Goua Beedi

🌐 Portfolio · ⦿ GitHub · in LinkedIn · ✉ gouabeedi@gmail.com · 📍 Paris, France

Profile

ENSAI engineering graduate specialized in Data Science and AI. Experienced in designing data/ML pipelines and model evaluation systems, particularly in generative AI. Seeking a Data Scientist position focused on applied AI projects, with strong emphasis on robustness, evaluation, and reliability.

Skills

Programming: Python, SQL, R
Machine Learning & Statistics: statistical modeling, supervised learning, model evaluation, interpretability
Generative AI: RAG systems, LLM evaluation and robustness analysis
Data Engineering: ETL/ELT pipelines, Bronze/Silver/Gold architecture, PostgreSQL, data quality monitoring
Deployment: FastAPI, Docker, CI/CD, Cloud (AWS, GCP)

Experience

Square Management — Square Research Center   Apr 2025 – Nov 2025
Generative AI & Data Science Intern

- Designed a complete RAG architecture for document analysis (OCR, chunking, embeddings, retrieval).
- Benchmarked 189+ configurations (retrieval, generation, reranking) to improve robustness and relevance.
- Developed an automated evaluation framework (faithfulness, relevance, traceability).
- Optimized retrieval using hybrid search (BM25 + embeddings).
- Production packaging (FastAPI, Docker, CI/CD) and technical documentation.

City of Paris — Mobility Agency   Jun 2024 – Sep 2024
Data Science Intern

- Built a supervised model to segment the most polluting vehicle fleets.
- Designed a multi-source SQL + Python pipeline on 30,000+ records.
- Implemented monitoring (latency, robustness, data quality checks).
- Delivered dashboards (KPIs, segmentation, prioritization) to support policy decision-making.

Projects

Credit Scoring — 12-Month Default Risk Model
Dataset: 530,000 accounts, 143 variables

- Data preprocessing (imputation, leakage detection).
- Feature selection for stability and interpretability.
- Logistic regression (AUC = 0.78; Gini = 0.57), evaluated against Random Forest and XGBoost.
- Deployed scoring API using FastAPI and Docker.

ETL Pipeline — Job Market Intelligence

- Multi-source ingestion with Bronze / Silver / Gold architecture.
- Cross-source deduplication, SCD2 historization, PostgreSQL warehouse.
- Delivered reliable KPIs on skill demand trends.

Hallucination Detection — LLM Summary Evaluation

- Automated evaluation pipeline on 300+ articles.
- Methodology to detect inconsistencies between source documents and generated summaries.
- Built annotation interface to categorize hallucination types.

Other projects: beedigoua.github.io

Education

ENSAI — Engineering Degree in Data Science and AI *(Master's Equivalent)*   2022 – 2025
Specialization in **Machine Learning**, NLP and Big Data, with strong foundations in **statistics**, **optimization**, and algorithms.

Preparatory Classes — Mathematics, Physics & Engineering Sciences   2020 – 2022
Intensive program focused on advanced problem-solving and mathematical modeling.

Languages

French (native) — English (professional proficiency)