

# Qi Wang

Data Scientist, NLP Engineer

Torcy, France  
☎ +33 (0)6 47 36 96 20  
✉ wq2021@outlook.com  
📄 wq2021.github.io

*Curieuse, rigoureuse avec l'esprit d'organisation et d'analyse acquis à travers mon parcours professionnel et académique. Titulaire du double master en Machine Learning for Data Science et Traitement Automatique des Langues, je suis passionnée par les techniques appliquées aux données textuelles.*

## Expériences professionnelles

- Février à Août 2022 **Stagiaire en Data Science**, Astek, Boulogne-Billancourt, France.
- Construire un système de recommandation d'articles techniques français personnalisé aux spécialistes du domaine en utilisant les techniques récentes en réseaux de neurones, notamment les modèles transformers.
  - Proposer un protocole d'évaluation pour le système construit au niveau de la pertinence et du classement.
  - Rédiger des documents techniques et des rapports de travaux.
  - Outils utilisés : Python(PyTorch, Tensorflow, Pandas, Selenium, BeautifulSoup), Markdown, GitHub,  $\LaTeX$ , etc.
- Avril à Octobre 2021 **Stagiaire en Text Mining**, BNP Paribas, Paris, France.
- Développer une fonctionnalité du plate-forme interne : détecter les faits d'actualité parmi un ensemble d'articles de presse du domaine économique et financier à partir d'une veille documentaire interne.
  - Implémenter et comparer trois différentes approches sur les propriétés lexicales (entités nommées, lemme), la statistique du corpus (IDF) ainsi que les réseaux de neurones (SBERT).
  - Outils utilisés : Python(Pandas, SpaCy, lxml), Excel,  $\LaTeX$ , etc.

## Études et formation

- 2021–2022 **Master Machine Learning for Data Science**, Université Paris Cité (ex. Université Paris-Descartes), Mention Bien, Supervised learning, Unsupervised learning, Deep learning, Text mining, Big data, Système de recommandation, Data visualization, Séries temporelles.
- 2019–2021 **Master Traitement Automatique des Langues (TAL/NLP) parcours Ingénierie Multilingue**, Inalco, Mention Bien, Algorithmique, Traitement de corpus, Text mining, Machine learning appliqué aux données textuelles, Ingénierie des connaissances, Base de données, Linguistique avancée.

## Récompense

- Juin 2023 Lauréat du HackaTAL (hackathon en Traitement Automatique des Langues) pour la mission de détection de citations dans un corpus de contenus du web.

## Certificat

- IBM Data Science Professional Certificate
- DeepLearning.AI & Stanford Online : Machine Learning Specialization
- DeepLearning.AI & AWS : Generative AI with Large Language Models
- Python3 : des fondamentaux aux concepts avancés du langage

## Compétences informatiques et outils

- Maîtrise Python (Pandas, Tensorflow, PyTorch, Scikit-learn, Matplotlib, Seaborn, SpaCy, NLTK, etc), SQL, HTML5, CSS3, XML, VS Code, MacOS, Linux, Microsoft Office, iWork
- Intermédiaire R, Shell, Git,  $\LaTeX$ , Praat, Power BI, Dataiku
- Notions Java, PHP, JavaScript

## Langues

- Natif **Mandarin**
- Avancé **Français (DALF C2), Anglais (B2)**