

ABDUL WASAY SIDDIQUE

Senior AI Engineer | Python Engineer | AI Infrastructure | Cloud DevOps

Karachi, Pakistan | waseysiddique11@gmail.com | +92 333 0266070 | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

PROFESSIONAL SUMMARY

Senior AI & Python Engineer with 6+ years of experience architecting production-scale AI systems, distributed backend infrastructure, and realtime AI platforms. Specialized in building enterprise RAG pipelines, agentic workflows, and cloud-native AI systems using OpenAI APIs, LangGraph, FastAPI, vector databases, and AWS. Proven experience leading engineering teams and shipping scalable AI products from architecture to production deployment.

Experienced in leading engineering teams, architecting high-performance backend systems, and delivering AI-powered products that automate workflows, reduce operational overhead, and improve business efficiency. Strong expertise in AI orchestration, multi-agent systems, streaming architectures, microservices, CI/CD automation, async systems, and cloud infrastructure optimization.

PROFESSIONAL EXPERIENCE

Senior Software Engineer & Team Lead

Xloop Digital Services | Karachi, Pakistan

Nov 2022 – Present

- Lead engineering initiatives focused on AI infrastructure, production-grade backend systems, and enterprise cloud architectures while managing a 12-member engineering team.
- Architected and deployed 8+ production-grade full-stack systems and 4+ enterprise AI platforms involving RAG pipelines, AI agents, realtime AI systems, and distributed backend architectures.
- Built a production-grade AI Voice Ordering Platform using OpenAI Realtime API, WebRTC, FastAPI, Redis, and AWS infrastructure, reducing order completion time by 40%.
- Designed enterprise AI knowledge systems with semantic retrieval pipelines using LangChain, OpenAI embeddings, vector databases, chunking strategies, and contextual search optimization.
- Developed multi-agent orchestration systems using LangGraph and CrewAI involving agent memory, task delegation, workflow orchestration, and distributed execution pipelines.
- Led backend architecture initiatives involving microservices, async processing systems, Redis-based orchestration, event-driven workflows, and scalable API infrastructure.
- Implemented containerized infrastructure using Docker, ECS, and GitHub Actions CI/CD pipelines, improving deployment efficiency by 40% while reducing infrastructure maintenance overhead.
- Designed scalable AWS infrastructure supporting high-concurrency AI workflows, realtime processing pipelines, and distributed microservices deployments.
- Built AI-powered HR analytics systems with intelligent insights, productivity scoring, scheduling automation, and AI-driven reporting pipelines.
- Established engineering standards around architecture reviews, deployment automation, testing workflows, infrastructure monitoring, and technical scalability.

Technologies Used: Python, FastAPI, Django, Flask, OpenAI APIs, LangChain, LangGraph, CrewAI, Redis, PostgreSQL, AWS, Docker, ECS, GitHub Actions, Qdrant, ChromaDB, React, TypeScript, WebRTC

Full Stack Developer & Cloud Engineer

Changes on the Fly | Toronto, Canada (Remote)

Jan 2025 – Present

- Worked on enterprise application modernization, cloud migration, infrastructure automation, and deployment optimization for large-scale backend systems.
- Migrated legacy PHP-based infrastructure from Linode to AWS using Dockerized deployment pipelines and automated CI/CD workflows.
- Developed and maintained enterprise-grade Angular + PHP applications integrated with AWS infrastructure and scalable backend services.
- Implemented zero-downtime deployment strategies, infrastructure automation scripts, and monitoring pipelines for production reliability.
- Optimized cloud infrastructure architecture using AWS EC2, RDS, API Gateway, Amplify, and Lambda services.

Technologies Used: Angular, PHP, Docker, AWS, EC2, RDS, Lambda, API Gateway, GitHub Actions, CI/CD Automation

Senior Software Engineer

AKSIQ Technologies | Karachi, Pakistan

Jan 2022 – Oct 2022

- Worked on banking and fintech systems involving backend engineering, fraud detection pipelines, deployment automation, and scalable infrastructure.
- Led backend engineering initiatives for banking and fintech applications using Django and Flask.
- Built fraud detection and transaction monitoring systems using Python automation and transaction pattern analysis.
- Containerized backend systems and deployment workflows using Docker and CI/CD pipelines, reducing manual deployment overhead by 50%.
- Automated data collection and analytics workflows using Selenium and Python-based scraping infrastructure.

