

Suraj Vishwakarma

Fullstack AI Engineer

career.surajv@gmail.com | +91-7488659074 | [linkedin.com/in/surajv354](https://www.linkedin.com/in/surajv354) | github.com/smartcraze | surajv.dev

SUMMARY

Full-Stack Engineer specializing in AI-powered applications, with production experience building LLM evaluation pipelines, agentic workflows, and cloud-native infrastructure on AWS. Shipped an LLM-as-Judge grading system using OpenAI APIs, LangChain, and the Vercel AI SDK, alongside full-stack platforms in React/Next.js, Node.js/Bun, and PostgreSQL/Redis, containerized with Docker and Kubernetes. SIH 2025 Finalist · WarpSpeed Top-20 · 380+ DSA problems solved. Seeking **Fullstack / AI Engineering** roles building production-grade AI products.

SKILLS

- **AI & Agentic:** OpenAI, LangChain, AI SDK, LLM Eval, LLM-as-Judge, RAG, Prompt Engineering, Agentic Pipelines
- **Languages:** JavaScript, TypeScript, Python, C++, Java, Go
- **Frontend:** React.js, Next.js, Tailwind CSS, Redux, REST APIs, GraphQL, WebSockets
- **Backend & Data:** Node.js, Express.js, Bun.js, BullMQ, JWT, OAuth2, PostgreSQL, MongoDB, Redis, Prisma ORM
- **DevOps & Cloud:** AWS (EC2, S3, IAM, ECR, ECS, EKS), Docker, Kubernetes, Terraform, CI/CD, GitHub Actions, Linux, Nginx

EXPERIENCE

Vimaan Robotics Pvt. Ltd. | **DevOps Engineer** | Bengaluru, Karnataka Jul 2026 – Present

- Designed and maintained CI/CD pipelines using GitHub Actions, automating build, test, and deployment workflows for production services.
 - Managed cloud infrastructure on AWS using Docker, EC2, ECR, S3, IAM, and Terraform, improving deployment consistency and reducing manual operational effort.
 - Deployed and monitored containerized applications on Kubernetes, implementing rolling updates, health checks, and automated recovery strategies to improve production reliability.
- Stack:** AWS, Docker, Kubernetes, Terraform, GitHub Actions, Linux, Nginx

Gradified | **AI Backend Engineer Intern** | Hybrid Noida Apr 2026 – Jun 2026

- Architected a production LLM evaluation pipeline for multi-subject exam grading (math, biology, general) with dual-pass LLM-as-Judge architecture, consensus gating, and type-aware rubric scoring — reducing manual grading effort to zero.
 - Engineered a math evaluation module supporting 4+ response formats (MCQ, short answer, equations, step-wise proofs), handling numerical precision and algebraic equivalence detection via OpenAI models.
 - Scaled concurrent grading throughput 3× using BullMQ-based async job queues with deterministic temperature settings, ensuring consistent, reproducible LLM scoring across high-volume runs.
- Stack:** Bun.js, Vercel AI SDK, OpenAI API, BullMQ, PostgreSQL, AWS S3

PROJECTS

RapidServe — Vercel-like Full-Stack Deployment Platform | [GitHub](#) | [Live](#) Dec 2025 – Mar 2026

- Designed a 5-service monorepo (API, Builder, Proxy, WS, Frontend) where the Builder clones a GitHub repo, installs dependencies, and runs the production build inside an isolated Docker container, uploading output to AWS S3.
 - Engineered a custom reverse proxy that routes subdomain requests to the correct S3 path, replicating Vercel's subdomain-per-deployment model without a CDN dependency.
 - Implemented real-time build log streaming via a scoped Redis Pub/Sub channel consumed over WebSocket, so users see live build output as the deployment runs.
- Stack:** Node.js, Express, React, PostgreSQL, Redis Pub/Sub, Docker, AWS S3, Terraform

SketchKaro — Real-Time Collaborative Whiteboard | [GitHub](#) | [Live](#) Nov 2025 – Dec 2025

- Architected a 3-app Turborepo monorepo spanning a Next.js frontend, an Express REST backend, and a Bun native WebSocket server, with shared UI (ShadCN), Prisma, and Zod packages reused across all apps.
 - Built the real-time drawing engine on the HTML5 Canvas API, broadcasting stroke events over Bun's native WebSocket server to all room participants with <100ms sync latency.
 - Containerized all three services with Docker Compose and provisioned AWS infrastructure via Terraform, enabling one-command environment setup with full dev/prod config parity.
- Stack:** Next.js, Bun.js, Express, WebSocket, JWT, PostgreSQL, Prisma, Docker, Terraform

ACHIEVEMENTS & CERTIFICATIONS

- **SIH 2025 Finalist** — Selected among 2000+ teams nationwide at Smart India Hackathon Jan 2025
- **380+ DSA Problems** — Solved on [LeetCode](#), Codeforces, and GFG May 2026
- **AI Foundations Associate** — [Oracle](#) Apr 2026
- **Full-Stack Web Development** — [100xDevs](#) Aug 2025
- **Generative AI with LangChain & OpenAI** — [Udemy](#) Oct 2024

EDUCATION

Lovely Professional University Phagwara, Punjab
B.Tech in Computer Science and Engineering | **CGPA: 7.66** Aug 2023 – May 2027