclear Pest Goa** – [Demo](#) | [Live](#): Enterprise business website with SEO optimization and a fully responsive UI.
- **SCULT Marketing** – [Demo](#) | [Live](#): Marketing landing platform for agency service offerings and lead generation.
- **Gloss Digital Entertainment** – [Demo](#) | [Live](#): Media platform frontend built with a scalable component architecture.
- **Passionate Records** – [Live](#): Music label website with a modern, responsive design.

Education

Dr. Virendra Swarup Institute of Computer Studies, Kanpur

2023 – 2026

Bachelor of Computer Applications (BCA)