

SOHAIL SHAIK

Full-Stack AI Engineer — AI Applications · Backend · LLMs · Distributed Systems

+91 96426 39515 — s.suhail9849@gmail.com — github.com/Sohail-9 — LinkedIn

PROFESSIONAL SUMMARY

Full-Stack AI Engineer with 4+ years of experience building and scaling production AI applications, backend systems, and modern web platforms. Experienced in designing end-to-end solutions spanning React/Next.js frontends, Node.js and Python microservices, PostgreSQL-backed architectures, LLM orchestration, RAG pipelines, and multi-agent systems. As a Founding Engineer, led the development of an AI platform from 0 to 1, delivering infrastructure, APIs, sandboxed AI code execution, and deployment workflows that scaled to 4,000+ beta users and 500+ daily active users. Strong background in cloud infrastructure, distributed systems, and secure payment gateway integrations with production-grade reliability.

TECHNICAL SKILLS

AI & LLM: RAG, Agentic AI, Multi-Agent Systems, LLM Orchestration, MCP (Model Context Protocol), OpenAI, Anthropic, Gemini

Backend: TypeScript, Node.js, Python, REST APIs, WebSockets, Microservices

Frontend: React, Next.js

Cloud & DevOps: AWS, Azure, Docker, Kubernetes, CI/CD, Prometheus, Grafana

Databases: PostgreSQL, Redis, MongoDB, Vector Search, pgvector

WORK EXPERIENCE

Founding Engineer — PrettiFlow

Jan 2026 – Present

Remote (San Francisco, USA)

- Owned end-to-end full-stack development of an AI platform enabling users to generate, deploy, and manage full-stack applications from natural language prompts, scaling to 4,000+ beta users with 500+ daily active users.
- Architected the AI application generation pipeline and deployment orchestration end-to-end—from prompt intake through code generation, infrastructure provisioning, and live deployment—reducing full-stack build time from 15 minutes to 8 minutes.
- Built sandboxed AI code execution with a pre-warming strategy that cut cold-start latency from ~3s to ~100ms, streaming results to the frontend via Redis Pub/Sub and WebSockets.
- Developed a cross-platform CLI for developers covering infrastructure provisioning, deployment automation, and AI-assisted development workflows.
- Shipped AI agent safety guardrails—input/output sanitization, action-level blocking, loop detection—built into the core execution layer from day one.

Software Engineer — OmniqAI

Aug 2022 – Dec 2025

Bengaluru, India, Full-time

- Owned the 0→1 backend architecture for an AI workflow platform, designing scalable services, PostgreSQL schemas, and REST APIs to power automated multi-step reporting workflows using LangChain-based RAG pipelines and vector databases.
- Built the Next.js frontend with WebSocket-based real-time workflow status and analytics dashboards, tightly integrated with backend job state.
- Developed an AI-powered collaborative documentation platform similar to Confluence, using RAG, pgvector, semantic search, document sharing, CRUD operations, and agent-friendly indexing.
- Built and integrated internal Model Context Protocol (MCP) servers, enabling AI agents to securely interact with internal tools and significantly expanding platform capabilities.
- Redesigned LLM job handling with Redis queues and WebSocket status updates, cutting perceived API response time by 30%.
- Resolved production DB bottlenecks with PgBouncer pooling and PostgreSQL partial indexes, reducing hot-path query latency by 40% and eliminating connection exhaustion under peak load.
- Integrated the Razorpay payment gateway with secure webhook signature verification and idempotent transaction processing.
- Contributed to on-prem → AWS migration (Docker, Kubernetes/EKS), enabling zero-downtime deployments.

TECHNICAL PROJECTS

OctaClaw — Multi-Agent Orchestration Platform

Node.js, TypeScript

- Designed a full-stack multi-agent system that breaks complex goals into a dependency graph and runs subtasks in parallel across specialist agents for research, code generation, review, and testing.
- Built tool integrations for filesystem, HTTP, email, databases, and Jira, enabling autonomous agents to interact with external systems.

EDUCATION

Brindavan Institute of Technology and Science

2019 – 2022

Bachelor's Degree