

# Fang ZHAO

fang.zhao@etu.u-paris.fr | 07.52.59.53.93 | site personnel

 mehechiger  mehechiger-zhao  0009-0002-8583-6642  
Docteur en linguistique computationnelle spécialisé en NLP/IA

## EXPÉRIENCES PROFESSIONNELLES

### LEAD DÉVELOPPEUR NLP/IA | BONNE LECTURE

2025 — présent | à distance

- Transition vers un système RAG (LangChain, RAGFlow).

2020 — 2021 | Hangzhou, Chine

- Développement d'une appli mobile Q&A pour apprenants de FLE.
- Méthodes : moteur de recherche neuronal, multilingual NLP.
- Gestion du projet ; collaboration avec l'équipe de collecte de données.

### STAGIAIRE EN R&D NLP | DEMIUM

2020 — 2021 | Madrid, Espagne (à distance)

- Développement d'un modèle neuronal d'extraction d'informations (IE) pour des articles scientifiques.
- Méthodes : heuristiques + NER neuronal.
- Collecte de données automatique via web scraping.
- Collaboration du projet via méthode agile (Jira).

## PROJETS

### NLP/IA FRUGAL | TENSORFLOW • PYTORCH • HUGGING FACE

Doctorat Linguistique Computationnelle

- Problématique : rareté de données d'entraînement.
- Entraînement de modèles avec des données disjointes.
- Méthodes explorées : apprentissage semi-supervisé, apprentissage par renforcement, adaptation de domaine, finetuning de LLM.

### AUTO-CORRECTION | TENSORFLOW • KERAS

Master 2 Linguistique Informatique

- Entraînement d'un analyseur (morpho-)syntaxique/sémantique capable de s'auto-corriger au cours du processus d'analyse.
- Études 1) des corrections effectués par le modèle et 2) de l'impact de l'auto-correction sur la performance du modèle.

### RÉSUMÉ AUTOMATIQUE | PYTORCH

Master 1 Linguistique Informatique

- Développement from-scratch d'un modèle de résumé automatique neuronal end-to-end avec des LSTMs seq2seq.
- Études de l'impact de l'intégration des informations (morpho-)syntaxiques sur la performance du modèle.

### RMNEWS | PYTHON

- Service de fil d'actualité synchronisé pour la tablette reMarkable.
- Déploiement sur Amazon AWS.

## PARCOURS

### DOCTORAT LINGUISTIQUE COMPUTATIONNELLE

2021 — 2025 | Université Paris Cité (Paris 7) | LLF | Paris, France

### MASTER LINGUISTIQUE INFORMATIQUE

2019 — 2021 | Université Paris Cité (Paris 7) | Paris, France

### LICENCE LINGUISTIQUE INFORMATIQUE

2018 — 2019 | Université de Paris (Paris 7) | Paris, France

## COMPÉTENCES

### PROGRAMMATION

Python • Java • R • JavaScript • Agile

### DATA & OPÉRATIONS

MySQL • Git/Github • AWS • GCP  
Azure • Docker • Linux/Bash • Pandas  
Jupyter Notebook • Pycharm • XML • JSON

### MACHINE LEARNING

Scikit-learn • Numpy • SciPy  
MLP • SVM • KNN • MRF/CRF  
Mean shift • K-means • SGD

### DEEP LEARNING

PyTorch • TensorFlow • Keras  
Transformer • Reinforcement learning  
RNN/LSTM • CNN • Calcul GPU

### NLP

RAGFlow • LangChain  
Hugging Face Transformers  
Stanza • SpaCy • NLTK  
BERT • GPT  
Large language models (LLMs)  
Pretraining • Finetuning  
Reconnaissance d'entités nomées (NER)  
Extraction d'information (IE)  
Résumé automatique  
Moteur de recherche  
Traduction automatique  
Speech recognition  
Graphe de connaissances  
Analyse syntaxique/sémantique

### FRONT-END

Vue.js • Flutter • Swift • HTML

## PUBLICATIONS

Zhao, F. & Bernard, T., TALN 2024

Zhao, F. & Bernard, T., TALN 2023

Zhao, F., RECITAL 2022

## LANGUES

Langue Maternelle    Communication  
Mandarin              Italien • Russe

Professionnel        Notions  
Anglais                Cantonais • Wu  
Français              Japonais