

MOUHAMED MBAYE

NLP Engineer

mouhamedmbaye371@gmail.com [+33 7 64 28 43 88](tel:+33764284388) [Montpellier, France](https://www.google.com/maps/place/Montpellier,+France)

[linkedin.com/in/mouhamed-mbaye-nlp-tal](https://www.linkedin.com/in/mouhamed-mbaye-nlp-tal) github.com/M-mbaye30



PROFIL

Spécialisé en traitement automatique des langues (NLP), je suis passionné par l'IA et la structuration d'informations textuelles complexes. Fort d'une expérience dans le développement de pipelines NLP robustes, l'intégration de LLMs et de systèmes RAG, ainsi que la conception d'interfaces interactives, je cherche un poste d'Ingénieur NLP à partir de septembre 2025. Mes compétences techniques en Python, Pytorch, spaCy, Transformers (Hugging Face), BDD SQL/NoSQL, et mon expertise en annotation, fine-tuning, extraction d'information et recherche sémantique me permettent de contribuer efficacement à des projets innovants en IA. Curieux, rigoureux et orienté solution, je m'investis pleinement dans la documentation technique et la veille scientifique.

EXPÉRIENCES

Spécialiste NLP

Jan-Jul 2025

IMGT (IGH, CNRS), Montpellier

- Développement d'une solution NLP pour l'extraction automatique d'entités dans des textes scientifiques
- Implémentation et fine-tuning de modèles Deep Learning (BiLSTM-CRF, BERT, GLiNER, scispaCy)
- Création d'une interface utilisateur interactive avec Streamlit incluant visualisation des entités, filtrage dynamique et export CSV.
- Déploiement via Streamlit Community Cloud

Data Scientist - NLP

Jan-Juin 2024

Université Mohammed V, Rabat

- Modélisation et classification automatique multilingue de documents
- Prétraitements : annotation, extraction de connaissances et structuration
- Implémentation et évaluation comparative de modèles LSTM et BERT

COMPÉTENCES

Langages & BDD :

Python, SQLite, SQL, MySQL, MongoDB, ChromaDB

NLP & IA :

spaCy, Hugging Face, scikit-learn, TensorFlow, PyTorch, OCR, Label Studio, RAG, LLMs

Outils & Frameworks :

pandas, matplotlib, Streamlit, Flask, Git/GitHub, LangChain, NLTK, requests, Docker

Méthodologies :

Annotation, Fine-tuning, OCR

Qualités professionnelles:

Adaptabilité, rigueur, capacité à comprendre les besoins métier, communication claire avec les équipes, Veille technologique, innovation continue, Gestion de projets NLP.

FORMATIONS

Master 2 en Traitement Automatique des Langues (TAL)

Université Marie & Louis Pasteur, Besançon, France

Juin 2025

Master 2 Data Science & NLP

Université Mohammed V, Rabat, Maroc

Juin 2024

LANGUES

Français

Courant

Anglais

Bien

Arabe

Courant

Wolof

Maternel

PROJETS

Classification Automatique et Analyse de sentiments

Développement d'un système robuste d'analyse de sentiments et de classification de revues de produits multilingues (Word2Vec, PCA, KNN, RL)

IA Générative et LLM

Implémentation Phi-3.5, Llama 3. Stratégies prompt engineering (zero-shot, few-shot). Intégration Ollama pour inférence locale.

Stack : Ollama, Transformers, Prompt Engineering, Phi-3.5, Llama 3

Système RAG

Développement d'un système RAG utilisant Llama 3.2 et une base de données vectorielle ChromaDB. Ce système est spécialisé dans l'analyse de documents budgétaires sénégalais, offrant une génération de réponses enrichie par la récupération d'informations spécifiques. L'application extrait, indexe et permet d'interroger le contenu des rapports PDF officiels via une interface web intuitive..

Stack : Python 3.8, LangChain, Llama 3.2, Ollama, ChromaDB, PyPDF2, Sentence-Transformers (all-MiniLM-L6-v2), Streamlit

Application NER Streamlit

Développement d'une application permettant d'identifier et d'extraire automatiquement des entités nommées

Stack : Python 3.9, Streamlit, spaCy, Pandas, Matplotlib, Pytorch, hugging face labelstudio

DataLake Explorer

une solution complète de gestion et d'exploration de données. Elle combine un système de traitement de données performant avec une interface utilisateur intuitive développée avec Streamlit. Les utilisateurs peuvent facilement filtrer, rechercher et analyser de vastes collections de documents grâce à un catalogue centralisé et un moteur de recherche avancé (catégorie, date, mots-clés)

Stack : Python 3.9, Streamlit, seaborn PyPDF2, BeautifulSoup

Agent de veille technologique spécialisé en NLP

Un système qui automatise la surveillance, l'analyse et la synthèse des dernières avancées du domaine. Il parcourt le web, utilise des techniques d'extraction pour collecter du contenu pertinent, puis emploie des algorithmes de NLP pour en extraire l'essence et générer des rapports de veille structurés.

Stack : OpenAI API, Web Scraping: bs4, Streamlit

CENTRES D'INTÉRÊT

- Exploration des LLMs agents autonomes
- Veille technologique en NLP et LLMs
- Linguistique computationnelle
- Multilinguisme