

Goua Beedi

[Portfolio](#) · [GitHub](#) · [LinkedIn](#) · gouabeedi@gmail.com · [Paris, France](#)

Profil

Diplômé de l'ENSAI, j'ai travaillé sur la conception de pipelines data/ML et de systèmes d'évaluation de modèles, notamment en IA générative. Je souhaite évoluer en tant que Data Scientist sur des projets d'IA appliquée, avec une attention particulière à la robustesse, à l'évaluation et à la fiabilité des résultats.

Compétences

Langages : Python, SQL, R

Machine Learning & Statistiques : modélisation statistique et apprentissage automatique, évaluation et explicabilité

IA Générative : systèmes RAG, évaluation et robustesse des LLM

Data Engineering : pipelines ETL/ELT, architecture Bronze/Silver/Gold, PostgreSQL, contrôle de la qualité des données

Déploiement : FastAPI, Docker, CI/CD, cloud (AWS, GCP)

Expérience

Square Management — Square Research Center

Avr 2025 – Nov 2025

Stagiaire IA Générative & Data Science

- Conception d'une architecture RAG complète pour l'analyse de documents (OCR, chunking, embeddings, retrieval).
- Comparaison de 189+ configurations (retrieval, génération, reranking) afin d'améliorer la robustesse et la pertinence.
- Mise en place d'un cadre d'évaluation automatique (fidélité, pertinence, traçabilité).
- Optimisation du retrieval via recherche hybride (BM25 + embeddings).
- Packaging technique (API FastAPI, Docker, CI/CD) et documentation.

Ville de Paris — Agence de la Mobilité

Juin 2024 – Sep 2024

Stagiaire Data Science

- Développement d'un modèle supervisé pour segmenter les flottes de véhicules les plus polluantes.
- Conception d'un pipeline SQL + Python multi-sources sur 30 000+ enregistrements.
- Mise en place d'un monitoring (latence, robustesse, qualité).
- Restitution via dashboards (KPI, segmentation, priorisation) afin de faciliter l'analyse et la décision.

Projets

Credit Scoring — Modèle de risque à horizon 12 mois

Base : 530 000 comptes, 143 variables

- Sécurisation des données (imputation, détection de fuite).
- Sélection de variables pour la stabilité et l'interprétabilité.
- Régression logistique (AUC = 0,78 ; Gini = 0,57), testée face à Random Forest et XGBoost.
- Déploiement du scoring sous forme d'API FastAPI dockerisée.

Pipeline ETL — Analyse du marché de l'emploi

- Ingestion multi-sources ; architecture Bronze / Silver / Gold.
- Déduplication, historisation (SCD2) et chargement dans PostgreSQL.
- Production de KPI fiables sur la demande par compétence.

Détection d'hallucinations — Évaluation de résumés LLM

- Pipeline automatisé d'analyse de résumés sur 300+ articles.
- Méthodologie d'évaluation des incohérences entre la source et le résumé.
- Développement d'une interface d'annotation pour analyser et catégoriser les incohérences.

Autres projets : beedigoua.github.io

Formation

ENSAI — Diplôme d'ingénieur en Data Science et IA — Bac+5 (Data Science / Ingénieur)

2022 – 2025

Spécialisation en Machine Learning, NLP et Big Data, avec des bases solides en statistiques, optimisation et algorithmes.

Classes Préparatoires — Mathématiques, Physique & Sciences de l'Ingénieur

2020 – 2022

Formation intensive centrée sur la résolution de problèmes et la modélisation.

Langues

Français (natif) — Anglais (professionnel)