

# Henri Goua

 Portfolio |  GitHub |  LinkedIn |  [gouabeedi@gmail.com](mailto:gouabeedi@gmail.com) | \* Paris, France

## Profile

Graduate in Data Science (ENSAI), with hands-on experience in building **end-to-end ML pipelines**, from modeling to deployment of **Python/FastAPI APIs** and cloud integration. I also have applied knowledge in **Deep Learning** and explainability methods. Currently seeking an opportunity as a **Data Scientist** to leverage these skills in impactful and real-world AI projects.

## Education

**ENSAI — Engineering Degree in Data Science and AI** 2022 – 2025  
Specialization in Machine Learning (PyTorch, TensorFlow), NLP (Transformers, LLMs, RAG), Big Data (Spark, Hadoop). Strong foundations in statistics, optimization, and algorithms.

**Preparatory Classes — Mathematics, Physics & Engineering Sciences** 2020 – 2022  
Intensive training focused on problem-solving, modeling, and analytical rigor.

## Skills

**Programming** : Python (ML APIs, FastAPI), Java, R, SQL (data matching), data structures & algorithms  
**ML & Algorithms** : Deep Learning (LSTM, CNN, PyTorch, TensorFlow), XGBoost, scikit-learn, supervised & unsupervised learning  
**NLP & Search** : Transformers (BERT, BART), RAG, BM25, embeddings, reranking, spaCy, LangChain, Docling  
**Cloud & DevOps** : REST APIs, FastAPI, Docker, AWS EC2, GCP (Cloud Run/Colab), GitHub Actions (CI/CD), ChromaDB, Streamlit, unit testing (pytest)  
**Tools & Practices** : Pandas, Matplotlib, Seaborn, SHAP, LIME, observability/logging, version control (Git)

## Professional Experience

**Intern — Generative AI & RAG** Apr 2025 – Present  
*Square Management — Square Research Center*

- Designed **modular RAG pipelines in Python** for enterprise search; benchmarked 189+ configurations (OCR × chunking × embeddings).
- Implemented hybrid search (BM25 + embeddings) with LLM reranking, improving retrieval quality.
- Automated evaluation with RAGAS, TruLens, and LLM-as-a-Judge (**faithfulness, relevance, traceability**) and delivered reproducible dashboards.
- Containerized services (FastAPI, Docker), CI/CD with GitHub Actions, and deployed a vector database (ChromaDB).
- Delivered a Streamlit prototype with integrated logging for Finance, Marketing, and ESG use cases.

**Intern — Behavioral Analysis & Urban Pollution** Jun 2024 – Sep 2024  
*City of Paris — Mobility Agency*

- Built a **supervised model (scikit-learn)** to identify the most polluting vehicle fleets.
- Integrated a **SQL + Python pipeline** for data matching (SIRENE API + regional datasets), achieving **85%+ accuracy on 30k+ records**.
- Implemented **latency p95 and quality logging** to monitor pipeline performance and robustness.
- Provided recommendations for sustainable mobility and sectoral emission scoring.

## Projects

**FraudTrack360 — Fraud Detection** Jan 2025 – Mar 2025  
*Python, scikit-learn, LSTM, SHAP, FastAPI, Docker, AWS EC2*

Developed an **ML engine** for transactional sequences with explainability. Scoring API deployed via FastAPI+Docker on AWS EC2, AUC-PR = 0.89 (+285k transactions). Interactive monitoring (Streamlit+SHAP dashboard).

**ReviewGuardian — Toxicity Detection** Mar 2024 – May 2024  
*Python, scikit-learn, SHAP, FastAPI, Streamlit, SQL*

Designed an **NLP pipeline** for text classification, evaluated on 50k reviews (accuracy = 91%). Deployed FastAPI API + explanatory interface (Streamlit+SHAP).

**InsightDetector — Hallucination Detection** Dec 2024 – Mar 2025  
*Python, HuggingFace Transformers (BART/T5), LLM, spaCy, Streamlit, OpenAI API*

Built an **end-to-end pipeline** RSS→Summarization→Hallucination Detection. Evaluated on **300+ news articles; annotation interface** in Streamlit; **reproducible workflow**.

**Bayesian Calibration — Lorenz-96 Model** Academic Project  
*Python, NumPy, TensorFlow, ABC-SMC, ABC-Rejection, ABC-MCMC*

Performed **Bayesian inference** to estimate parameters of chaotic systems (Lorenz-96). Compared Bayesian approaches (ABC-Rejection/SMC, MCMC) with **30k simulations/samples; parameter error** 0.25 on synthetic data.

**Hybrid Music Recommender** Jan 2024 – Feb 2024  
*Python, pandas, Surprise, Streamlit*

Designed a **hybrid recommendation engine** (collaborative + content-based) with cold-start fallback strategy. Reproducible pipeline and **interactive demo** deployed with Streamlit.

## Languages

French (native), English (professional)