

HAMZA BASHARAT

AI / Computer Vision Engineer — Deep Learning · Edge AI · MLOps · LLM & RAG Systems

Lahore, Pakistan · Open to Relocation & Visa Sponsorship (Germany / EU) · +92 300 6547302 · hamzabasharat2004@gmail.com

linkedin.com/in/hamzabasharat26 · github.com/hamzabasharat26 · Recommendations: linkedin.com/in/hamzabasharat26/details/recommendations

PROFESSIONAL SUMMARY

AI / Computer Vision Engineer with 2+ years of production experience owning vision systems end to end — dataset curation, PyTorch and YOLO model training, ONNX / TensorRT optimisation, FastAPI serving, Docker deployment and post-deployment monitoring. Shipped MagicQC, live on AWS EC2 processing 500+ items/day at 90%+ accuracy, and holds 92%+ mAP at 15+ FPS on NVIDIA Jetson, Raspberry Pi and Luxonis OAK edge hardware. Additional delivery in LLM and RAG agent systems using LangChain, LangGraph and the Claude and OpenAI APIs. Invited trainer for a university CPD workshop on Vision AI in manufacturing. 1st Place, IEEE Hackathon 2024 — converted into a paid commercial deployment.

TECHNICAL SKILLS

Computer Vision: Object detection, object tracking, image classification, semantic & instance segmentation, pose estimation, anomaly and defect detection, OCR, stereo depth & spatial localisation, real-time video analytics · YOLOv5 / v8 / v11 / v26 · OpenCV · PatchCore · Vision Transformers (ViT) · Vision-Language Models (VLM)

Deep Learning & ML: PyTorch, TensorFlow, Keras, scikit-learn, Hugging Face Transformers & Diffusers, CNNs, transfer learning, fine-tuning, data preprocessing & augmentation, diffusion-based synthetic data, hyperparameter tuning, model evaluation (mAP, precision / recall)

Edge AI & Optimisation: NVIDIA Jetson Nano, Raspberry Pi, Luxonis OAK-D / OAK-D Pro W / OAK-1W (DepthAI), Stereolabs ZED X Mini (ZED SDK) · ONNX, TensorRT, INT8 / FP16 quantisation, model compression, latency & inference optimisation, threading and multiprocessing, on-device inference

MLOps & Deployment: Docker, AWS (EC2, S3), Microsoft Azure AI, FastAPI, Flask, Streamlit, REST API design, MLflow, model versioning, monitoring, CI/CD, n8n workflow automation, Git & GitHub, Linux (Ubuntu, JetPack), Bash

LLM, NLP & Agentic AI: Anthropic Claude API, OpenAI API, LangChain, LangGraph, RAG pipelines, vector databases (FAISS, Chroma), NLP, prompt engineering, context management, conversational & voice agents, ElevenLabs text-to-speech, speech synthesis integration

Data, Languages & Tools: Python (primary), C++, SQL, JavaScript · NumPy, Pandas · MySQL, MongoDB · Django, React, Next.js · Roboflow, CVAT, dataset curation & annotation · ROS / Gazebo · English (Professional), Urdu (Native)

