UMER SIDDIQUI

Senior AI Engineer | ML Engineering & Production AI Systems

● um3rsiddiqui99@gmail.com ● +92-341-8094081 ● Karachi, Pakistan ● linkedin.com/in/um3r-siddiqui
● github.com/UmerSidiqui ● umersidiqui.github.io/Portfolio

■■ PROFESSIONAL SUMMARY

Senior AI Engineer specializing in production-grade AI systems, agentic workflows, and enterprise voice AI solutions. Expertise in building scalable RAG systems, LLM-based applications, and ML deployment pipelines.

Proven track record of delivering end-to-end AI solutions from development to production deployment.

■■ PROFESSIONAL EXPERIENCE

Senior AI Engineer

Saynt AI

06 /2025 – Present | Remote

- ■ Building intelligent voice agents and AI receptionists for enterprise clients using ElevenLabs and LLM frameworks
- ■ Developing agentic workflows with LangGraph for production-grade voice AI solutions
- ■ Implementing RAG systems and tool-calling capabilities via webhooks for context-aware conversations
- ■ Integrating voice AI with telephony systems (Twilio) and PMS (Practice Management Systems) like OpenDental and Dentrix for automated booking workflows
- ■ Optimizing AI agent performance, reliability, and latency for real-world enterprise deployment

Jr. AI Developer

Erly Stage Studios

10 /2024 – 06 /2025 | Remote Part-Time

- ■ Led AI development efforts using GenAI and LLMs, building sophisticated RAG systems
- ■ Developed agentic RAG architectures and custom AI agents for specialized use cases
- ■ Implemented LangChain and LangGraph workflows to boost project impact and efficiency
- ■ Created the Meeting Scheduling Agent: an intelligent WhatsApp bot using OpenAI API, LangGraph, and prompt engineering
- ■ Deployed solutions on GCP with Firestore for real-time updates and seamless natural language processing

AI and Data Analytics Intern

NCL NEDUET

04 /2024 - 10/2024 | Karachi, Pakistan

- ■ Contributed to AI and EDA (Exploratory Data Analysis) projects with real-world applications
- ■ Applied expertise in data exploration, AI model development, and optimization
- ■ Delivered impactful solutions to complex data challenges through statistical analysis and machine learning

■■ TECHNICAL SKILLS

Core AI/ML	Deep Learning • Machine Learning • Computer Vision • NLP • Data Analysis
LLM & GenAI	OpenAI API • Anthropic Claude • RAG Systems • Agentic AI • LangChain • LangGraph • Prompt Engineering
ML Engineering	MLflow • Weights & Biases • Model Versioning • Experiment Tracking • Model Optimization • ONNX
Development	Python (Advanced) • Django • Flask • FastAPI • PostgreSQL • SQL & NoSQL
DevOps	Docker • Kubernetes • CI/CD (GitHub Actions, Jenkins) • GCP • AWS
Vector DBs	Pinecone • Weaviate • Redis • Query Optimization • In-memory Caching
Voice & Integration	ElevenLabs • Twilio • Telephony • PMS Integration (OpenDental, Dentrix) • Webhooks

■■ KEY PROJECTS

Enterprise Voice AI Agent @Saynt AI

ElevenLabs | Twilio | LangChain | PostgreSQL | PMS Integration

- Built intelligent appointment scheduling system with natural language understanding
- Integrated Twilio for telephony and ElevenLabs for human-like voice interactions
- Implemented RAG-based context retrieval for personalized patient interactions
- Created webhook-based tool calling for real-time PMS data access and booking

Meeting Scheduling Agent @Erly Stage Studios

OpenAI API | LangGraph | GCP | Firestore | WhatsApp API

- Developed agentic workflow using LangGraph for complex scheduling logic
- Collaborated with Anthropic AI engineer on architecture design
- Implemented context-aware conversations with prompt engineering
- Deployed on GCP with Firestore for real-time synchronization

Patient Face Recognition System

DeepFace (VGG-Face) | Flask | SQLite | RAG | Computer Vision

- Built registration system with facial embeddings using DeepFace and VGG-Face model
- Implemented real-time face recognition matching live video against patient database
- Created RAG-based chatbot for appointment booking assistance

Violence Detection System

YOLOv8 | Deep Learning | Computer Vision | Flask

- Retrained YOLOv8 model on violence detection dataset
- Built Flask web application for video upload and real-time analysis
- Implemented streaming interface showing annotated frames with violence alerts

■■ EDUCATION

Bachelor of Engineering in Electrical Engineering

NED University of Engineering and Technology

02 /2022 - 02/2026

Specialization in AI, Machine Learning, and Data Analytics • Grade: 2.5/4.0 CGPA