Beedi Goua

🌐 Portfolio | ⚙ GitHub | in LinkedIn | ✉ beedi.goua@eleve.ensai.fr | ▇ +33 779 872 812

Profile

Early-career Data Science engineer passionate about machine learning, NLP, and generative AI. Hands-on experience designing RAG pipelines, predictive modeling, and behavioral analysis. Eager to apply my skills to real-world use cases by building robust, interpretable, and business-oriented solutions.

Professional Experience

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

- Designed an industrial-grade RAG pipeline to query large-scale business PDFs (300+ pages parsed)
- Optimized semantic chunking and evaluated answer quality using RAGAS and TruLens
- Built an interactive prototype with Streamlit, with traceability through logs and automated evaluation
- **Stack**: Python, LangChain, Docling, OpenAI API, ChromaDB, GitHub Actions

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

- Identified polluting entities using supervised clustering (pollution maps + sectoral scoring)
- Matched heterogeneous data sources (SIRENE API, regional datasets) with 95% accuracy
- Delivered actionable policy recommendations to improve sustainable urban mobility
- **Stack**: Python, scikit-learn, pandas, DigDash, SIRENE API

Education

Engineering Degree – Data Science   2022 – 2025
ENSAI – National School of Statistics and Data Analysis (France)

- Multidisciplinary training in statistics, modeling, generative AI, and big data
- **Key courses**:
  - GLMs, time series, survival analysis, stochastic calculus
  - Machine learning, deep learning, NLP, RAG
  - Hadoop, Spark, SQL, OOP (Python, Java)

Projects

ReviewGuardian – Toxic comment detection with local explanation
→ *scikit-learn, SHAP, FastAPI, Streamlit*
MultinomialNB model with SHAP explainability, deployed via FastAPI and Streamlit interface.

Hybrid Music Recommender – Content-based and collaborative system
→ *pandas, Surprise, Streamlit, fallback logic, KNNBasic*
Hybrid recommender combining KNN-based collaborative filtering and content similarity, with dynamic fallback for cold users.

Bayesian Calibration – Lorenz-96 Model
→ *Python, NumPy, matplotlib, ABC-SMC, ABC-MCMC*
Bayesian inference to calibrate chaotic system parameters (Lorenz-96), including performance and computational cost analysis.

InsightDetector – Hallucination detection in generated texts
→ *BART, BERTScore, spaCy, Streamlit, RSS, OpenAI, LLM-as-a-judge*
End-to-end pipeline for summarization, fact-checking and user annotation; detected hallucinations in 300+ articles (news, politics); Streamlit interface, open-source-ready.

FraudTrack360 – Explainable transaction fraud detection
→ *pandas, scikit-learn, LSTM, SHAP, FastAPI, Docker, GitHub Actions, AWS EC2*
Sequential LSTM model with anomaly scoring; containerized FastAPI service, CI/CD with GitHub Actions, deployed on EC2; interactive Streamlit dashboard with local SHAP interpretation.

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 proficiency)