





# Maria Afara

ML Engineer

 portfolio  mariaafara  mariaafara  maria-afara5@hotmail.com  650082715

## EXPERIENCE

### WOLTERS KLUWER | ML ENGINEER

October 2023 – Present | Paris, France

- Developed a new feature for analyzing legal documents leveraging Azure OpenAI LLMs and Retrieval-Augmented Generation (RAG) systems, enhancing the application's ability to provide precise answers to legal queries.
- Built services for document parsing, chunk storage in a vector database, and question answering by retrieving top-k chunks and generating answers.
- Used Langfuse to monitor the performance of the RAG system, storing user requests in PostgreSQL for comprehensive tracing and analysis.
- Designed and managed microservices architecture using RabbitMQ for inter-service communication, Kubernetes, Helm charts, Terraform and leveraging Azure cloud services for infrastructure management.
- Orchestrated Docker-based deployments and automated CI/CD pipelines with GitHub Actions.
- Collaborated closely with project managers and software engineers to ensure alignment of development and deployment processes.

### ACOLAD | NLP ENGINEER

March 2021 – October 2023 | Paris, France

- Developed internal text-processing libraries for neural machine translation models, enhancing text processing efficiency and accuracy.
- Built and automated data pipelines with Airflow, enabling seamless integration with ML models and APIs.
- Containerized and deployed ML models on Google Cloud Platform using Docker, GitHub Actions, and Kubernetes, improving deployment speed and reproducibility.
- Designed and optimized machine learning models using Keras, scikit-Learn, Spacy, and PyTorch, achieving significant performance improvements.
- Developed a video summarization prototype leveraging OpenAI's LLMs, deployed on an on-premise Kubernetes cluster with a CI/CD pipeline.

### SAP LABS | NLP DATA SCIENTIST

March 2020 – August 2020 | Sophia Antipolis, France

- Researched and developed SOTA techniques for information extraction and sentence-level classification.
- Performed weak supervision to label unlabelled data.
- Designed and tested topic classifiers using transfer learning on a confidential French dataset.
- Created custom Named Entity Recognition models using spaCy.
- Used PySpark SQL to process large datasets from Wikipedia and DBpedia.
- Built a Relation Extraction pipeline to identify sensitive information as per GDPR.

## SKILLS

Python • Pandas • SpaCy  
NLTK • SparkNLP • Keras  
TensorFlow • Scikit-learn  
GenAI • RAG • HuggingFace  
Airflow • FastAPI • Docker  
GitHub Actions CI/CD • GCP  
AZURE • Terraform  
Kubernetes • Helm Chart  
PySpark • Git • SQL

## EDUCATION

### UNIVERSITÉ DE LORRAINE





MASTER 2 IN NLP

### LEBANESE UNIVERSITY

MASTER 1 IN AI AND BIG DATA

BACHELOR'S IN COMPUTER SCIENCE



## CERTIFICATION

Advanced Deployment  
Scenarios with TensorFlow   
Natural Language Processing  
Specialization   
TensorFlow Developer  
Specialization   
Deep Learning Specialization 

## LANGUAGES

Arabic	Native
English	Fluent
French	Beginner

## PROJECTS

Dungeon Map Generator   
Document Search System   
Question Answering System 