

SANGRAMJEET GHOSH

+91 9059150937 | sangramjeet47@gmail.com | linkedin.com/in/sangramjeetghosh
github.com/JeetInTech | sangramjeet.me | neuroviai.com | [Certificates Portfolio](#)

Professional Summary

Versatile AI & Full-Stack Engineer with hands-on experience building and shipping production-grade SaaS platforms, autonomous AI systems, and content automation pipelines. Designed and launched Neuroviai — a live multi-provider LLM automation SaaS. Built Sebastian, a fully autonomous assistant with real-time voice, vision, OS control, and 110+ API integrations. Applied AI researcher specializing in constraint-driven system design with proven ability to take products from concept to deployment across cloud platforms. Actively seeking opportunities to build impactful, scalable technology.

Education

Bachelor of Technology (B.Tech) in Computer Science and Engineering *Expected June 2026*

Specialization in Artificial Intelligence and Machine Learning

Malla Reddy University, Hyderabad, Telangana

GPA: 8.0 / 10.0

- **Relevant Coursework:** Data Structures & Algorithms, Machine Learning, Deep Learning, Natural Language Processing, Computer Vision, Database Management Systems, Cloud Computing, Statistics & Probability
- **Thesis:** *Design and Implementation of Cost-Constrained AI Automation Systems* (Feb 2026) — Full-stack SaaS platform (Neuroviai) with 15,000+ lines of Python and 12,000+ lines of TypeScript; regulatory-aligned data stewardship (India's DPDPA 2023, EU AI Act).

Technical Skills

Languages	Python, JavaScript, TypeScript, SQL, HTML5, CSS3, Bash
AI / ML	LangChain, LangGraph, Gemini API, OpenAI API, Groq, RAG Pipelines, pgvector, Prompt Engineering, Multi-Agent Systems, Scikit-learn, XGBoost, LightGBM, Pandas, NumPy
Backend	FastAPI, REST APIs, WebSockets, JWT/OAuth 2.0, Webhook Handlers, Microservices Architecture, Async Python
Frontend	React 18, Next.js 14 (App Router), Tailwind CSS, Zustand, ShadCN/UI, Vite, TypeScript
Databases	PostgreSQL, Supabase, pgvector, MongoDB (basics), Redis, SQLAlchemy ORM
DevOps/Cloud	AWS (EC2, S3, IAM), Docker, Docker Compose, GitHub Actions CI/CD, Render, Vercel, Netlify, Nginx, FFmpeg
Tools	Git, GitHub, Postman, VS Code, Playwright, PyAutoGUI

Work Experience

Freelance AI & Full-Stack Developer

Aug 2025 – Present

Remote

- Delivered end-to-end web applications and AI-powered MVPs for 3+ startup founders, handling requirements gathering, architecture design, development, deployment, and post-launch iteration.
- Built SaaS products with subscription billing (Razorpay), role-based access control (JWT/OAuth), and multi-platform API integrations.
- Deployed applications across AWS, Render, Vercel, and Netlify with custom domains, SSL certificates, and CDN configuration.
- Managed LLM provider costs through intelligent quota splitting, response caching, and request batching strategies.
- Created integration guides and handoff documentation enabling clients' in-house teams to maintain and extend delivered products.

Hyderabad

- Developed AI-powered backend features for CraftingBrain (craftingbrain.com), a live content-technology platform, using Python and FastAPI.
- Built and deployed NLP pipelines for real-time text analysis, sentiment scoring, topic extraction, and content optimization in production environments.
- Managed AWS infrastructure: EC2 provisioning, S3 bucket policies, IAM role configuration, and Nginx reverse-proxy setup.
- Containerized Python services using multi-stage Docker builds, reducing image size and cold-start latency.
- Authored CI/CD pipelines with GitHub Actions automating linting, testing, and zero-downtime production deployments.

Projects

Neuroviai — AI-Powered Social Media Automation SaaS

*Mar 2025 – Present**Next.js 14, FastAPI, PostgreSQL, Supabase, Gemini, OpenAI, Razorpay, Docker*neuroviai.com

- Designed and built a full-stack AI SaaS platform from scratch — product design, database schema, API architecture, AI pipelines, payment integration, and multi-platform publishing.
- Architected 4-tier LLM cascade (Gemini → GROQ → HuggingFace) achieving 99.7% text generation reliability; autonomous engine posts across 5 platforms daily, cutting operational effort by 90%.
- Integrated 8+ external APIs (LinkedIn Publishing, Twitter/X, Razorpay, Supabase Auth) with OAuth flows, webhook listeners, error handling, and rate-limit management.
- Designed token-economy billing: per-user credit tracking, usage alerts, low-balance upsell triggers, and plan-upgrade flows.
- Deployed as 4 microservices (Render, Supabase, Netlify, AWS S3) with GitHub Actions CI/CD and engagement analytics dashboard.

