



Thiziri SACI

Data scientist NLP

PROFIL

Je suis actuellement stagiaire en data science, passionnée par le Traitement Automatique des Langues (TAL). Je suis à la recherche de mon premier emploi dans ce domaine afin d'apprendre davantage et de contribuer activement, avec mes compétences techniques et mon engagement, au sein d'une équipe dynamique et innovante.

EXPÉRIENCE PROFESSIONNELLE

Data science - stagiaire

Mars 2024 - jusqu'à présent

Malakoff Humanis - site de Chevaleret, Paris

- Intitulé du projet : Génération de contrats et de notices avec l'IA générative.
- Compétences mobilisées / développées :
 - Analyse et préparation des données.
 - Implémentation d'un RAG pour générer du contenu pertinent.
 - Optimisation des performances du modèle par adaptation de prompts.
 - Collaboration avec les équipes métier afin d'identifier leurs besoins, de prioriser les tâches et de s'assurer de l'adéquation des solutions envisagées avec leurs exigences.

PROJETS ACADÉMIQUES

- Mémoire de recherche sur l'évaluation des biais sociaux dans les modèles de langue masqués BERT dans le cadre du projet Your Stereotypical Mileage may Vary : Practical Challenges of Evaluating Biases in Multiple Languages and Cultural Contexts (Fort et al., 2024).
 - Traduction et adaptation du corpus à l'arabe standard.
 - Evaluation de quatre LLMs de la librairie Transformers.
 - Collaboration interdisciplinaire avec l'équipe du projet afin de relever les défis inhérents à la constitution du corpus de l'étude.
- Reconnaissance d'Entités Nommées - déterminer le rôle des instances gouvernementales dans les scénarios de la transition énergétique.
 - NER et POS-tagging avec SpaCy et NLTK.
 - GoogleTrans pour l'enrichissement du corpus.
- Analyse de sentiments - corpus des Tweets.
 - Exploration de différents algorithmes de Machine Learning : SVM, Perceptron, Naive Bayes, Random Forest...
- Autres : création d'ontologies, création d'un moteur de recherche avec Solr Apache, annotation automatique de corpus avec GATE, traitement du signal (reconnaissance de la parole, détection de mots-clés dans un flux de parole, synthèse vocale).

+33 7 55 96 60 02

saci.thiziri1999@gmail.com

linkedin.com/in/thiziri-saci/

Île-de-France, France

Disponibilité : octobre 2024

FORMATION UNIVERSITAIRE

Master en Langue et Informatique

Sorbonne Université
2022 - 2024

Licence en SDL

Sorbonne Université
2021 - 2022

OUTILS NLP

- Bibliothèques Python : scikit-learn, Pandas, Gensim, NLTK, SpaCy, NumPy, Pytorch.
- Bases de données : SQL | Neo4j | Cypher
- Ontologies : Protégé.
- Logiciels de TAL : TXM | GATE | Unitex | GREW-match.
- Dataïku | Visual Studio Code | Eclipse

CENTRE D'INTÉRÊT

Lecture | vélo | cuisine

LANGUES

- Anglais - intermédiaire
- Français - bilingue
- Arabe - bilingue
- Kabyle - langue maternelle