



# Marius Roger

Massy 91300, France  
marius.roger@outlook.fr  
linkedin.com/in/marius-roger

## Ingénieur Machine Learning

Je cherche un premier emploi en Data Science, centré sur le ML Ops ou la R&D.

### Expérience professionnelle

Avril 2024

Septembre 2024

#### Ingénieur ML R&D stagiaire *GitGuardian*

- Contextualisation de secrets (incidents de cybersécurité) par NLP (HuggingFace, llama.cpp, AWS EC2)
- Construction d'un jeu de données de secrets génériques (Prodigy, Snowpark, DVC, OpenAI)
- Contribution au moteur de détection de secrets (Python, git, PDM, Linear)

Octobre 2022

Juin 2023

#### Assistant de recherche *LMU Munich*

- Analyse et comparaison de nombreux LLMs français et allemands (HuggingFace)
- Écriture et revue d'articles de recherche sur les techniques de NLP et leurs biais : [article](#)
- Relecture du cours et correction de TP pour la classe Deep Learning for NLP

Avril 2022

Juillet 2022

#### Ingénieur ML R&D stagiaire *HrFlow.ai*

- Bibliographie et expériences sur la recommandation, les réseaux de graphes et les LLMs
- PoC pour relier des candidats et des offres d'emploi (PyTorch, HuggingFace)

Mars 2021

Août 2021

#### Stagiaire de recherche en NLP *Dauphine PSL*

- Conception de modèles récurrents (LSTM & GRU) basés sur Neural ODE (PyTorch)
- Entraînement du modèle sur le PennTreeBank dataset et évaluation de sa perplexité

### Études

2019 – 2024

#### Diplôme de l'ENS Paris-Saclay

École d'excellence de recherche scientifique, études en informatique et science des données

2023 – 2024

#### Master 2 IASD *Dauphine PSL, Mines Paris, ENS PSL*

Machine Learning, extraction et stockage de données, NoSQL et calcul distribué

- Projet Data Science : Systèmes de recommandation et GANs avec NumPy, PyTorch et git
- Projet BDDs : Extraction de collaboration entre chercheurs depuis DBLP avec XML SAX et Neo4j
- Projet NoSQL : Implémentation de iSAX (indexation de séries temporelles) en Python

2021 – 2022

#### Master 2 MVA *ENS Paris-Saclay, Institut Polytechnique de Paris*

Machine Learning, NLP, analyse des séries temporelles, traitement du signal/audio

- Analyse d'articles : NLP, traitement du signal, séries temporelles, réseaux de graphes, biostatistiques

2020 – 2021

#### Master 1 MPRI *ENS Paris-Saclay*

Vision par ordinateur, deep learning, génie logiciel, optimisation, chaînes de Markov

- Projet Génie Logiciel : Programme de synthèse/traitement de l'audio modulaire pour Linux en C++

2017 – 2019

#### Classes Préparatoires aux Grandes Écoles *Lycée Camille Guérin*

Préparation intense en maths, physique et informatique pour les concours des Grandes Écoles

### Compétences

#### Programmation

Python (PyTorch, Pandas, Transformers), Scala (Spark), SQL, OCaml, Bash

#### Outils

git, SSH, LaTeX, Notion, DVC, Linear, Snowflake

#### Relationnel

Communication, Organisation, Flexibilité, Autonomie, Capacité de synthèse, Curiosité

#### Langues

Anglais (C2), Espagnol (B2), Français

#### Autres intérêts

Production musicale, électronique pour l'audio, handball