

Beedi Goua

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

Profile

Early-career Data Science engineer passionate about machine learning, NLP, and generative AI. Practical experience in building RAG pipelines, predictive modeling, and behavioral analytics. Eager to apply my skills to real-world use cases through robust, explainable, and business-oriented solutions.

Professional Experience

Internship – Generative AI & RAG for Business Applications (Finance, CSR, Marketing) Apr 2025 – Present
Square Management – Square Research Center

- Designed an industrial RAG pipeline with a benchmark of 189 configurations (7 OCR × 3 chunking × 3 embeddings) on a multi-domain corpus
- Developed hybrid search (vector + BM25), LLM-based reranking and optimized semantic chunking
- Automated evaluation using RAGAS, TruLens and LLM-as-a-Judge (faithfulness, relevance, traceability)
- Deployed a Streamlit prototype with business-specific UI (Finance, CSR, Marketing) and integrated performance logging
- **Stack:** Python, LangChain, Docling, OpenAI API, ChromaDB, Streamlit, GitHub Actions

Internship – Behavioral Analysis & Urban Pollution June 2024 – Sept 2024
City of Paris – Mobility Agency

- Identified the most polluting vehicle fleets using supervised clustering (mapping + sectoral pollution scores)
- Automated matching of heterogeneous datasets (SIRENE API, regional sources) with 95% match rate
- Delivered actionable recommendations to inform sustainable mobility policies
- **Stack:** Python, scikit-learn, pandas, DigDash, SIRENE API

Education

Engineer's Degree – Data Science & Artificial Intelligence 2022 – 2025
ENSAI – National School for Statistics and Information Analysis (Grande École – affiliated with INSEE, France's national statistical institute)

- **AI Specialization:** supervised & unsupervised learning, deep learning (PyTorch, TensorFlow), advanced NLP (Transformers, LLMs, RAG)
- **Statistics & Applied Math:** GLMs, time series modeling, Bayesian inference, stochastic calculus
- **Big Data & Engineering:** Spark, Hadoop, SQL, Python, Java, APIs (FastAPI), MLOps, cloud deployment
- **Bilingual curriculum (French/English)** – International track (Erasmus+ eligible)

Projects

Hybrid Music Recommender System – Content-based and collaborative approach Jan – Feb 2024
→ *pandas, Surprise, Streamlit, fallback logic, KNNBasic*

Built a hybrid recommendation engine combining KNN-based collaborative filtering and content similarity, with a dynamic fallback for cold-start users.

ReviewGuardian – Toxic comment detection with local explainability Mar 2024 – May 2024
→ *scikit-learn, SHAP, FastAPI, Streamlit*

MultinomialNB model explained with SHAP, deployed as a FastAPI and Streamlit interface.

Bayesian Calibration – Lorenz-96 Model Oct 2024 – Mar 2025
→ *Python, NumPy, matplotlib, ABC-SMC, ABC-MCMC*

Bayesian inference for calibrating parameters of the chaotic Lorenz-96 system; analyzed both performance and computational cost.

InsightDetector – Hallucination detection in generated texts Dec 2024 – Mar 2025
→ *BART, BERTScore, spaCy, Streamlit, RSS, OpenAI, LLM-as-a-judge*

End-to-end summarization and fact-checking pipeline; annotated 300+ articles for hallucinations; open-source-ready with Streamlit interface.

FraudTrack360 – Explainable transaction fraud detection Jan 2025 – Mar 2025
→ *pandas, scikit-learn, LSTM, SHAP, FastAPI, Docker, GitHub Actions, AWS EC2*

LSTM model for sequential anomaly detection; deployed with FastAPI and EC2, CI/CD via GitHub Actions; includes a SHAP-powered Streamlit dashboard.

Skills

Languages: Python, R, SQL, SAS

ML / DL: scikit-learn, XGBoost, TensorFlow, CNN, LSTM, BERT

Generative AI / NLP: Hugging Face, LangChain, OpenAI API, Whisper, BART

Engineering: FastAPI, Docker, Git, CI/CD, REST APIs

Cloud / MLOps: AWS, GCP, pipeline automation

Visualization: Matplotlib, Seaborn, ggplot2, Streamlit

Explainability: SHAP, LIME

Languages

French (native), English (professional working proficiency)