Technologies Used: Python, Django, Flask, Docker, PostgreSQL, Selenium, CI/CD, AWS

Python Developer

Linkstar | Karachi, Pakistan

Sep 2020 – Dec 2021

- Worked on scalable backend systems, automation workflows, and deployment pipelines for production web applications.
- Developed backend architecture for a freelancing platform using Flask and React.
- Built automated data extraction and scraping systems using Selenium and Flask.
- Managed deployment workflows and CI/CD pipelines for multiple production applications.

Technologies Used: Python, Flask, React, Selenium, PostgreSQL, Heroku, CI/CD

AI & ENGINEERING PROJECTS

Enterprise AI Knowledge Platform

Designed and deployed a scalable enterprise RAG platform capable of semantic document retrieval, contextual search, embeddings generation, and intelligent AI-assisted querying.

- Built chunking and embedding pipelines
- Implemented semantic search using vector databases
- Designed scalable ingestion workflows
- Optimized retrieval latency and context relevance
- Integrated OpenAI APIs for contextual generation

Stack: FastAPI, LangChain, OpenAI APIs, Qdrant, PostgreSQL, Docker, AWS

Realtime Voice AI Ordering Platform

Built a realtime multilingual voice AI platform capable of handling live customer interactions using streaming AI pipelines.

- Realtime audio streaming with WebRTC
- Low-latency AI response orchestration
- Interruption handling and session memory
- Scalable backend session management
- AI-powered ordering workflows

Stack: OpenAI Realtime API, WebRTC, FastAPI, Redis, AWS, Docker

Multi-Agent AI Workflow System

Developed distributed multi-agent orchestration systems capable of autonomous task delegation and workflow execution.

- Agent memory systems
- Distributed workflow orchestration
- Tool-calling pipelines
- Multi-step reasoning flows
- Redis-backed state management

Stack: LangGraph, CrewAI, FastAPI, Redis, PostgreSQL

AI Productivity & Workforce Analytics Platform

Built an AI-powered workforce analytics and productivity tracking platform involving intelligent scheduling, activity analysis, and operational insights.

- Activity tracking systems
- AI-driven analytics pipelines
- Intelligent scheduling workflows
- Operational reporting dashboards
- Scalable backend APIs

Stack: FastAPI, PostgreSQL, React, OpenAI APIs, AWS

CORE TECHNICAL EXPERTISE

Generative AI & AI Systems: OpenAI APIs (Responses API, Function Calling, Realtime API), RAG Pipelines, AI Agents & Agentic Workflows, Multi-Agent Systems, LangChain, LangGraph, CrewAI, MCP Servers, Prompt Engineering, Structured Outputs, Tool Calling, Context Engineering, AI Evaluation Pipelines, Vector Search & Semantic Retrieval, Embeddings & Chunking Strategies, AI Orchestration Systems, Realtime Voice AI, Streaming AI Systems, AI Workflow Automation

Backend Engineering: Python, FastAPI, Django, Flask, REST APIs, GraphQL APIs, AsyncIO, AioHTTP, Microservices Architecture, Distributed Systems, Event-Driven Systems, WebSockets, Background Processing, Queue-Based Architectures, API Gateway Design, High-Performance API Optimization

Cloud & Infrastructure: AWS ECS, AWS EC2, AWS Lambda, AWS API Gateway, AWS RDS, AWS CloudWatch, AWS Route53, AWS S3, AWS IAM, AWS Elastic Beanstalk, AWS Amplify, Docker, Kubernetes, GitHub Actions, CI/CD Automation, Nginx, Linux Systems, Infrastructure Automation, Monitoring & Logging, Containerized Deployments

Databases & AI Storage: PostgreSQL, Redis, MongoDB, MySQL, MSSQL, Qdrant, ChromaDB, Pinecone, Vector Databases, Semantic Search Systems, Caching Architectures

Frontend & Integration: React, TypeScript, JavaScript, Angular, Redux, API Integrations, Realtime Frontend Systems

Engineering Practices: System Design, Scalable Architecture, Performance Optimization, Agile/Scrum, Technical Leadership, Code Reviews, TDD, Unit Testing, Integration Testing, CI/CD Best Practices, Technical Documentation

EDUCATION

Bachelor of Computer Engineering

Mohammad Ali Jinnah University

2018 – 2022

Relevant Focus Areas: Artificial Intelligence, Database Systems, Algorithms & Data Structures, Software Engineering, IoT Systems, Distributed Computing

GPA: 3.46

CERTIFICATIONS

- Cloud Native Developer — Emeritus
- IoT Developer — PIAIC
- Web Development Using Python & JavaScript — Harvard Online

ADDITIONAL INFORMATION

Languages: English (Fluent), Urdu (Native)