

Henri Goua

 Portfolio |  GitHub |  LinkedIn |  gouabeedi@gmail.com | * Paris, France

Profil

Diplômé en Data Science (ENSAI), J'ai acquis une bonne expérience dans la conception de **pipelines ML de bout en bout**, depuis la modélisation jusqu'au déploiement d'**APIs Python/FastAPI** et l'intégration cloud. Je possède également des connaissances appliquées en **Deep Learning** et en méthodes d'explicabilité. Je recherche aujourd'hui une opportunité en tant que **Data Scientist**, afin de mettre ces compétences au service de projets IA concrets et à fort impact.

Formation

ENSAI — Diplôme d'ingénieur en Data Science et IA 2022 – 2025
Spécialisation en Machine Learning (PyTorch, TensorFlow), NLP (Transformers, LLMs, RAG), Big Data (Spark, Hadoop). Bases solides en statistiques, optimisation et algorithmes.

Classes Préparatoires — Mathématiques, Physique & Sciences de l'Ingénieur 2020 – 2022
Formation intensive axée sur la résolution de problèmes, la modélisation et la rigueur analytique.

Compétences

Programmation : Python (ML APIs, FastAPI), Java, R, SQL (correspondances de données), structures de données & algorithmes
ML & Algorithmes : Deep Learning (LSTM, CNN, PyTorch, TensorFlow), XGBoost, scikit-learn, apprentissage supervisé & non supervisé
NLP & Recherche : 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, tests unitaires (pytest)
Outils & Pratiques : Pandas, Matplotlib, Seaborn, SHAP, LIME, observabilité/logging, contrôle de version (Git)

Expérience Professionnelle

Stagiaire — IA Générative & RAG Avr 2025 – Présent
Square Management — Square Research Center

- Conception de **pipelines RAG modulaires en Python** pour la recherche d'entreprise ; benchmark de 189+ configurations (OCR × chunking × embeddings).
- Mise en œuvre d'une recherche hybride (BM25 + embeddings) avec reranking LLM, améliorant la qualité de la recherche.
- Évaluation automatisée avec RAGAS, TruLens et LLM-as-a-Judge (**fidélité, pertinence, traçabilité**) et mise en place de tableaux de bord reproductibles.
- Conteneurisation des services (FastAPI, Docker), CI/CD avec GitHub Actions, et déploiement d'un moteur vectoriel (ChromaDB).
- Livraison d'un prototype Streamlit avec journaux intégrés pour les cas d'usage Finance, Marketing et ESG.

Stagiaire — Analyse comportementale & Pollution urbaine Juin 2024 – Sep 2024
Ville de Paris — Agence de la Mobilité

- Construit un **modèle supervisé (scikit-learn)** pour identifier les flottes de véhicules les plus polluantes.
- Intégré un pipeline **SQL + Python** pour la correspondance de données (API SIRENE + jeux régionaux), obtenant une **précision de plus de 85% sur 30k+ enregistrements**.
- Implémenté un système de **logging de latence p95 et qualité** pour monitorer performance et robustesse du pipeline
- Recommandations fournies pour la mobilité durable et le scoring sectoriel des émissions.

Projets

FraudTrack360 — Détection de Fraude Jan 2025 – Mar 2025
Python, scikit-learn, LSTM, SHAP, FastAPI, Docker, AWS EC2

Développement d'un **moteur ML** (séquences transactionnelles) avec explicabilité. API de scoring déployée via FastAPI+Docker sur AWS EC2, AUC-PR = 0.89 (+285k transactions). Monitoring interactif (dashboard Streamlit+SHAP).

ReviewGuardian — Détection de Toxicité Mar 2024 – Mai 2024
Python, scikit-learn, SHAP, FastAPI, Streamlit, SQL

Conception d'un **pipeline NLP** pour classification texte, évalué sur 50k avis (précision = 91%). Déploiement API FastAPI + interface explicative (Streamlit+SHAP).

InsightDetector — Détection d'hallucinations Déc 2024 – Mar 2025
Python, HuggingFace Transformers (BART/T5), LLM, spaCy, Streamlit, OpenAI API

Mise en place d'un **pipeline end-to-end** RSS→Résumé→Détection d'hallucinations. Évalué sur **300+ articles de presse** ; **interface d'annotation** Streamlit ; **workflow reproductible**.

Calibration Bayésienne — Modèle Lorenz-96 Projet académique
Python, NumPy, TensorFlow, ABC-SMC, ABC-Rejection, ABC-MCMC

Réalisation d'une **inférence bayésienne** pour estimer des paramètres de systèmes chaotiques (Lorenz-96). Comparaison d'approches bayésiennes (ABC-Rejection/SMC, MCMC) avec **30k simulations/échantillons** ; **erreur paramètre** 0,25 sur données.

Recommandateur Hybride de Musique Jan 2024 – Fév 2024
Python, pandas, Surprise, Streamlit

Conception d'un **moteur de recommandation hybride** (collaboratif + contenu) avec stratégie de fallback cold-start. Pipeline reproductible et **démo interactive** déployée via Streamlit.

Langues

Français (natif), Anglais (professionnel)