

HASHAM AKRAM

Lahore, Pakistan

+923249709214 hashamakram50@gmail.com linkedin.com/in/hasham-akram

AI/ML Engineer driving impactful solutions with expertise in NLP, CV, GenAI leveraging MLOps and data-driven Machine Learning. Focused on orchestrating cutting-edge AI systems dynamically.

Skills

- **Programming Languages:** Python, FastAPI, C++, SQL, Django REST
- **Data Science Toolkit:** Pandas, NumPy, NLTK, Matplotlib, SciPy, Playwright
- **ML/DL Libraries & Frameworks:** PyTorch, TensorFlow, Scikit-learn, HuggingFace Transformers, LangChain, LangGraph
- **Core AI/ML:** NLP, Computer Vision, Deep Learning (Transformers, Generative AI incl. Diffusion Models PEFT/LoRA for LLMs, Agentic AI), Scientific Machine Learning
- **MLOps & Cloud:** AWS (SageMaker, Lambda, Bedrock), GCP (VertexAI, Cloud Run), Docker, Git, MLflow, DVC, Kubernetes, CI/CD (GitHub Actions), Prometheus
- **Databases:** MySQL, MongoDB, Pinecone, ChromaDB, Qdrant, Weaviate

Experience

Silicon Nexus

Dec 2025 – Present

AI & ML Engineer (WMD Team)

Lahore, Pakistan · On-Site

- Built AI-powered compliance detection system using Django REST Framework with Corrective RAG and vector-based rule validation for national registration workflows.
- Led AI architecture for an AI/AR Virtual Garment Fitting Mirror platform supporting multi-store deployment.
- Experimented with DM-VTON and HR-VTON; optimized real-time virtual try-on pipeline for low-latency, high-fidelity inference.

Sprouto Groups

May 2025 – December 2025

AI & ML Engineer

Lahore, Pakistan · On-Site

- Built an AI-Human document classification pipeline using fine-tuned Gemma-3B, BERTa, and QLoRA (logits ensemble) as instruct; integrated with Gemini API for hybrid predictions, achieving 90% accuracy and 0.90 F1-score.
- Deployed FastAPI microservice on GCP Cloud Run via Docker for scalable, low-latency inference.
- Led 100% automation of social media lead-generation using LangGraph agents, LLM APIs, and real-time web scrapers (Selenium, Playwright), reducing manual intervention by 90%.

AXON Technologies

Nov 2024 – May 2025

Machine Learning Engineer

Lahore, Pakistan · On-Site

- Developed and deployed ML models (XGBoost, Random Forest) on AWS SageMaker, improving R^2 by 30% and streamlining multi-disease models, reducing complexity by 40%.
- Automated model inference with AWS Lambda & EventBridge, boosting efficiency by 25%, while CloudWatch monitoring ensured 99.9% uptime. Integrated ML predictions into user interfaces, enhancing engagement by 30%.
- Built and deployed AI agents on Amazon Bedrock for multi-purpose tasks, including generating custom contracts between two signing parties, reducing contract generation time by 50% and improving accuracy by 35%.

Projects

Kidney Multi-Disease Classification | VGG-16, TensorFlow, DVC, MLflow, DagsHub, CI/CD, Pytest

Aug 2023

- Built a web app for multi-disease kidney classification from MRI images using VGG-16 (80% acc). Implemented robust MLOps with DVC, MLflow on DagsHub, CI/CD pipelines, and pytest for validation. 🔄

Small Language Model (SLM) from Scratch | Transformers, PyTorch, Tokenization, Training Dynamics

2026

- Designed and trained a 30M parameter transformer language model on a small corpus, observing emergence of grammar and narrative structure. Used GELU activation and GPT-style BPE (50k vocab) Encoder.
- Tracked perplexity against qualitative text generation, identifying divergence between loss improvements and actual output coherence.
- Extending toward RLHF and policy tuning to align generation behavior beyond next-token prediction.

Galaxy Morphology Classification | ViT, ResNet, Pytorch, Diffusion Models, Galaxy Zoo Dataset

Dec 2024

- Enhanced galaxy morphology classification on Galaxy Zoo, leveraging diffusion models for data augmentation and achieving 63% with Vision Transformers (ViT) by systematically progressing from VGG-16 and ResNet. 🔄