

Maria Afara

ML/NLP Engineer

 portfolio  mariaafara  mariaafara  maria-afara5@hotmail.com  650082715

EXPERIENCE

ACOLAD | NLP ENGINEER

March 2021 – Present | Paris, France

- Developed and maintained internal text-processing libraries to support the development of neural machine translation models, improving efficiency and accuracy of text processing tasks.
- Built and managed data pipelines using Airflow, automating data processing and enabling seamless integration with machine learning models and APIs.
- Containerized machine learning models and APIs using Docker and deployed them to Google Cloud Platform, leveraging GitHub Actions CI/CD for automated deployments, resulting in faster release cycles and improved reproducibility.
- Developed and implemented an automated deployment process for Kubernetes applications using GitHub Actions and Helm Chart, resulting in a significant increase in deployment speed and efficiency.
- Designed, developed, and optimized machine learning models and deep learning architectures using Keras, scikit-Learn, Spacy, and PyTorch in CPU/GPU environments, resulting in significant improvements in model performance and accuracy.
- Resolved business challenges, including classifying domains of translation projects using classification models, identifying personal information based on GDPR using NER models, and automating the creation of neural machine translation models.
- Developed a video summarization prototype using OpenAI's LLMs (GPT-4 and ChatGPT) for transcription, summarization, and translation, and deployed it on an in-premise Kubernetes cluster through a comprehensive CI/CD pipeline with GitHub Actions workflows and Helm charts for efficient management.

SAP LABS | NLP DATA SCIENTIST

March 2020 – August 2020 | Sophia Antipolis, France

- Researched & developed SOTA techniques for information extraction & sentence level classification.
- Performed weak supervision approach to label unlabelled data.
- Designed, developed, and tested several topic classifiers using transfer learning for confidential French dataset.
- Created custom Named Entity Recognition models using spaCy.
- Leveraged PySpark SQL to process and join large datasets of real people biographies extracted from wikipedia dumps and DBpedia.
- Built a Relation Extraction pipeline that combines Named Entity Recognizer & Relation Classifier to identify the sensitive information according to GDPR.
- Mostly worked with Keras, Tensorflow, Spacy, Pandas, Numpy, Matplotlib, Sklearn, mBert, CamemBert, Snorkel, PySpark, BiLSTM, CNN, FastText, SPARQL, Git, Docker.

SKILLS

Python • Pandas • SpaCy
NLTK • SparkNLP • Keras
TensorFlow • PyTorch
Scikit-learn • HuggingFace
Airflow • FastAPI • Docker
GitHub actions CI/CD • GCP
Kubernetes • Helm Chart
PySpark • Datadog • 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 