Sebastian — Autonomous AI Personal Assistant

*Jan 2026 – Present**Python, FastAPI, PostgreSQL, pgvector, Gemini Live, Groq, Playwright, Edge TTS, PyAutoGUI*

- Architected a 10-phase fully autonomous assistant with 59+ tools across OS control, browser automation, real-time bidirectional audio streaming (Gemini Live), camera/screen vision, and 110+ API integrations.
- Built a multi-layer memory system (episodic, semantic, behavioral) using PostgreSQL + pgvector with cosine similarity search, time-decay importance weighting, and automatic memory deduplication.
- Implemented a 3-tier permission gate (GREEN/YELLOW/RED) with automatic escalation and full audit logging for every tool execution.
- Added Discord voice channel bridge via py-cord with real-time PCM audio resampling between Discord (48kHz stereo) and the Gemini Live pipeline (16kHz mono).
- Exposed a FastAPI server with REST + WebSocket endpoints and a React PWA dashboard for cross-device access.

AI YouTube Shorts Automation Pipeline

*Jan 2026 – Present**Python, Ollama (local LLM), FFmpeg, edge-tts, Pexels API, YouTube Data API v3*[GitHub](#)

- Built a fully automated short-form video pipeline: trend scraping (Reddit, Hacker News, Google Trends) → LLM script generation with quality scoring → voiceover synthesis → stock footage assembly (FFmpeg) → YouTube upload with metadata.
- Engineered Windows Task Scheduler integration for unattended daily posting; multi-channel mode generates 4+ videos per run.
- Implemented a script quality scoring module to filter low-confidence AI scripts before production, with a Streamlit analytics dashboard for tracking upload history and performance metrics.

Neurovia Careers — AI-Powered ATS Resume Builder

Nov 2025 – Present

React 18, TypeScript, FastAPI, Supabase, Groq LLM, Vite

[GitHub](#)

- Built a full-stack ATS resume builder with a weighted scoring engine: keyword match (40%), content completeness (35%), formatting quality (25%) targeting 90%+ ATS compatibility across 15+ job categories.
- Engineered a custom LaTeX export system with 4 professional templates (Simple, ModernCV, Awesome-CV, Deedy) producing clean, recruiter-safe PDFs.
- Integrated Groq LLM API for AI-assisted bullet generation, summary writing, and skill recommendations with Supabase auth, cloud CV storage, and shareable CV links.

NeuroviaI — Multi-Agent AI Data Analytics Platform

Oct 2025 – Present

Python, PySide6 (Qt6), Groq, Gemini, Ollama, Scikit-learn, XGBoost, LightGBM, Matplotlib

- Developed a desktop analytics platform using a 6-agent LLM pipeline (Profiler, Cleaner, Feature Engineer, Model Selector, Visualizer, Orchestrator) each executing think → act → observe loops.
- Benchmarked 11 ML models (Random Forest, XGBoost, LightGBM, SVM, etc.) with cross-validation and automated feature importance analysis.
- Built 20+ chart types (2D + 3D with matplotlib) across 5 themeable styles; implemented LLM provider failover: Groq → Gemini → Ollama for full offline capability.

Virtual Therapist — LangGraph Agentic AI System

Sep – Oct 2025

LangGraph, LangChain, Python, PostgreSQL

[GitHub](#)

- Designed a stateful 3-agent LangGraph system (intake, emotion-detection, therapeutic-response) with deterministic state transitions, structured prompt contracts, and JSON output validation.
- Integrated speech-to-text input and adaptive LLM responses tuned to detected emotional context; persisted conversation memory in PostgreSQL.

Publications & Research

Design and Implementation of a Cost-Constrained Multi-Platform AI Automation System

Independent Researcher — Manuscript in Preparation (Case Study on NeuroviaI)

- Engineered a deterministic four-tier AI fallback strategy (Gemini, GROQ, HuggingFace) achieving 99.7% text generation reliability through provider-level redundancy.
- Documented orchestration-free microservice patterns to maximize free-tier resource allocations across ten independent server processes.
- Formalized 23 distinct engineering trade-offs regarding cold-start mitigation, ephemeral storage, and security in zero-budget environments.

Achievements & Recognition

- **Winner — IDEATHON 2.0** (University-level Ideation Competition): Developed a high-impact AI solution evaluated by industry experts for technical execution, feasibility, and market potential.
- **Participant — METALLOOP 2K23** (National-level Hackathon): Built a functional prototype under intensive time constraints; received official certification for technical execution.
- **Self-Published Author — Amazon KDP:** Authored “*The Future of Touchless Technology*,” managed end-to-end digital publishing lifecycle including automated manuscript formatting and global distribution. amazon.com/dp/B0FBLC1RLT

Certifications

All certificates verified and available at: [Certificate Portfolio \(Google Drive\)](#)

- **Amazon Web Services (AWS)** — Cloud Practitioner Essentials / Foundations
- **Google Cloud** — Cloud Computing Fundamentals
- **Coursera** — Machine Learning Specialization / AI & Data Science tracks
- **Microsoft Learn** — Azure AI Fundamentals / Developer certifications

- **NPTEL** — Programming and Data Science courses (IIT-certified)

Extracurricular & Leadership

- **Open-Source Contributor:** Actively maintain public repositories on GitHub (github.com/JeetInTech) with multiple AI, web, and data projects.
- **Technical Mentor:** Guide junior students in AI/ML concepts, project development, and hackathon preparation.
- **Hackathon Participant:** Regular competitor in national-level hackathons; strong track record of building functional prototypes under 24–48 hour deadlines.
- **Freelance Consultant:** Advise early-stage startup founders on technology stack selection, architecture design, and MVP scoping.

Languages

English (Professional Proficiency) | Hindi (Native) | Bengali (Native)