



## Vincent Duchauffour

### Machine Learning Engineer

🏠 Paris, France

☎ 06.09.78.66.27

✉ [vincent.duchauffour@proton.me](mailto:vincent.duchauffour@proton.me)

in [LinkedIn](#)

🐙 [Github](#)

## Profile

I work in data/AI field since 2019.

Over the years, I have specialized closely on software engineering aspects of machine learning projects. This includes maintaining and deploying ML librairies for production use, developing and monitoring ML pipelines and building robust CI/CD guidelines.

I'm currently looking towards contributing to open-source software and learning high-performance languages (C++/Rust).

## Skills

- Python, PyTorch
- Linux, Shell scripting
- Docker, GitHub Actions
- Git, DVC, W&B, MLFlow

## Languages

- French (Native)
- English (Conversational)

## Experience

Since Oct. 2022 **Machine Learning Engineer - Kinetix**

- **Design, build and deploy production-level** ML applications in collaboration with the R&D team, this include:
  - **Rewrite** large part of the ML codebase with **strong OOP principles** (SOLID principles / Design patterns / Refactoring)
  - **Set up good coding practices** along the R&D team (code quality, shell scripting, CI pipeline with GitHub Actions, Docker, unit tests, integration tests, packaging, deployment)
  - Split the ML monolith into multiple independent repositories in order to move our projects towards **microservice** approach
  - Write from scratch **ML libraries** for using public pre-trained models and in-house trained and fine-tuned models
- **Be in charge of MLOps** related topics, this involve to design and setup the production environment in accordance with DevOps team for running our ML pipelines (AWS EKS, CD pipelines with Github Actions, experiment tracking and reproducibility, model versioning)
- **Lead and help** R&D engineers adopt good coding practices by conducting multiple workshops on code quality and tooling
- **Reduce by 50% the inference time** on the main ML pipeline (**PyTorch**, profilers, model serving optimization)

Sep. 2021 – Sep. 2022 **Data Scientist - Yanport**

- **Design, build and deploy** an improvement on existing **NLP** system with ordinal classification with focus on explainability (**NLTK**, **Spacy**, **PyTorch**, **HuggingFace**)
- Design, build and deploy **new features** on the home value estimate engine (**CatBoost**), validate these additions by ensuring that they don't have side effects on other features
- Maintain and deploy **data pipelines** and internal scripting

Sep. 2019 – Sep. 2021 **Data Engineer - Qare**

- **Design, build and deploy** a monitoring tool for data and ETLs quality
- **Analytics engineering** (DBT)
- **Contributing to ETLs** and RETLs (MongoDB, AWS Redshift, Salesforce)
- **EDA** and ad hoc analyses, **interact with** concerned **corporate departments** and make **public presentation** of results

## Education

2021 – 2022

**Msc Data Science (IASD)** - Paris Dauphine - *Apprenticeship at Yanport*

2019 – 2021

**Msc Digital Humanities** - CNAM - *Apprenticeship at Qare*