



# Martin LE CORRE

Specialized Master **BIG DATA & AI, Télécom Paris - FRANCE**

As a general engineer currently studying for a master's degree in Big Data & AI at Télécom Paris, I am currently focusing on **RAG & ELK pipeline projects**.

(personal projects ongoing or completed with demonstrators, see details below)

Seeking employment as a Data Engineer or Data Scientist starting in Jan 2026 (worldwide).

## Contact

- martinlecorre@live.fr
- (+33) 06.72.92.51.88
- 28 years old
- Driving license (international)
- [LinkedIn.com/in/martinlecorre/](https://www.linkedin.com/in/martinlecorre/)

## Languages

- French (native)
- English (TOEIC: 840/990)
- Spanish (occasional practice)
- Portuguese & Chinese (basics)

## Skills

**Languages:** Python, Java, R, SQL, PHP, HTML/CSS

**Data Science & ML:** Machine Learning, Deep Learning, Computer Vision, NLP, Transformers

**Python Stack:** Pandas, NumPy, Scikit-learn, Seaborn, Selenium, Django

**Databases:** PostgreSQL, MySQL, Oracle, SQL Server

**Big Data:** Spark, Hadoop MapReduce, Hive, Elasticsearch

**DevOps:** Docker, Kubernetes, Linux

**Tools:** Power BI, Git, Trello, Flutter, Laravel, Symfony

**Methods:** Agile, Project Management

## Engineering

**Simulation:** ANSYS, COMSOL, MATLAB

**CAO:** Catia V5, Solidworks

## Personal & Main Projects



**Year:** 2025

**Status:** personal

**Demonstrator:** ready

**Data:** All courses my courses (flexible data input, multi-format ingestion)

Advanced **RAG system** for confidential or local data processing with on-premise deployment (Ollama) or multi-cloud LLM integration (Claude, GPT...) enabling real-time performance comparison and further cost analysis.

Hybrid semantic search combining dense vector retrieval (PostgreSQL / PGVector, HNSW indexing), sparse BM25, and RRF fusion with cross-encoder reranking.

Automated ingestion pipeline with multi-format parsing (PDF, DOCX, TXT, Markdown, py, ipynb, jpg, mp3...), intelligent chunking, quality filtering, and GPU-accelerated embeddings (BAAI/bge-m3).

### RAG Project Tech Stack & Tools:

RAG/LLM: LangChain • Local LLMs (Ollama) + Cloud (APIs) • PostgreSQL/PGVector • Hybrid search (Vector+BM25+RRF) • NLP pipeline • LangChain • Sentence-Transformers • PyTorch/CUDA • HNSW indexing • CrossEncoder reranking • Whisper • Streamlit UI • Docker • embeddings

- 100% local for sensitive data or hybrid (cloud/local)
- Demonstrator ready, fully deployable (ask Martin for Git)



**Year:** 2025

**Status:** personal

**Demonstrator:** ready

**Data:** (see RGPD for usage)

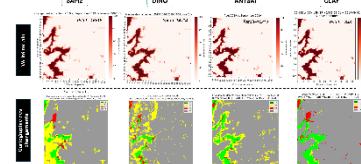
Automated big data collection and analysis platform with **ETL pipeline**, intelligent multi-pass scraping, ML-based gender prediction, and real-time visualization dashboard using Elastic & Kibana.

Containerized infrastructure deployable locally and on cloud servers (Hetzner/Oracle Cloud).

### ETL Tech Stack & Tools:

Big data ETL pipeline • Airflow (DAG orchestration & scheduling) • Docker • Spark/PySpark • Parquet columnar format (Snappy) • Selenium • PostgreSQL • Flask • ML (gender prediction) • Pandas • Jupyter/Notebooks • Docker • Redis (caching) • CI/CD • SQL avancé (optimization queries) • Data quality • Elastic • Kibana • Cloud

- Demonstrator ready
- Deployment-ready infrastructure (Hetzner/Oracle Cloud automated scripts)



**Year:** 2024/2025

**Status:** Academic & confidential

**Partner:** Airbus Defense & Space

**Data:** satellite images

Automated environmental change detection in satellite imagery time series using foundation models.

