

Calixte HENRY

Data Scientist

github.com/calhenry · calixte.henry@protonmail.com · linkedin.com/calixte-henry · +33632694310

Formation

Expert en ingénierie et science des données (*OpenClassrooms*) 2024 - 2025

- 9 projets professionnalisants de machine learning

Master 2 Development Economics (*Université Paris 1 Panthéon Sorbonne*) 2022 - 2023

- Analyse économique appliquée au développement : évaluation de politiques, enjeux de mondialisation et développement durable. Économétrie, théorie économique
- Traitements BigData

Master 1 Économie Quantitative (*CY Cergy-Paris Université*) 2021 - 2022

- Économétrie, évaluation de politiques publiques, mathématiques, séries temporelles

Expérience

Projets professionnalisants (*OpenClassrooms*)  2024 - 2025



- Développement et déploiement de modèles ML supervisés/non-supervisés avec optimisation d'hyperparamètres
- Segmentation de clients
- **Moteur de classification multimodal** (*image + texte*)
Contexte : Étude de faisabilité d'un système de classification automatique combinant analyse visuelle et textuelle pour la catégorisation de produits e-commerce.
NLP : BERT pour classification de texte avec tokenisation et optimisation des hyperparamètres
Computer Vision : Implémentation VGG16 pré-entraîné avec data augmentation
Résultats : Précision de 95% sur dataset de test, modèle NLP meilleur que multimodal
Librairies : `tensorflow` et `transformers`
- **MLOps** (*versionnage MLflow, déploiement cloud, API, monitoring de drift*)
Objectifs : Prédire la probabilité de faillite d'un client. Analyser les features importantes du modèle pour assurer la transparence des calculs pour les clients. Déployer le modèle à l'aide d'une API
Librairies : `scikit-learn`, `mlflow` et `fastapi`
- **Traitement dans un environnement BigData Cloud** (*AWS, Databricks*):
Contexte : Développement d'un pipeline Big Data pour l'extraction et l'analyse de features d'images à partir de grands volumes de données, en utilisant cluster et des outils cloud.
Librairies : `pyspark`, `tensorflow`

Chargé d'étude et d'évaluation (*Ministère du Travail et de la Santé*)  09/2023 - 08/2024

- Études statistiques, traitement, analyse des données RH.
- Création d'un modèle de prévision de l'âge de départ à la retraite (modèle de survie Cox)


Stage Chargé d'étude de données sociales (*Ministère du Travail et de la Santé*)  05/2023 - 09/2023

- Étude empirique des inégalités de genre et de rémunération au sein du ministère.
Présentation devant la direction.


Assistant de recherche (*LIEPP - SciencesPo Paris*)   12/2022 - 04/2023
05/2022 - 08/2022

- Recherche sur les discriminations de genre et ethniques
- Collecte et traitement des données. Statistiques descriptives

Projets personnels

Wildchat-1M clustering - NLP sur un dataset d'1 million de conversations ChatGPT  11/2025

- Objectifs : Segmenter les conversations par similitude sémantique. Optimiser le pipeline pour un volume de données supérieur à la RAM disponible.
- Bibliothèques : [Duckdb](#), [polars](#), [scikit-learn](#)

job-scraper-llm-vs-tradition - web scrapping et LLM en local  08/2025

- Objectifs : Automatiser l'extraction et le traitement d'offres d'emplois. Tester les performances d'un petit LLM versus les méthodes traditionnelles de traitement des données.
- Scraper des données et rétro-ingénierie d'API publique
- Avec [langchain](#), [crawl4ai](#), [polars](#)

commit-ai - Assistant IA pour git commits  09/2025

- Objectifs : Développer un outil accessible dans le terminal qui génère un draft de commit via LLM local (éditable). Faire une interface CLI
- Avec [ollama](#) pour le LLM en local, [textual](#) et [typer](#) pour l'application CLI

Compétences

Programmation	Python, R, SQL, VBA
Statistiques	Machine learning, Économétrie, Deep learning
Machine learning	Scikit-learn, PyTorch, Tensorflow, MLFlow
Dashboards	Shiny, Streamlit, Marimo
Développement	Git, GitHub Actions
Documentation & Publication	Github-pages, LaTeX, Typst, PackOffice

Langues

- Anglais C1
- Espagnol A2

Centres d'intérêt

- Escalade
- Course à pied
- Cyclisme