

Sahil Singh

B.Tech, Mechanical Engineering, IIT Kanpur
CPI: 7.93/10 (2020–2024)

sahil85.10s@gmail.com | +91-9874082672
GitHub: sahilssingh | LinkedIn: sahilssingh

Work Experience

Spine AI (YC S23) | SDE | Remote

Aug'25 – Apr'26

- Architected the **backend orchestration engine** powering Spine Swarm, coordinating 1–10 specialized AI agents/task with **500+ agents running concurrently** for parallel multi-step workflows at production scale.
- Built **intelligent model routing** dynamically selecting across **300+ LLMs** per execution step, processing 106,000+ intelligence nodes cumulatively to reduce single-model bias and improve output reliability across diverse reasoning, coding, and retrieval tasks.
- Designed backend APIs powering **6,400+ active canvas environments** with real-time collaborative agent execution, artifact persistence, branching, and full inspectability of intermediate agent state across concurrent user sessions.
- Went **0 to 1** across full stack (React canvas, backend, DevOps), scaling Spine to **10k+ users** and **151 paying customers** (2.5% conversion) with **zero-downtime blue-green deployments** across AWS/GCP/Azure.

Signi.AI | SDE, Founding Member | Gurgaon

Jun'24 – Jul'25

- Migrated backend services from PM2 to a **containerized ecosystem**, improving scalability and deployment consistency.
- Built a **high-concurrency websocket microservice** handling **10k concurrent requests** and a Redis-powered real-time leaderboard delivering low-latency user-engagement insights across the platform's service usage rankings.
- Integrated **Stripe** for domestic/international payments (GRE product); architected **Datadog APM** with centralized logging across AI webserver, improving real-time monitoring and reducing outage response times via proactive alerting.
- Executed blue-green deployments with Docker/PM2 ensuring zero downtime; led and mentored a team of interns across backend and DevOps, delegating short-term tasks, reviewing code, and driving continuous team development.

Signi.AI | SDE Intern | Remote | PPO Offered

Oct'23 – May'24

- Built backend APIs/websockets in **Django** for AI chat interactions and a **Redis leaderboard** with logarithmic-time access.
- Developed custom notification panel with **FCM + Celery** (bulk, CSV upload, analytics, internal testing); implemented deeplink-based article sharing via Firebase with auto-generated images stored in AWS S3.

GIVA Indiejewels | SDE Intern | Bangalore | PPO Offered

May'23 – Jul'23

- Full-stack work on internal POS tools (**Next.js/Node.js**); shipped mission-critical UI flows to the React Native mobile app.
- Integrated **Shopify** webhooks for real-time order sync and **Razorpay** for secure domestic payments; discovered and patched an indefinite-OTP vulnerability via timeout invalidation to prevent unauthorized access.

Key Projects

JEEenius AI | Founder & Sole Engineer | landing.jeenius.tech

Sep'25 – Mar'26

- Built a **JEE question-prediction engine** using **RAG over a vector-indexed corpus** of 20+ years of PYQ/syllabus data with semantic similarity search and concept-graph traversal — achieving **83% topic coverage** on JEE Main 2026 January session with **63 near-exact question matches** (Physics 92%, Chemistry 87%), ground-truth validated post-exam.
- Architected a **multi-model reasoning stack** routing across GPT-4o, o-series, Gemini 2.5 Pro, and Fathom AI per query type, optimising for quality/latency/cost across **100,000+ practice questions**.
- Shipped the entire product solo — Android app, landing page, AI backend, adaptive study-plan engine, **real-time 1v1 duel system**, image-based doubt solver, and analytics dashboard.
- Grew to **5,000+ users** with a **4.8 Play Store rating** entirely via organic growth; paying customers on a freemium model (Pro Rs.499/mo, Pro+ Rs.999/mo).

AI-based Labor Room Smart Assistant | Google AI4SG '21-'22 | Prof. Vipul Arora, EE

Jan'22 – Jan'23

- Secured **USD 30,000 funding** from Google's AI for Social Good program (out of 1000+ global applicants); built a Flutter + Flask app with ASR for timely data documentation in low-resource labor rooms in Bihar.

Feed Harmony | GDSC IITK, Google Solution Challenge | Overall Manager

Feb'24

- Led Product/Design, Development, and ML teams to ship a mobile app tackling **global hunger with predictive ML models** within a 3-week sprint; contributed core mobile components and drove cross-team integration end-to-end.

Awards & Leadership

- **Founder & Lead, Google Developer Student Club, IIT Kanpur** — grew to **1,000+ members** with **5,000+ event participants**; led a 9-member core team & 50 volunteers across Ideathons, Cloud Jams, and Generative AI workshops.
- **Bronze medal, Inter IIT Tech Meet 10.0** — built ISRO's solar burst identifier tool for X-Ray light curves.
- **Google Code Jam 2021**: Global Rank **794**; **Google Kick Start 2021**: best rank **1544** (56k+ global participants).

Technical Skills

- **Languages**: C++, C, Python, Java, JavaScript, TypeScript, Dart, C#, SQL
- **Backend & Data**: Django, Flask, Node.js, Express, REST, WebSockets, Redis, MongoDB, PostgreSQL, Firestore, Celery
- **Frontend & Mobile**: React, Next.js, React Native, Flutter, Redux, Tailwind, HTML/CSS, D3.js, Chart.js
- **AI/ML**: RAG, Vector DBs, LLM Orchestration, OpenAI/Gemini/Anthropic APIs, Pandas, NumPy, Scikit-learn, ASR
- **DevOps & Integrations**: AWS, GCP, Azure, Docker, CI/CD, Datadog, Firebase, Stripe, Razorpay, Shopify, FCM, Git, Linux