

Muhammad MOIZ

AI/ML Engineer | Data Science & MLOps | Software Engineering

[in linkedin.com/in/m2moiz](https://www.linkedin.com/in/m2moiz) github.com/m2moiz
+33-786799795 @ MuhammadMoiz.Work@proton.me

Mechatronics engineer turned AI lead with 7+ years across embedded systems, edge deployment, and production AI. Led a team that shipped a multilingual voicebot and real-time video analytics pipelines on constrained industrial hardware. Experience with multimodal models, object detection and tracking, using both cloud APIs and open-source on-device solutions. AWS certified with production deployments across Azure, GCP, and AWS. Actively seeking AI/ML, data science, or software engineering roles in France.

PROFESSIONAL EXPERIENCE

AI/ML Engineer | Publicis Resources | Paris, France

11/2025–02/2026

- › Production LLM agents with LangChain : RAG pipelines, multi-provider gateway (OpenAI/Gemini), automated brand compliance.
- › Built automated multi-model benchmarking in CI/CD to track response quality over time.
- › Rewrote a legacy codebase into a LangChain agent architecture; deployed with Docker on Azure.
- › Fine-tuned vision and language models; integration with Adobe SDKs (PSD, Illustrator, Premiere).

Python FastAPI LangChain RAG LLM Agents PyTorch Docker Azure OpenAI Gemini CI/CD

Senior AI Lead | Zaheen Systems | Remote, Singapore

11/2022–10/2025

- › Built a voicebot with STT-LLM-TTS pipeline optimized to sub-400ms latency across four languages.
- › Designed conversational agents grounded in business data via RAG for decision support.
- › Deployed real-time video analytics pipelines across 10+ cameras with GPU-accelerated inference (DeepStream).
- › Optimized models with TensorRT and OpenVINO for deployment on Jetson edge devices.
- › Led a team of 5 engineers delivering AI systems for government clients across multiple contracts.

Python PyTorch TensorRT OpenVINO LangChain RAG DeepStream Docker GCP FastAPI MLFlow

Data Science Intern | ResilEyes Therapeutics | Remote, France

02/2022–07/2022

- › Built a multimodal pipeline fusing video, audio, and text streams for clinical PTSD assessment.
- › Curated an annotated clinical dataset (1000+ clips, 20h) from survivor interviews and documentaries.
- › Compared a three-branch late-fusion approach against a unified end-to-end Perceiver Transformer.
- › Identified audio spectrograms as the dominant modality for PTSD detection.

PyTorch Multimodal Signal Processing NLP Transformers Weights & Biases

Research Assistant | Technische Universität München | Munich, Germany

07/2021–12/2021

- › Fused mocap and Lidar data to build spatial safety maps for human-robot shared workspaces.
- › Improved human trajectory prediction by 30% with a bidirectional LSTM; validated via ROS bag replay.
- › Ported the Signed Distance Field computation from MATLAB to Python to unify the ROS/Gazebo framework.
- › Contribution to the EU Horizon 2020 HR-Recycler project (€7M, 12 partners) on WEEE disassembly.

ROS Sensor Fusion LSTM Lidar Gazebo Python MATLAB Industrial Safety

Robotics Engineer | GEM Learning | Islamabad, Pakistan

01/2020–07/2021

- › Modified PX4 firmware to support custom VTOL and powered-paraglider airframes.
- › Deployed real-time perception on Jetson Nano with TensorRT-optimized models via ROS.
- › Built a sim-to-real pipeline (Gazebo/PX4 SITL) to validate drone behavior before field tests.

Embedded Firmware IoT Edge AI TensorRT Jetson Nano ROS PX4 C++ Python

Robotics Engineer | Quest Labs | Islamabad, Pakistan

07/2018–12/2019

- › Designed and built a 200kg autonomous vehicle with Lidar, stereo cameras, and Jetson Nano; full SLAM navigation.
- › Coordinated a fleet of 5 drones with environment cameras for real-time tracking and formation control.
- › Developed a 6-axis robotic arm with trajectory planning (ROS/MoveIt) and vision for part detection.

Hardware IoT ROS MoveIt Lidar Jetson Nano Edge Computing C++ Python

SKILLS

Generative AI & Agents	LLM Agents, RAG, Prompt Engineering, LangChain, Agentic AI, Voice Agents, NLU, STT/TTS
Machine Learning & Vision	PyTorch, Fine-tuning, HuggingFace, Computer Vision, OpenAI, Gemini, Mistral
MLOps & Cloud	Docker, CI/CD, MLFlow, FastAPI, REST APIs, Azure, GCP, AWS, PostgreSQL
Edge & IoT	TensorRT, OpenVINO, Jetson Nano, Sensor Fusion, MQTT, Real-Time Systems
Languages	Python, C++, SQL, Bash, JavaScript/TypeScript

PROJECTS

AI TRAINING AVATAR PLATFORM

- › Built two voice pipelines (cascaded STT-LLM-TTS and end-to-end multimodal) optimized to sub-400ms latency.
- › Grounded responses in company training materials via RAG; supported 4 languages with localized accents.
- › Delivered an SDK enabling clients to deploy their own training courses at scale.

STT/TTS LangChain RAG FastAPI Voice Agents Multilingual AI Real-Time Inference

DOCUMENT DIGITIZATION WITH VISUAL LLMs

- › Unified physical receipts, email, and WhatsApp into a single pipeline; reduced staff from 2 FTE to 1 part-time.
- › Fine-tuned LLaVA for on-device extraction (client privacy constraints) on a manually annotated dataset.
- › Designed line-by-line total verification to catch discrepancies missed by human operators.

LLaVA Fine-tuning TensorRT Distillation RAG WhatsApp API Python AWS

MULTI-CAMERA VIDEO ANALYTICS

- › Deployed real-time detection, tracking, and behavioral alerts across 10 cameras; incidents reduced by 70%.
- › Built a smart tollbooth across 8 cameras with OCR and vehicle classification on a fully on-premise edge architecture.
- › NVIDIA DeepStream YOLO + DeepSORT pipeline optimized via TensorRT for concurrent multi-stream processing.

YOLO DeepSORT DeepStream TensorRT CUDA Edge AI Python

AUTONOMOUS GROUND VEHICLE

- › Designed and built a 200kg vehicle (steel frame, drivetrain, full sensor suite) with autonomous SLAM navigation.
- › Deployed a ROS stack on Jetson with Lidar, stereo cameras, and GPS; validated in Gazebo before field tests.
- › Integrated a front-mounted robotic arm with object pickup demonstration via stereo vision.

IoT Sensor Fusion Lidar Edge AI Jetson Nano ROS Hardware Python C++

EDUCATION

Masters in Computer Vision and Robotics | Université Bourgogne-Europe | *Le Creusot, France*, 2020–2022

- › Ranked 1st in the M1 cohort and graduated with honors.
- › Received the EIPHI Excellence Scholarship, a nationally recognized merit-based award.

Multi-Sensor Fusion Signal Processing NLP Machine Learning Software Engineering Advanced Image Analysis

Bachelor of Mechatronics Engineering | Air University | *Islamabad, Pakistan*, 2015–2019

- › Thesis on designing and building an autonomous delivery drone from scratch received Distinction.
- › Awarded the Dean's Scholarship for academic performance.

Embedded Systems IoT Microcontrollers Control Engineering AI Systems DSP Sensors & Actuators

CERTIFICATIONS

AWS Certified Machine Learning - Specialty | *Amazon Web Services (AWS)* 02/2025
AWS Certified Solutions Architect - Associate | *Amazon Web Services (AWS)* 12/2024