Abdelilah Younsi

Data & AI Scientist

Paris-Palaiseau ✉ abdelilah.younsi@polytechnique.edu ☎ (+33) 775-84-06-12 ⚲ AbdelilahYounsi

Skills

Programming & Big Data: Python, SQL, C++, Java, PostgreSQL, MongoDB, PySpark, Elasticsearch, Apache Hadoop
ML&AI: scikit-learn, XGBoost, PyTorch, Optuna, Unsloth, HG Transformers, CrewAI, LangChain, Ollama
MLOps: Snowflake, Databricks, AWS, Azure ML, MLflow, Airflow, Docker, Git, Kubernetes, CI/CD
ML & Statistics : Fine-tuning, RAG, Agentic AI, Computer Vision, Time Series, Statistical Analysis, Hypothesis Testing
Soft Skills : Agile/Scrum methodology, Analytical mindset, Rigor, Autonomy, Teamwork, Communication and presentation

Education

l'X Ecole Polytechnique, Paris-Palaiseau, France *M2 Data Science* *Sept 2024 – Sept 2025*
- **Relevant Coursework:** Computer Vision, Optimization for data science, Generative models, Graphs Machine Learning, Causal inference, Big Data, Cloud architectures...

EMINES, Mohammed VI Polytechnic University, Morocco *Engineering Degree in* *Sept 2021 – Sept 2024*
Industrial Management, Data Science Minor
- **Relevant Coursework:** Probability Theory and Statistics, Intensive Data-science courses, Macroeconomics and Finance

Lycée CPGE Ibn Timiya, Marrakech, Morocco *Sept 2019 – June 2021*
Classes Préparatoires aux Grandes Écoles (CPGE)

Experience

IDEMIA PUBLIC SECURITY - *Iris Recognition* *Courbevoie, France*
Libraries and Skills: PyTorch, PyTorch Lightning, MLflow, Optuna, Accelerate *April 2025 – Oct 2025*
- Trained a **Vision Transformer** model for iris-based biometric identification, reducing comparison time versus traditional methods and achieving competitive performance on **open-source** datasets, with a **0.26% false rejection rate**.
- Fine-tuned the model using a **DINO-style self-distillation** approach, enhancing embedding quality by pulling representations of irises from the same individual closer in latent space, thereby improving recognition accuracy.

ARDIAN FRANCE - *Cash Flow Forecasting* *Paris, France*
Libraries and Skills: Databricks, sklearn, Pandas, PySpark, PyTorch Forecasting, Time Series *Jan 2025 – March 2025*
- Developed a hybrid modeling framework combining a temporal fusion transformer (TFT) with the Takahashi-Alexander financial model to **forecast private equity fund cash flows**, capturing both temporal dynamics and economic fundamentals.

CLEVERLYTICS - *Iris Recognition* *Ben Guerir, Morocco*
Libraries and Skills: PyTorch, OpenCV, C++, OpenIRIS *April 2024 – Sept 2024*
- Developed a complete iris-based identification pipeline using OpenIris (Worldcoin) **open-source code**, boosting accuracy on public datasets by **60%** through integration of advanced techniques from recent research.
- Reduced **False Rejection Rate** from **50%** to **20%** on the public UBIRIS dataset using a two-stage deep learning approach: an Autoencoder for feature learning and a **Siamese CNN** for iris pair classification.

AI/ML Projects

Multimodal Agentic RAG System *Sept 2025*
Libraries and Skills: CrewAI, Whisper, Gemini API, Milvus, Docker, Streamlit
- Built a voice-enabled RAG system using CrewAI multi-agents to process multimodal data through Whisper transcription, Gemini API, and Milvus vector database for intelligent query responses.

AI Agent-Powered Flight Search Application [GitHub ↗] *June 2025*
Libraries and Skills: CrewAI, Selenium, Streamlit, BeautifulSoup, Gemini CLI, LLM
- Developed an AI-powered flight search tool using a CrewAI multi-agent system and Selenium for web scraping, enabling automatic flight retrieval, comparison, and recommendation from Kayak via a Streamlit interface.

Fast style transfer with instance normalization [GitHub ↗] *May 2025*
Libraries and Skills: PyTorch, Git
- Developed a real-time neural style transfer system using a CNN autoencoder for styled image generation, optimized with perceptual losses (content + style) and instance normalization to reduce batch artifacts.

Fine-tuning a Small Language Model for Summarization [GitHub ↗] *April 2025*
Libraries and Skills: Unsloth, Hugging Face Transformers, AutoAWQ, PEFT, LLM, NLP
- Fine-tuned Qwen 2.5-0.5B on CNN/DailyMail dataset using LoRA and 4-bit quantization, implementing efficient batch processing for summary generation and ROUGE-based evaluation within Colab's resource constraints.