

PERRINE QUENNEHEN

Ingénieure Machine Learning | Data Scientist NLP

Pantin, 93500 | 06.65.61.98.38 | perrine.quennehen@gmail.com | [in perrine-quennehen](https://www.linkedin.com/in/perrine-quennehen) | [portfolio](#)

PROFIL

Spécialisée en NLP avec une appétence pour le machine learning et l'intelligence artificielle, je finalise actuellement mon stage de fin d'études chez Bouygues Construction au sein du LabTP. Diplômée en décembre 2025, je recherche un poste en CDI d'ingénieure machine learning ou data scientist NLP. Mes expériences m'ont permis de mener à bien plusieurs projets en NLP et ML avec rigueur et autonomie. Je souhaite rejoindre une équipe où continuer à développer mes compétences tout en contribuant à des projets concrets.

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 **Transformers**, **Databricks**, **MLflow** et **Optuna**.
 - **Prétraitement** avancé (filtrage, masquage de mots-clés, gel de couches, nouveaux tokens), création d'un dataset représentatif selon les domaines BTP
 - **Évaluation multi-tâches** (recherche sémantique, clustering, distance cosinus).
 - **Amélioration de la précision** de recherche sémantique sur les documents d'appels d'offre BTP.
 - **Optimisation des hyperparamètres** avec Optuna réduisant le temps d'entraînement.
 - **Déploiement** du modèle via **OpenSearch**.

Technologies : Python, PyTorch, Transformers, Databricks, MLflow, Optuna, Docker, PySpark, Plotly, OpenSearch

Compétences : Fine-tuning, LLMs, Embeddings, MLOps, IA Générative, Optimisation, Travail d'équipe

- **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.

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

Compétences : Autonomie, Capacité d'adaptation, Rigueur

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)

INALCO | Paris - Mention Bien

SEP 2020 – JUIN 2023

COMPETENCES TECHNIQUES

- **Machine Learning & Deep Learning :** PyTorch, TensorFlow, Scikit-learn, optimisation (Optuna), PCA
- **IA Générative & NLP :** Transformers (Hugging Face), BERT, XLM-RoBERTa, embeddings, contrastive learning, fine-tuning, continued pretraining, semantic search, génération de texte, spaCy, NLTK, TF-IDF
- **MLOps & Production :** MLflow, Docker, Databricks, PySpark, OpenSearch
- **Développement & Automatisation :** Python, Git, Bash
- **Analyse & Visualisation :** Pandas, NumPy, Plotly, HTML reports
- **Langues :** Français (natif), Anglais (B1), Chinois (B2)