



# Perrine QUENNEHEN

**Data Scientist - NLP**

📞 06 65 61 98 38

✉ [perrine.quennehen@gmail.com](mailto:perrine.quennehen@gmail.com)

📍 Pantin (93)

🎂 14/02/1997

🌐 [perrine-quennehen](https://perrine-quennehen.github.io)

🐙 [PerrineQhn](https://github.com/PerrineQhn)

Attirée par les avancées technologiques dans le domaine du Traitement Automatique du Langage, je souhaite mettre à profit mes compétences en Python, Machine Learning et linguistique computationnelle pour contribuer à des projets d'innovation en NLP. Mon esprit d'analyse et ma proactivité me permettent de développer des solutions génératives performantes, tout en assurant la robustesse et l'efficacité des modèles déployés. Disponible à partir de mai 2025, je recherche un stage de fin d'études pour consolider et appliquer mes compétences.

## Expériences Professionnelles

**Ingénieure TAL - Stage** Novembre 2023  
Laboratoire Modyco (CNRS), Nanterre Juillet 2024

- Multiplication par deux de la taille du corpus Naija Syncor grâce à la collecte et l'intégration de nouvelles données.
- Implémentation d'outils de reconnaissance automatique de la parole (ASR) tels que SPPAS pour extraire et analyser les données intonatives, à l'aide de scripts Python.
- Automatisation de l'alignement audio-transcription via une suite de scripts Python pour optimiser le temps d'annotation.
- Création d'un outil pour mettre à jour les données prosodiques dans les fichiers CoNLL-U.
- Utilisation de modèle Whisper fine-tuné pour entraîner sur la détection des pauses.

**Ingénieure TAL - Stage** Juin 2023  
Laboratoire Modyco (CNRS), Nanterre Juillet 2023

- Conception d'un lexique prosodique du Naija.
- Enrichissement des données prosodiques par l'identification des schémas prosodiques (semi-tons, durées, pentes).
- Présentation du lexique dans un article soumis aux conférences LREC-COLING 2024 et TALN.
- Développement d'arbres de décision pour catégoriser les lemmes en fonction de leurs caractéristiques prosodiques, avec automatisation sous Python et expérimentation avec Scikit-Learn.

## Projets

**Extraction de mots-clés TF-IDF vs. LDA** Avril 2024  
Projet académique

- Développement de modèles TF-IDF et LDA pour l'extraction de mots-clés.
- Comparaison des performances avec des métriques comme précision, rappel, et F1-score.
- **Techniques** : Scikit-learn, Gensim, NLTK, Pandas.
- **Soft Skills** : Esprit d'équipe, autonomie, rédaction.

**Évaluation d'OpenNMT (Deep Learning)** Avril 2024  
Projet académique

- Évaluation de OpenNMT avec des corpus lemmatisés et non lemmatisés.
- Analyse des résultats avec les métriques BLEU, METEOR, BERT.
- **Technique** : OpenNMT PyTorch, NLTK, Scikit-learn
- **Soft Skills** : Esprit d'équipe, autonomie, rédaction

## Formations

**Master TAL spécialité** Septembre 2023  
**Ingénierie Multilingue** Décembre 2025  
Inalco, Paris

- NLP avec Python
- Machine & Deep Learning (IA)
- Statistiques et lois de probabilité
- Linguistique
- Web (HTML/CSS/JS)
- Programmation Objet (C++/C#)

**LLCER Chinois spécialité TAL** Septembre 2020  
Inalco, Paris Juillet 2023

## Informatique

- Python : intermédiaire
- Machine Learning : intermédiaire
- Deep Learning : intermédiaire
- Intelligence Artificielle : intermédiaire
- HTML/CSS : intermédiaire
- Git : intermédiaire
- C++ : débutant
- C# : débutant
- SQL : débutant
- Bash : débutant
- SPPAS : débutant
- Praat : débutant

## Bibliothèques Python

- NLTK, SpaCy
- Pytorch, Tensorflow
- Pandas, Scikit-Learn

## Soft Skills

- Persévérance
- Autonomie
- Rigueur
- Curiosité

## Langues

- Anglais : C1
- Chinois : B2