PROFESSIONAL EXPERIENCE

AI & Computer Vision Engineer — Robionix Technologies

Islamabad, Pakistan | Aug 2024 – Jul 2026

- Replaced a manual measurement process with MagicQC, an automated vision platform processing 500+ garments/day at 90%+ accuracy, by owning dataset curation, PyTorch model training and a documented FastAPI inference API containerised with Docker on AWS EC2.
- Sustained 15+ FPS real-time inference at 92%+ mAP on constrained edge hardware, by converting models to ONNX, applying INT8 / FP16 quantisation and model compression, and restructuring capture, inference and post-processing to run concurrently through threading and multiprocessing.
- Deployed detection and pose-estimation models across NVIDIA Jetson Nano, Raspberry Pi and Luxonis OAK devices, tuning memory and CPU / GPU usage per board so one model family held each device's thermal and power budget.
- Improved spatial measurement accuracy on 3D targets by integrating stereo-depth capture from ZED X Mini and OAK-D cameras, using depth to resolve the scale and occlusion cases that RGB-only detection consistently mis-measured.
- Enabled natural-language search of enterprise knowledge bases by designing RAG pipelines and conversational agents with LangChain, LangGraph and the Claude and OpenAI APIs over vector databases, served as REST inference APIs into Laravel, React and MySQL backends.
- Selected as trainer for a two-day university CPD workshop on Vision AI in manufacturing, teaching YOLO detection and classification, real-time industrial camera interfacing, Python / Django inference APIs and an end-to-end fabric-defect-detection build with live ERP quality dashboards.

AI Engineer (Contract, concurrent) — Essenceware Technologies

Pakistan | Jul 2025 – Sep 2025

- Delivered a real-time PPE and workplace-safety detection system end to end in a 10-week contract, by cleaning and annotating the dataset, training and evaluating YOLO models, and deploying a Flask and Streamlit inference dashboard.
- Enabled the client team to retrain without support by adding model versioning and documenting every training run, evaluation result and handoff procedure.

AI Engineer, Intern (concurrent) — NESCOM, National Development Complex

Islamabad, Pakistan | Jun 2025 – Sep 2025

- Built SkyResQ, a UAV disaster-response perception pipeline evaluated by government stakeholders for national emergency preparedness, by developing near-real-time detection and tracking in Python and OpenCV, validated through Gazebo simulation and hardware integration testing.

SELECTED PROJECTS

AI-Enabled Fabric Defect Detection — Final Year Project with MEB Karachi. Reached mAP50 0.81 on a self-curated 1,400+ image dataset by fusing dual-branch YOLO classification with PatchCore (WideResNet-50) anomaly detection via IoU matching, deployed to an OAK-1W edge camera with severity grading and automated CSV / JSON reporting. github.com/hamzabasharat26/Automated-Fabric-Defect-Detection-System

Dock Vision AI — 1st Place, IEEE Hackathon 2024. Real-time vessel-docking safety system using video analytics for spatial localisation and alignment detection; converted into a paid commercial deployment. github.com/hamzabasharat26/DockVision-AI

Nexus AI — Fuel-Price RAG Agent — Retrieval-augmented agent answering natural-language queries over fuel-price data, built on LLM APIs and a vector store with citation-backed responses.

AI Portfolio Voice Agent — Conversational portfolio assistant with ElevenLabs text-to-speech voice synthesis, retrieval-backed answers over a structured project record, and an audio-reactive React / Next.js interface.

Content Automation Pipeline (n8n) — Automated workflow that pulls scheduled posts from a Notion database, formats them, and publishes to LinkedIn without manual handling.

CV Recruiter RAG Agent — Resume-screening agent parsing PDF and DOCX into a FAISS vector store, ranking candidates against a job description with citation-backed scoring; LangGraph, Claude API, FastAPI.

EDUCATION

BS Computer Engineering — National University of Technology (NUTECH), Islamabad | 2022 – 2026

TRAINING, CERTIFICATIONS & AWARDS

Trainer — CPD Workshop “Applications of Vision AI in Manufacturing Industries”, Punjab Tianjin University of Technology (PTUT), Lahore, Jul 2026 — delivered by NUTECH with Robionix Technologies.

Certifications: AWS Certified Machine Learning Engineer – Associate · Microsoft AI & ML Engineering Professional Certificate (Azure AI / Computer Vision) · Advanced Computer Vision with TensorFlow · LangChain for LLM Application Development (DeepLearning.AI) · Introduction to Embedded ML (Edge Impulse)

Awards: 1st Place — IEEE Hackathon 2024 (Dock Vision AI) · 1st Place — ICAT National Robotics Competition · Runner-Up — National AI Sprint 2025 (NCP)