Benchmark of foundation models: SAM2, DinoV2, Clay, and Anysat for unsupervised anomaly detection and prioritization.

### Tech Stack & Tools:

- Foundation models: SAM2, DinoV2, Clay, Anysat
- Computer Vision: Semantic segmentation, embedding extraction
- Time series analysis
- ML: Unsupervised learning, LOF clustering, anomaly detection
- PyTorch • Multi-spectral imagery
- Dataset: Dynamic Earth Net (Planet Labs)

- Data preprocessing
- Dataset exploration
- Foundation models implementation & benchmarking
- ROC-based performance evaluation showing in-domain models outperform general models for satellite change detection

## EDUCATION & EXPERIENCES

### September 2024 – Ongoing

Télécom Paris (Academic) | Paris – FRANCE  
Specialized Master in Big Data & Data Science



**Description:** Statistics, Advanced Machine Learning, Deep Learning, ML Ops, Distributed Systems for Big Data, DevOps, IoT, AI, Computer Vision, Airbus Academic project, Early-stage ETL Project, NLP, Computer Vision, Time Series, HADOOP, Docker / Kubernetes, Python, JAVA, SQL

### 2023 – 2024

Year off – IKO Kitesurfing Instructor | Sri Lanka, Brazil, Zanzibar, Vietnam

**Description:** International kitesurfing instructor, traveling across 4 countries: Brazil, Sri Lanka, Vietnam and Zanzibar (teaching in English/French/Spanish)

# Associations

## **Student union:**

- 2018, in charge of the organization of a ski trip (600 students in Vars & Risoul)
- 2017, in charge of the organization of events & parties
- 2017, participation of the organization of nautical activities

## **kitesurfing:**

- 2022, member of the FKS0 board, an association responsible for kitesurfing in the Landes region of France

# Hobbies



Kitesurfing, surfing, Bodysurfing, boxing



Sports & trips youtube channel (100k+ views, sponsorships & event)

## 2022 – 2023

Sofren (CMA-CGM) | Marseille – FRANCE

Naval Engineer | R&D Ships Department



**Description:** In charge of the integration studies for systems using the wind to propel CMA-CGM's fleet of ships: Rigid sails, kites, Flettner Rotors studies.

I'm also part of the technical team in charge of optimizing autopilots.

- Technical Assessment of WAPS integration & Autopilots development
- Proposal of new implantation solutions applicable to the specific case of CMA-CGM fleet

## 2021 – 2022

Ailes Marines (Iberdrola) | Saint-Brieuc – FRANCE

Assistant Director Development & territories



**Description:** Ailes Marines is responsible of the development, construction and operation of the Saint-Brieuc Offshore Wind Farm (500 MW).

- Meetings with government departments (Maritime Prefecture, DDTM)
- In charge of the Call for Expressions of Interest (CEI) for the Saint-Malo maintenance ports
- Participation in maintenance port projects (Saint-Malo, SQP, Lezardrieux)

## 2020 – 2021

SOFRESID Engineering | Lorient – FRANCE

Naval Engineer



**Description:** In charge of the integration studies of wind propulsion systems on various commercial vessels (oil tankers, LNG Carriers, and VLCCs).

- Technical and economical assessment of the integration of different WAPS in accordance with class, flag, applicable standards (local and international regulations)

## 2020 – 2020

CorPower Ocean | Stockholm – SWEDEN

Internship | Test bench & Analysis Engineer



- Carrying out dry test campaign prior to development of the prototype at sea

- Matlab programming

- Early stop due to COVID: test bench delivery delay problems

## 2015 – 2020

EPF (Academic) | Paris – FRANCE

General engineering school, Specialization in Energy & Environment



**Description:** French "Grande École" offering generalist engineering education with hands-on projects and international exposure. Specialized in Energy & Environment

## 2018 – 2018

UTM | Johor Bahru – MALAYSIA

Internship | Engineering student



**Description:** The objective of this internship was to determine the best incidence of wave energy converters (WECs) systems to maximize their efficiency

- Matlab programming
- Study of suitable locations for system implementation (offshore)
- Numerical simulations (ANSYS Fluent)