

SOHAIL SHAIK

s.suhail9849@gmail.com — [LinkedIn](#) — [GitHub](#) — +91 9642639515

EXPERIENCE

OmniqAI - *Backend Engineer*

January 2025 – Present

- Architected scalable backend systems using Python (FastAPI, asyncio) supporting high-concurrency asynchronous APIs and real-time WebSocket services for distributed workloads.
- Designed and implemented AI-powered backend architecture combining retrieval-augmented generation (RAG), semantic knowledge pipelines, and vector search infrastructure for intelligent decision support systems.
- Developed autonomous AI agents for technical interviews with advanced features: resume parsing, behavioral analysis, code evaluation, and multi-step reasoning with persistent memory layers.
- Optimized relational and NoSQL databases (PostgreSQL with read replicas, MongoDB, Redis) implementing connection pooling, caching strategies, and schema migrations ensuring zero-downtime deployments.
- Led infrastructure migration from on-premises to Kubernetes-based cloud environment, containerizing workloads with Docker, provisioning secured VPC clusters, automating deployments via Helm charts, and implementing Horizontal Pod Autoscaling.
- Implemented distributed system patterns including message queues (Redis, RabbitMQ), background task processing with Celery, and event-driven architectures for reliable asynchronous workflows.
- Engineered comprehensive CI/CD pipelines using GitHub Actions enabling zero-downtime releases, automated testing, migrations, and artifact management—reducing deployment time by 40%.
- Automated infrastructure provisioning with Terraform and CloudFormation for multi-environment, repeatable deployments across AWS resources (EKS, EC2, VPC, IAM, S3, Secrets Manager).
- Integrated end-to-end observability stack with Prometheus metrics, Grafana dashboards, centralized logging, and distributed tracing for proactive monitoring and rapid incident response.
- Optimized container orchestration and DevOps workflows, reducing cloud infrastructure costs by 20% while improving deployment resilience and system reliability.
- Collaborated cross-functionally to troubleshoot production incidents, optimize API latency and throughput, and implement disaster recovery procedures.

OmniqAI - *Backend Engineer Intern*

August 2024 – December 2024

- Developed and maintained backend microservices using Python, building scalable REST APIs and contributing to production systems.
- Implemented comprehensive monitoring and alerting infrastructure using Prometheus and Grafana, reducing mean time to detection (MTTD) by 30% and incident response times.

SKILLS

Backend Development: Python (FastAPI, Flask, asyncio), Node.js, RESTful & WebSocket APIs, System Design, Distributed Systems, Microservices Architecture

Databases & Data Stores: PostgreSQL (read replicas, connection pooling), MongoDB, Redis (caching, pub/sub), RabbitMQ, Neo4j

Cloud & Infrastructure: AWS (EKS, EC2, VPC, IAM, S3, CloudWatch, Secrets Manager), Docker, Kubernetes (Helm, HPA, probes), Terraform, CloudFormation

DevOps & CI/CD: GitHub Actions, GitOps, Prometheus, Grafana, Infrastructure as Code (IaC), Automated Testing

Programming Languages: Python, TypeScript, C++, Shell Scripting

Core Competencies: Problem-Solving, System Architecture, Scalable Design, Asynchronous Programming, Message Queues, DevOps Automation, Agile Collaboration

PROJECTS

- **TraceFox AI — AI-Powered Code Review & Test Platform** [\(Code\)](#)
 - Architected end-to-end quality workflow control center using FastAPI, orchestrating webhooks, modular backend services, and pipeline eventing for automated pull request analysis, test generation, root-cause analysis, compliance checking, and drift detection.
 - Designed robust API gateway with dynamic service registry for domain orchestration; integrated resilience patterns (retries, circuit breakers) and structured event bus for inter-service decoupling enabling distributed systems scalability.
 - Automated GitHub repository onboarding for cloning and indexing, supporting both simulated and real-world clone flows with async task tracking and persistent state management.
 - Developed AI review and test generation workflows leveraging advanced LLM models (DeepSeek R1, Gemma-3-27B-IT) with deterministic JSON parsing, robust fallback mechanisms, and intelligent caching strategies.
 - Built comprehensive dashboard UI with Next.js exposing real-time review data, test artifacts, RCA correlation, drift/compliance feedback aggregation, and engineering visibility dashboards.
 - Automated multi-cloud deployment infrastructure using Docker Compose and Terraform (AWS/Azure), integrated secure token/session management, and implemented structured end-to-end logging.
- Technologies:** Python, FastAPI, Node.js, Next.js, Redis, PostgreSQL, Neo4j, SQLite, Terraform, Docker, GitHub API, LLMs

EDUCATION

- **Bachelor of Technology, Computer Science** 2019 - 2022
Brindavan Institute of Technology and Science

ACHIEVEMENTS

- SAP Hackfest 2024: Top 4 out of 30+ teams (SAP India) — Developed scalable backend architecture for enterprise solution
- Flipkart Robotics Challenge: Finalist — Advanced vision-based quality inspection system with distributed processing pipeline