

Osama Altaf

Founder & CEO, Wisko (wiskoai.com) | AI Engineer - LLM & Voice AI

Pakistan | (+92) 318 7661096 | osamaaltaf.pk@gmail.com
linkedin.com/in/osamaaltafpk | github.com/osamaaltaf-pk

SUMMARY

Founder and CEO of Wisko (wiskoai.com), a multi-tenant AI SaaS whiteboard-animation platform, and an AI Engineer specializing in production-grade large language model systems, real-time voice AI, and agentic automation. Experience ranges from multilingual voice platforms handling live telephony to a government-grade real-time translation system spanning 30+ languages. Focused on making models deployment-ready: fine-tuning LLMs and vision-language models, hyperparameter tuning, and accelerating inference with KV caching and quantization using PipeCat, LiveKit, LangChain, LangGraph, and CrewAI.

TECHNICAL SKILLS

LLM & GenAI:	Fine-tuning, RAG, quantization, KV cache, prompt engineering, transformers, vision-language models
Frameworks:	LangChain, LangGraph, CrewAI, AutoGen, PipeCat, LiveKit, Hugging Face
Voice AI:	Whisper ASR, STT/TTS pipelines, SIP trunking, WebRTC, real-time telephony
Inference & MLOps:	vLLM, LMCache, AIBrix, Docker, model deployment, MLOps
Backend & Data:	Python, FastAPI, Redis, REST APIs, Pinecone, Weaviate, Qdrant, vector databases
Automation:	n8n, Make, Zapier

PROFESSIONAL EXPERIENCE

Founder & CEO - Wisko (wiskoai.com)

Multi-tenant AI SaaS - Whiteboard Animation Studio

Current

- Founded and lead Wisko, a multi-tenant SaaS platform where an agentic pipeline turns a prompt into a finished whiteboard-animation video.
- Own product direction, system architecture, and the agentic content-generation stack.

AI Engineer - LLM Systems Research & Analyst (Contract, Remote)

London, United Kingdom

Jul 2025 - Jan 2026

- Deployed and scaled large language models for smart assistants and task automation at company level.
- Used vLLM for high-throughput, low-latency inference with efficient KV cache use.
- Integrated AIBrix multi-agent orchestration to coordinate tools and APIs into goal-driven agent workflows.
- Fine-tuned LLMs on proprietary datasets to align outputs with company-specific terminology.
- Built scalable FastAPI services with in-memory caching via LMCache and Redis.

AI Voice Agent Developer & Research Analyst - Government Real-Time Translation System

Confidential GovTech Project, Remote - Morocco

Feb 2025 - Jun 2025

- Co-developed a real-time multilingual communication system with bi-directional translation across 30+ languages.
- Integrated Whisper ASR with custom speech pipelines, achieving sub-1500 ms latency using LiveKit and PipeCat.
- Engineered context-aware memory modules to retain conversational history and ensure cross-language consistency.
- Met government-grade security, privacy, and performance standards.

AI Engineer - Voice AI Agent Platform (US Project, Remote)

Miami, United States

Nov 2024 - Feb 2025

- Designed a scalable Voice AI platform supporting inbound/outbound telephony via SIP trunk routing.
- Built modular, no-code agent workflow builders enabling users to define dynamic call flows.
- Integrated LangGraph-powered agents for context-aware, multi-turn conversations.
- Hardened Speech-to-Text and Text-to-Speech pipelines for production reliability.

Associate Business Development (Internship, Remote)

iAxon Software House

Oct 2023 - Feb 2024

- Conducted market analysis that drove a 10% increase in client acquisition.
- Coordinated product launches, consistently meeting or exceeding deadlines.
- Improved client communication strategies, increasing retention and satisfaction.

Junior AI Engineer (Intern)

iAxon Software House

Apr 2023 - Oct 2023

- Optimized model-training workflows, cutting compute time 30% via improved feature engineering.
- Built Python predictive-modeling solutions, improving forecasting accuracy 15%.
- Fine-tuned ML models across diverse domains and automated data processing (+20% team productivity).

EDUCATION

BS Physics - Computational Physics & Computer Programming

Khwaja Fareed University of Engineering & IT (KFUEIT), Pakistan

CGPA 3.63 / 4.00

Thesis: Object Detection Transformer model for disease detection on cotton leaves.

Additional: Machine Learning, MLOps & LLM Zoomcamps - DataTalks.Club (2023). F.Sc. Pre-Medical, BISE Bahawalpur - 89%. Matriculation, BISE Bahawalpur - 95%.

SELECTED PROJECTS

- Airport Vehicle Detection, Vision vs. LiDAR (freelance research, UK): Benchmarked a range of computer-vision models for detecting vehicles at airports, assessing whether camera-based systems can replace LiDAR by estimating depth, distance, and detection quality across varied lighting, seasons, and weather.
- Cotton Disease Detection Transformer (Final Year Project): Vision Transformer with an AI digital microscope for accurate cotton-crop disease detection from visual symptoms; outperformed CNN baselines.
- Autonomous Financial Analyst Swarm: Multi-agent CrewAI system that scrapes market news, runs sentiment analysis, and drafts investment memos with self-correction loops and tool use.
- Secure Enterprise RAG Knowledge Base: Privacy-preserving document assistant on local Llama 3 and Qdrant for semantic search with no external data leakage.
- Edge-AI Industrial Safety Monitor: YOLOv8 PPE detection on NVIDIA Jetson Orin Nano at <40 ms offline inference for low-bandwidth factory environments.
- Portable Radar System: Raspberry Pi remote-sensing radar with real-time Python signal processing for distance and velocity estimation.

CERTIFICATIONS

DeepLearning.AI: Fine-Tuning LLMs; LLMOps - Deployment & Inference; Building AI Voice Agents for Production; Practical Multi-AI Agents with CrewAI; AI Agentic Design Patterns with AutoGen; Evaluating AI Agents; Vector Databases: From Embeddings to Applications; LangChain for LLM App Development; Building Coding Agents. edX (IBM): Mastering Generative AI - Agents with RAG & LangChain. DataTalks.Club: MLOps & LLM Zoomcamp.

LANGUAGES

English: IELTS 7.5 (C2 Listening & Reading; C1 Speaking & Writing)

Urdu: Native

LEADERSHIP & ACTIVITIES

- President, Natural Resources Defense Council & Society, KFUEIT.
- Organizer of ICPSE-2021 and ICPSE-2022 (International Conferences on Physical Sciences & Engineering) and ICSMAND (9th International Conference on Semiconductor Materials & Nano-Devices).
- Organized the 1st South Punjab Science Mela 2021 and a Blood Donation Camp at Sehja Public Health Care Centre.
- Student Proctor at the 1st KFUEIT Convocation; Organizer and Cricketer at the Annual Sports Gala.