

**YCH**

**Data Scientist / AI Engineer**

**Machine Learning – Generative AI – Python**

7 ans d'expérience

## **DOMAINES DE COMPETENCES**

---

- Retrieval-Augmented Generation (RAG).
- Évaluation automatique de modèles génératifs.
- Mise en place de pipelines de traitements de données
- Modélisation Machine / Deep Learning.
- Exploratory Data Analysis (EDA).
- Traitement et modélisation de séries temporelles.
- CI/CD (intégration et déploiement continus), Dockerisation de projets.
- Gestion de projet de bout en bout.

## **COMPETENCE TECHNIQUES**

---

Python, SQL

AWS Certified Cloud Practitioner (CLF-C02) | Amazon Web Services

Docker, GitHub / GitLab

PyTorch

Scikit-Learn, PyTorch, Pandas, Numpy, Langchain, LlamaIndex, HuggingFace Docling, Transformers, Openai, Anthropic, Matplotlib, Streamlit, OpenCV, Yolo

## **FORMATION**

---

**Mastère Spécialisé : AI Expert Data & MLOps – Ecole Télécom ParisTech**

**Master 2 : Ingénierie Mécanique et Modélisation Numérique – Université Lyon 1**

## **LANGUES**

---

**Anglais : Courant**

## **CERTIFICATIONS**

---

AWS Certified Cloud Practitioner (CLF-C02) | Amazon Web Services

Fundamentals for Data Science | Institut Mines-Telecom.

## EXPERIENCE PROFESSIONNELLE

09/2024 – 09/2025 (12 mois)

INSTITUT LOUIS BACHELIER

*Data Scientist / AI Engineer*

- **Développement d'une solution d'IA générative permettant de répondre à des questions à partir de documents fournis.**
- Implémentation du RAG avec orchestration des différentes étapes via LangChain
- Mise en place d'une base de données Single vector store FAISS
- **Elaboration du Dataset d'évaluation**
- Mise en place d'un mode opératoire d'annotation de réponse et collaboration avec des analystes en finance pour l'élaboration d'une base de donnée pour l'évaluation du RAG.
- Mise en place d'un système de contrôle qualité des annotations.
- **Automatisation de l'évaluation**
- Mise en place d'une méthode d'évaluation automatique des réponses générées par le modèle RAG, en utilisant un LLM comme juge de la pertinence des réponses.
- **Optimisation du système RAG**
- Intégration de modèles OCR via Docling pour l'extraction précise de tableau à partir de documents.
- Conversion automatisée des documents en format Markdown avec Docling afin de faciliter le traitement.
- Optimisation du découpage (chunking) des documents pour améliorer la pertinence et la rapidité de la recherche.
- Mise en place d'un système de recherche hybride combinant 75% d'analyse sémantique et 25% d'analyse lexicale pour un retrieval performant.
- **Tests de différents modèles LLM**
- Réalisation de tests comparatifs sur différents modèles LLM, incluant Llama 3 et Nemo via Huggingface
- Évaluation des performances du modèle OpenAI GPT-4o à travers une intégration API.
- Expérimentation avec le modèle Claude Sonnet 3.5 via API pour analyser ses capacités.
- **Rédaction d'article scientifique**
- Rédaction et publication d'un article scientifique, accessible via <https://doi.org/10.48550/arXiv.2505.22752>.

**Environnement :** Python, Langchain, Transformers, FAISS, Pandas, Docling, HuggingFace, Openai, Anthropic, GitLab.

09/2023 – 07/2024 (10 mois)

VEOLIA

*Data Scientist*

- Réalisation d'une analyse exploratoire statistique pour comprendre les comportements de consommation et identifier les anomalies.
- Imputation des données manquantes afin d'assurer la cohérence et la qualité des séries temporelles.
- Génération de données synthétiques pour enrichir l'ensemble d'entraînement et améliorer la robustesse des modèles.
- **Modélisation de la détection d'anomalies**

- Mise en place d'un modèle de détection d'anomalies basé sur un algorithme de machine learning non supervisé Isolation Forest.
- Évaluation des performances du modèle à l'aide de métriques adaptées et optimisation des hyperparamètres pour améliorer la détection des fuites.
- **Estimation du volume d'eau perdu**
- Prédiction des séries de consommation d'eau en l'absence d'anomalies à l'aide du modèles ARIMA
- Estimation du volume d'eau perdu en comparant les séries prédites aux données réelles contenant des fuites.

**Environnement :** Python, Scikit-Learn, Statmodels, Pandas, Numpy, GitLab.

---

**09/2017 – 09/2023 (6 ans)**

**FRAMATOME ( ex AREVA)**

***Ingénieur en Data science et modélisation numérique***

- **Développement d'une solution IA pour la détection de motifs sur images / vidéos**
- Elaboration du Dataset d'entrainement par des annotations en utilisant Labelimg
- Mise en place d'un modèle de computer vision YOLO basé sur des réseaux de neurones convolutifs pour la détection de motifs sur vidéo, afin d'assister le contrôle visuel en centrales nucléaires.
- **Maintenance Prédictive des Capteurs par Machine Learning**
- Réalisation d'une analyse exploratoire des données (EDA) et mise en place d'un pipeline complet de traitement de données
- Modélisation en machine learning pour la prévision du moment de défaillance, dans le but d'anticiper les pannes et optimiser la maintenance.
- **Développement d'application python pour le traitement du signal et l'imagerie ultrasonore**
- Conception et développement d'une application Python dédiée au traitement du signal et à l'imagerie ultrasonore, réalisée from scratch.
- Mise en place d'une chaîne CI/CD intégrant développement, tests, documentation et gestion de version.
- **Optimisation des méthodes de contrôles non destructifs par simulation numérique**
- Développement et mise en place de modèles de calculs numériques multiphysiques
- Proposition d'idées d'optimisation, réalisation de tests et validation des solutions innovantes.
- **Réalisation d'une veille technologique pour alimenter l'innovation et orienter les choix techniques.**
- Surveillance spécifique des avancées en data science, intelligence artificielle et simulation numérique
- **Gestion de projets**
- Gestion de projet en environnement international : définition des besoins, budget, risques, planification, développement et validation de solutions.
- **Invention et dépôt de brevet**
- Invention d'une méthode innovante améliorant le contrôle non destructif par ultrasons, renforçant la précision et la fiabilité des inspections.
  - Dépôt d'un brevet français et d'un brevet international relatif à une innovation dans le contrôle non destructif.

**Environnement :** Python, YOLOv5, Labelimg, Scikit-learn, OpenCV, Pytorch, SciPy, Numpy, Pandas, Sphinx, GitHub