


PERRINE QUENNEHEN

Ingénieure Machine Learning | Data Scientist NLP

Pantin, 93500 | 06.65.61.98.38 | perrine.quennehen@gmail.com |  [perrine-quennehen](#)

PROFIL

Passionnée par les technologies d'IA, machine learning et NLP (LLMs, Deep Learning, Agents), suite à mon stage de fin d'études chez Bouygues Constructions au sein du LabTP, je suis à la recherche d'un poste en CDI en tant qu'ingénieure machine learning ou data scientist NLP, afin de contribuer à des projets innovants dans votre entreprise. Rigoureuse et proactive, mes expériences précédentes mon permises de mettre à profit mes compétences mais de mener également à bien des projets concrets de NLP/ML. Je suis disponible à partir de début décembre 2025.

EXPERIENCES

Stagiaire Ingénieure Machine Learning - Data Scientist NLP

Bouygues Construction | Guyancourt (78)

MAI 2025 – NOV 2025

- **Fine-tuning de XLM-RoBERTa** pour la **recherche sémantique** dans le domaine du BTP :
 - Pipeline de **continued pretraining** (MLM + contrastive learning) sur corpus interne (~1 M de paragraphes) avec **Databricks**, **MLflow** et **Optuna**.
 - Prétraitement avancé (filtrage, masquage de mots-clés, gel de couches, nouveaux tokens), création d'un dataset équilibré selon les domaines BTP
 - Évaluation multi-tâches (masked accuracy, recherche sémantique) et génération de rapports interactifs.
 - Amélioration de 25% de la précision de recherche sémantique sur les documents d'appels d'offre BTP.
 - Optimisation des hyperparamètres avec Optuna réduisant le temps d'entraînement de 30%.

Technologies : Python, PySpark, Databricks, OpenSearch, Plotly, Mlflow, Transformers, Numpy, HTML, Embeddings, Semantic Search, Docker

- **Analyse NLP d'enquêtes clients :**
 - Traitement et analyse de réponses ouvertes d'enquêtes de satisfaction dans le secteur de la construction.
 - Structuration par thèmes, extraction de mots-clés (**TF-IDF**), détection d'outliers par **ACP**
 - Visualisation interactive avec intégration de verbatims augmentant l'efficacité d'analyse de 40%.

Technologies : Python (Pandas, spaCy, numpy (PCA), scikit learn (TF-IDF)), visualisation de données (Plotly)

Compétences : Autonomie, Capacité d'adaptation, Rigueur, Travail d'équipe

Stagiaire Ingénieure NLP

Laboratoire MoDyCo (CNRS) | Nanterre (92)

MAI 2023 – AOUT 2024

- **Augmentation du corpus** Naija Syncor par deux, avec l'utilisation d'alignement automatique des silences et tokens transcrits pour améliorer la couverture des données linguistiques.
- **Automatisation de l'annotation** de fichiers TextGrid afin d'optimiser l'alignement audio-transcription, réduisant le temps de traitement manuel.
- **Conception d'un lexique prosodique** pour le Naija, incluant la catégorisation linguistique via des **arbres de décisions**, structurant les données pour des analyses avancées.

Technologies : Python, SPPAS, Whisper, Praat, Natural Language Processing

Compétences : Autonomie, Méthodologie, Proactivité

FORMATIONS

Master NLP - Ingénierie Multilingue

INALCO, Sorbonne Nouvelle, Paris Nanterre | Paris, Nanterre

SEP 2023 – DEC 2025

LLCER Chinois - spécialité Natural Language Processing (NLP) - Mention Bien

INALCO | Paris

SEP 2020 – JUIN 2023

COMPETENCES TECHNIQUES

- **Natural Language Processing (NLP) :** NLTK, SpaCy, SPPAS, BERT, XLM-RoBERTa, NER, génération de texte, embeddings, recherche sémantique
- **Machine Learning & Deep Learning :** Scikit-learn, PyTorch, TensorFlow, Transformers, Optuna, fine-tuning, continued pretraining, contrastive learning
- **Développement et Programmation :** Python, Git, Bash, C# (notion), C++ (notion)
- **Cloud & Infrastructure :** Databricks, MLflow, PySpark, OpenSearch, Docker
- **Traitement et Analyse de Données :** Pandas, Numpy, Praat, WebRTC-VAD, Whisper, TF-IDF, ACP (PCA)
- **Visualisation :** Plotly, rapports interactifs HTML
- **Langues :** Français (natif), Anglais (C1), Chinois (B2)