

Beedi Goua

 Portfolio |  GitHub |  LinkedIn |  beedi.goua@eleve.ensai.fr |  +33 779 872 812

Profil

Ingénieur Data Science en début de carrière, passionné par le machine learning, le NLP et l'IA générative. Expérience concrète en conception de pipelines RAG, modélisation prédictive et analyse comportementale. Je souhaite appliquer mes compétences à des cas d'usage réels en développant des solutions robustes, explicables et adaptées aux besoins métier.

Expérience Professionnelle

Stage – IA Générative & RAG appliquée aux cas métiers (Finance, RSE, Marketing) Avril 2025 – Aujourd'hui
Square Management – Centre de Recherche

- Conception d'un pipeline RAG industrialisable pour explorer des rapports PDF métier (300+ pages analysés)
- Optimisation du découpage sémantique et évaluation de la pertinence des réponses via RAGAS et TruLens
- Déploiement d'un prototype interactif avec Streamlit, traçabilité assurée via logs et évaluation automatique
- **Stack** : Python, LangChain, Docling, OpenAI API, ChromaDB, GitHub Actions

Stage – Analyse comportementale & pollution urbaine Juin – Septembre 2024
Ville de Paris – Agence de la Mobilité

- Identification d'entités polluantes via clustering supervisé (cartographie + scoring pollution sectoriel)
- Appariement automatique de bases hétérogènes (API SIRENE, bases régionales) : 95% de taux de match
- Recommandations synthétiques livrées aux décideurs pour adapter la politique de mobilité durable
- **Stack** : Python, scikit-learn, pandas, DigDash, API SIRENE

Formation

Diplôme d'ingénieur – Data Science 2022 – 2025
ENSAI – École Nationale de la Statistique et de l'Analyse de l'Information

- Formation pluridisciplinaire en statistique, modélisation, IA générative et Big Data
- **Cours clés** :
 - GLM, séries temporelles, analyse de survie, calcul stochastique
 - Machine learning, deep learning, NLP, RAG
 - Hadoop, Spark, SQL, POO (Python, Java)

Projets

ReviewGuardian – Détection de toxicité dans les avis avec explication locale
→ *scikit-learn, SHAP, FastAPI, Streamlit*

Modèle MultinomialNB avec explication SHAP, exposé via API FastAPI et interface Streamlit interactive.

Recommandation musicale hybride – Système collaboratif et par contenu
→ *pandas, Surprise, Streamlit, fallback logic, KNNBasic*

Recommandateur hybride combinant KNN (collaboratif) et similarité de contenu, avec fallback utilisateur dynamique.

Calibration bayésienne – Modèle Lorenz-96
→ *Python, NumPy, matplotlib, ABC-SMC, ABC-MCMC*

Inférence bayésienne pour calibrer les paramètres du modèle chaotique Lorenz-96, avec analyse des performances et coûts calculatoires.

InsightDetector – Détection d'hallucinations dans des textes générés
→ *BART, BERTScore, spaCy, Streamlit, RSS, openai, LLM-as-a-judge*

Pipeline de résumé + fact-checking + annotation utilisateur ; détection d'hallucinations sur 300+ articles (actualité, politique) via interface Streamlit, code prévu en open-source.

FraudTrack360 – Détection de fraude transactionnelle explicable
→ *pandas, scikit-learn, LSTM, SHAP, FastAPI, Docker, GitHub Actions, AWS EC2*

Modèle LSTM sur séquences transactionnelles avec scoring d'anomalie ; API conteneurisée, CI/CD GitHub + déploiement EC2 ; tableau de bord explicatif Streamlit.

Compétences

- **Langages** : Python, R, SQL, SAS
- **ML / DL** : scikit-learn, XGBoost, TensorFlow, CNN, LSTM, BERT
- **IA générative / NLP** : Hugging Face, LangChain, OpenAI API, Whisper, BART
- **Engineering** : FastAPI, Docker, Git, CI/CD, REST API
- **Cloud / MLOps** : AWS, GCP, automatisation de pipelines
- **Visualisation** : Matplotlib, Seaborn, ggplot2, Streamlit
- **Explicabilité** : SHAP, LIME

Langues

Français (natif), Anglais (professionnel)