Atmane Ayoub

Senior AI/ML Engineer · RAG · Agentic AI · Document AI · Computer Vision

Abu Dhabi, UAE atmaneayoub10@gmail.com atmaneayoub.xyz linkedin.com/in/atmaneayoub github.com/atmaneayoubdev

huggingface.co/atmaneayoub

SUMMARY

Senior AI/ML engineer with 8+ years of software engineering experience, including 5+ years in AI and machine learning. I design and deploy enterprise AI across generative and agentic AI, RAG, document intelligence, computer vision and predictive modelling — including the evaluation, inference optimisation, backend APIs, MLOps and secure on-premises deployment that make those systems dependable.

EXPERIENCE

Senior AI/ML Engineer · First.tech Group

Jun 2025 – Present · Abu Dhabi, UAE

Building and deploying production AI systems for insurance workflows across LLM applications, agentic automation, document intelligence and multimodal inference.

- Designed a multi-agent workflow for corporate client onboarding and endorsement, orchestrating document processing, insurer quotation comparison, benefit analysis, recommendation generation and multi-stage workflow execution.
- Developed a LangGraph-based insurance assistant using document-grounded RAG, embeddings, semantic retrieval and tool-driven workflows to answer questions about health insurance quotations and policy benefits.
- Developed components of an agentic workflow for financial claims recovery, automating document review, case analysis, structured decision steps and multi-stage processing with controlled workflow execution.
- Built an enterprise-wide document extraction service using fine-tuned open-source multimodal models to extract text and structured fields from sensitive documents, keeping processing on-premises.
- Developed and optimised GPU-accelerated inference APIs with FastAPI and Docker, reaching sub-5-second inference for image-based extraction workloads.
- Built evaluation, monitoring and observability workflows measuring extraction quality, structured-output consistency, latency, failure patterns and production reliability.

AI/ML Engineer · eData Information

Jun 2023 – Jun 2025 · Dubai, UAE

Built and deployed production AI systems for automotive valuation, vehicle data intelligence and insurance automation.

- Developed and maintained a production vehicle price prediction engine using ensemble regression and deep learning, with automated daily market-data ingestion, feature engineering, model evaluation and valuation updates.
- Built VIN and chassis-number recognition systems using computer vision and deep learning to extract vehicle identifiers from images.
- Delivered vehicle valuation and VIN recognition as scalable commercial APIs integrated into the company's automotive and insurance products.
- Designed automated pipelines for ingestion, web scraping, validation, deduplication and feature engineering, turning large volumes of automotive listings into structured datasets.
- Developed RAG applications and agentic workflows, fine-tuned open-source LLMs, and deployed containerised AI services.

AI/ML Engineer & Data Scientist · Independent contractor

May 2021 – Jun 2023 · Oran, Algeria · remote

Project-based ML and data engineering for startups and small businesses: predictive analytics, automation and decision support.

- Built ML models for regression, classification, forecasting and decision support.
- Owned data cleaning, feature engineering, model selection and hyperparameter tuning.
- Built ETL pipelines turning raw data into structured datasets for ML workflows.
- Evaluated models with cross-validation, error analysis and task-appropriate metrics.

Full-Stack & Mobile Developer · Independent contractor

Mar 2018 – Apr 2021 · Oran, Algeria · remote

End-to-end web and mobile applications for startups and SMEs across the MENA region, with clients in Saudi Arabia and Algeria.

- Built cross-platform mobile apps with Flutter, backend APIs, authentication and databases.
- Developed REST APIs with Django, Node.js and ASP.NET, including third-party integrations.
- Designed PostgreSQL and Firebase data models and ran deployment and maintenance.

SELECTED OPEN SOURCE

- Jev-AR (2026): 322M-parameter Arabic-first routing model for agents and RAG across MSA, Gulf dialects and code-switched Arabic. 83.4% on 60 held-out Gulf-Arabic routes at ~10 ms p50 on GPU, 38/40 on an unseen-domain RAG routing benchmark. Public model, benchmark and code.
- evalcascade: open-source adaptive evaluation for LLM, RAG and agentic systems. jev-mem: long-term memory for AI agents.

TECHNICAL SKILLS

LLM & agentic AI	RAG, LangGraph, Multi-agent systems, Tool calling, Fine-tuning (LoRA, QLoRA), Embeddings & semantic search, Multimodal models, Hugging Face Transformers
Machine learning	Regression & classification, Ensemble methods, Forecasting, Computer vision, NLP, PyTorch, TensorFlow, scikit-learn
Serving & optimisation	vLLM, GPU inference optimisation, Quantisation, Latency optimisation, FastAPI
Data & backend	Python, SQL, PostgreSQL, Pandas, Data pipelines & ETL, Microservices, Django, Node.js
MLOps & infrastructure	MLflow, Docker, Kubernetes, CI/CD, Prometheus, Airflow, AWS, Azure, GCP

EDUCATION

MSc, Artificial Intelligence 2020 – 2022
University of Science and Technology of Oran – Mohamed Boudiaf (USTO-MB), Algeria
Thesis: a machine-learning time-series forecasting system for regional rainfall prediction using real-time NASA meteorological data.

BSc, Computer Science 2017 – 2020
University of Science and Technology of Oran – Mohamed Boudiaf (USTO-MB), Algeria
Thesis: a recycling platform with web and mobile apps connecting households with waste collectors.

CERTIFICATIONS & LANGUAGES

- Certified Data Science, Data Analytics & Machine Learning Professional — LIRC, Dubai
- Advanced Learning Algorithms — DeepLearning.AI & Stanford
- Unsupervised Learning, Recommenders, Reinforcement Learning — DeepLearning.AI & Stanford
- Generative AI: Introduction and Applications — IBM
- Introduction to Artificial Intelligence — IBM

Languages: Arabic (native), Berber (Tamazight) (native), French (fluent), English (fluent)