

FLORENT LIN

Data Scientist - Machine Learning & NLP

📍 6 Av. Haroun Tazieff, Bussy-Saint-Georges, France · ✉ lflorentlin@gmail.com · 📞 (+33) 7 81 63 66 63
in linkedin.com/in/florentlin · 🐙 github.com/flin1

Education

ENSAE Paris - IP Paris

Engineering Diploma (Master's Degree) in Data Science & Statistics

Palaiseau, France

Sept 2021 - June 2024

- **Core Coursework:** Deep Learning, NLP, Machine Learning, Bayesian Statistics, Time Series, Optimization.

Lycée Saint-Louis

Preparatory Classes for French "Grandes Écoles" (Major in Physics & Chemistry)

Paris, France

Sept 2019 - June 2021

Professional Experience

Data Science Consultant (on contract at SNCF Réseau)

Aubay Data & AI

Boulogne-Billancourt,

France

Jan 2025 - Present

- Leading Data Science and ML initiatives to optimize predictive maintenance for railway infrastructure.
- Engineering complex data pipelines, processing and harmonizing multi-source datasets.
- Building predictive models with supervised algorithms (XGBoost, Random Forest) to forecast track defects.
- Optimizing models on imbalanced data using SMOTE and tailored metrics (Recall, F1-Score, ROC-AUC).
- Communicating technical results to non-technical teams through reports and operational recommendations.

Data Science Intern - Machine Unlearning

Aubay

Boulogne-Billancourt,

France

Aug 2024 - Dec 2024

- Conducted a state-of-the-art review of Machine Unlearning techniques for LLMs (e.g., Concept Erasure, DPO).
- Fine-tuned a LLaMA model (LoRA approach) and implemented a RoBERTa-based prompt classifier.
- Measured model performance using specialized metrics (Truth Ratio, conditional probabilities, ROUGE score).

Academic Consulting Project - Route Optimization

Deloitte

Lèves, France

Nov 2022 - May 2023

- Modeled patient and nurse constraints to mathematically formalize a complex vehicle routing problem.
- Developed algorithmic solutions to optimize staff schedules under multiple constraints.
- Delivered a final report with strategic recommendations and a functional prototype in Python.

Personal Projects

AI-Powered Matching Algorithm

[global-franchises](#) 🌐


- Designed and architected a multi-agent AI system to automate data extraction and lead qualification. Developed a hybrid matching algorithm combining hard filters, weighted scoring, and semantic search.
- **Stack:** Python, Supabase (PostgreSQL), Vector Embeddings, LLM APIs (OpenAI & Google GenAI).

Large-Scale Text & Image Summarization

2025

- Performed OCR and semantic analysis on hundreds of pages of documents (text, images, diagrams), implementing a feedback loop to improve system reliability and storing data on the cloud.
- **Stack:** Python, Firebase, Supabase (PostgreSQL), LLM & OCR APIs (Mistral AI & Google GenAI).

Online Shopper Purchase Intent Prediction

classification-shoppers 

- Executed feature engineering, implemented various classification algorithms, and tuned hyperparameters.
- Addressed class imbalance to accurately predict online shoppers' purchase intentions.

Skills

Languages: French (Native), English (Professional Proficiency, C1), Mandarin Chinese (Intermediate, B2)

Programming: Python (PyTorch, JAX, Scikit-learn, Pandas), R, SQL, HTML/CSS, JavaScript (basic)

Tech Stack: AWS, Docker, Git, Firebase, Supabase, Vercel, Power BI, Office365, LaTeX