

Zain Habib

MLOps Engineer — Dubai, UAE

Email: zain.habib28@example.com | Phone: +971554185971

Professional Summary

Experienced MLOps Engineer with a strong track record of building production-grade machine learning systems, from data engineering and model training to deployment and monitoring. Skilled in both research prototyping and scalable MLOps practices. Proven ability to lead cross-functional teams and deliver business impact.

Technical Skills

PyTorch, Computer Vision (YOLO, Detectron2), NumPy, Time Series, CI/CD (Jenkins/GitHub Actions), Kubernetes, PyTorch Lightning, Spark, Kafka, Tableau

Professional Experience

AI Engineer — DataForge (1 yrs)

- Deployed scalable inference service on AWS with autoscaling, containerization (Docker), and serverless endpoints; reduced latency to <120ms.
- Built recommendation system using collaborative filtering + content-based features; increased CTR in A/B test by 12%.
- Designed and implemented an anomaly detection system for fraud detection using autoencoders and XGBoost for post-filtering.
- Optimized model inference using TensorRT and mixed precision; achieved 2.5x throughput improvement on GPU.

AI Engineer — DataForge (2 yrs)

- Fine-tuned multilingual speech recognition model (wav2vec2) for domain-specific calls with 5% error reduction.
- Implemented continuous training pipeline using DVC and GitHub Actions to automate model retraining and versioning.
- Optimized model inference using TensorRT and mixed precision; achieved 2.5x throughput improvement on GPU.

Deep Learning Engineer — VisionaryTech (9 yrs)

- Implemented continuous training pipeline using DVC and GitHub Actions to automate model retraining and versioning.
- Led an end-to-end image segmentation project for medical imagery; built data pipeline, trained U-Net variants, and reduced labeling time by 40%.
- Deployed scalable inference service on AWS with autoscaling, containerization (Docker), and serverless endpoints; reduced latency to <120ms.

Selected Projects

- Built recommendation system using collaborative filtering + content-based features; increased CTR in A/B test by 12%.
- Led an end-to-end image segmentation project for medical imagery; built data pipeline, trained U-Net variants, and reduced labeling time by 40%.
- Optimized model inference using TensorRT and mixed precision; achieved 2.5x throughput improvement on GPU.
- Implemented knowledge distillation to create lightweight transformer models for edge deployment with 3x speedup.

Education

BSc in Computer Science — National University, 2014

Certifications

- Coursera: Deep Learning Specialization

Languages

French (Conversational), English (Fluent)

Generated sample resume — 2025-09-30