

## Rami Nassar

AI Engineer — Remote

Email: rami.nassar4@example.com | Phone: +971589161505

### Professional Summary

Experienced AI 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

NoSQL (MongoDB), Pandas, API Development (FastAPI), CUDA, CI/CD (Jenkins/GitHub Actions), Azure ML, GCP (BigQuery), Linux, DVC, AutoML, Hugging Face Transformers

### Professional Experience

#### Computer Vision Engineer — EdgeAI Solutions (11 yrs)

- Optimized model inference using TensorRT and mixed precision; achieved 2.5x throughput improvement on GPU.
- Architected resume-parsing NLP pipeline using transformer-based models, custom NER, and rule-based post-processing to extract structured candidate data.

#### ML Researcher — AIWorks (2 yrs)

- Optimized model inference using TensorRT and mixed precision; achieved 2.5x throughput improvement on GPU.
- Deployed scalable inference service on AWS with autoscaling, containerization (Docker), and serverless endpoints; reduced latency to <120ms.
- Designed and implemented an anomaly detection system for fraud detection using autoencoders and XGBoost for post-filtering.

#### Machine Learning Engineer — EdgeAI Solutions (2 yrs)

- Architected resume-parsing NLP pipeline using transformer-based models, custom NER, and rule-based post-processing to extract structured candidate data.
- Built recommendation system using collaborative filtering + content-based features; increased CTR in A/B test by 12%.
- Implemented knowledge distillation to create lightweight transformer models for edge deployment with 3x speedup.

### Selected Projects

- 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.
- Led an end-to-end image segmentation project for medical imagery; built data pipeline, trained U-Net variants, and reduced labeling time by 40%.

### Education

BSc in Computer Engineering — University of Science, 2016

### Certifications

- Coursera: Deep Learning Specialization
- Certified Kubernetes Application Developer (CKAD)

### Languages

French (Conversational), Arabic (Native)

