

Abdoulaye Balde

Montréal, QC, Canada

abdoulahi.pro96@gmail.com — +1 (514) 709-3612
github.com/abdoulayegk — linkedin.com/in/abdoulayegk

Profil professionnel

Finissant à la maîtrise en informatique à l'UQAM (diplomation prévue en août 2026), spécialisé en apprentissage automatique et en science des données, avec de l'expérience dans la conception de systèmes d'IA robustes et industrialisables (MLOps) pour l'analytique du risque financier. Habile à vulgariser des concepts techniques complexes pour des décideurs d'affaires (VP/DG). Engagé envers l'IA responsable, éthique et transparente. Bilingue français-anglais.

Expérience professionnelle

Scientifique de données, IA et gestion des risques – Stage

Janvier 2026 – Aujourd'hui

Desjardins, Direction analytique en gestion des risques

- Conçu **Vigie Bancaire**, une solution d'IA automatisant l'analyse comparative trimestrielle des divulgations financières des six grandes banques canadiennes, auparavant réalisée manuellement avec Kofax Power PDF.
- Architecturé deux pipelines complémentaires : un système multi-agents pour l'appariement des tableaux et indicateurs financiers et un moteur de comparaison sémantique pour détecter et prioriser les changements narratifs.
- Développé les mécanismes d'extraction, de structuration, d'appariement et de classification des changements, avec validation humaine afin d'assurer la traçabilité et la fiabilité des résultats.
- Ramené le temps d'analyse estimé de 5 à 10 jours à quelques heures par cycle et réduit fortement les faux positifs, permettant aux analystes de concentrer leurs efforts sur les changements à risque élevé.
- Présenté la solution à des gestionnaires et dirigeants lors d'une démonstration complète et contribué à son adoption par l'équipe de divulgation en gestion intégrée des risques.

Formateur en science des données

Janvier 2023 – Juin 2023

Orange Digital Center, Guinée

- Animé des sessions de formation en IA pour le personnel d'Orange Guinée, portant sur l'apprentissage automatique, le TALN (traitement automatique du langage naturel) et les modèles transformeurs.

Organisateur principal – Conférence IA (Deep Learning IndabaX Guinée) Octobre 2022 – Aujourd'hui

Deep Learning Indaba

- Organisé et planifié l'événement de bout en bout : recherche de commanditaires, partenaires, conférenciers et mentors, et animation d'un hackathon en apprentissage automatique durant la conférence.

Projets sélectionnés

Plateforme de données OpenFoodFacts et moteur de recherche avancé — Développeur principal Septembre 2025 - Janvier 2026

github.com/abdoulayegk/Platform.OpenFoodFacts

- Développé une plateforme de données de bout en bout ingérant l'ensemble du jeu de données OpenFoodFacts via un **pipeline ETL Airflow** dans une infrastructure conteneurisée (MinIO, PostgreSQL, MongoDB, JupyterLab), permettant des rafraîchissements de données reproductibles et planifiés.
- Conçu un **moteur de recherche multicritères** (pays, catégorie, ingrédient, marque, nom de produit) combinant l'appariement flou (basé sur la distance de Levenshtein) avec un score de pertinence multifactoriel, améliorant la qualité de récupération par rapport à une recherche par mots-clés naïve.
- Exposé le moteur via une API **FastAPI** personnalisée et une interface **Dash** réactive, permettant aux utilisateurs non techniques d'explorer et de filtrer eux-mêmes des milliers de produits; entièrement conteneurisé avec **Docker Compose**.

Assistant multi-agents en génie logiciel propulsé par l'IA — Chef d'équipe

2025

github.com/abdoulayegk/MGL7320_AI_Assistant

- Architecturé un **système multi-agents** (LangGraph/LangChain) automatisant l'ensemble du cycle de vie du génie logiciel — collecte des exigences, conception de l'architecture, génération de code et configuration du déploiement — coordonné par un agent orchestrateur parmi quatre agents spécialisés (ingénieur des exigences, concepteur d'architecture, ingénieur de code, DevOps).
- Pris en charge la **génération de code multilingue** (Python, Node.js, Java) et **plusieurs modèles LLM** (Gemini, OpenAI, Groq, Ollama), permettant aux utilisateurs de changer de modèle selon le coût ou la performance recherchée.

- Livré une interface **Streamlit** fonctionnelle pour l'ensemble du pipeline, avec déploiement basé sur Docker et intégration continue via GitHub Actions.
- Validé la fiabilité avec une **suite de 33 tests pytest**, couvrant l'orchestration des agents, l'exactitude de la génération de code et la sortie de la configuration de déploiement.

Détection de fraude par carte de crédit (Kaggle) — Participant individuel 2023
Classé dans le top 16% (100^e/641) d'un défi de classification binaire sur données tabulaires, utilisant LightGBM sur Kaggle.

Formation

Maîtrise en informatique (M.Sc.) 2024 – Août 2026
Université du Québec à Montréal (UQAM), Canada
Concentration : science des données, MLOps

Baccalauréat en technologies de l'information (B.Tech.) 2018 – 2022
Université Sharda

Compétitions en science des données (Zindi)

- 2^e place – Data Science Week Guinée 2022
- 18^e/62 – Réclamations d'assurance AutoInland
- Gagnant national – Umojahack 2022, 2023
- 78^e/336 – Prédicteurs de réussite scolaire précoce
- 39^e/493 – Capteur défectueux de qualité de l'air
- Top 10 dans plusieurs défis de détection de fraude

Publications

An Efficient Approach for Predicting Financial Inclusion
2022, *Conférence IEEE ICCSEA* — DOI : 10.1109/ICCSEA54677.2022.9936452

Certifications

- Applied Machine Learning in Python – Université du Michigan (Coursera)
- Introduction to Data Science in Python – Université du Michigan (Coursera)
- Technical Support Fundamentals – Google (Coursera)

Distinctions et leadership

- **Bourse Inclusive AI du CIFAR**, NSERC CREATE en IA responsable – Bourse pour la conférence Canadian AI 2025 Été 2025
- Organisateur principal, Deep Learning IndabaX Guinée, 2023 – Aujourd'hui
- Bourse de voyage – Deep Learning Indaba 2022 (Tunisie)
- Expert Kaggle et contributeur en données ouvertes

Compétences techniques

IA générative et NLP : LangGraph, LangChain, LLM (OpenAI, Gemini, Groq, Ollama), Transformers, Embeddings, Systèmes multi-agents, RAG

Science des données : Python, SQL, Scikit-learn, LightGBM, PyTorch, TensorFlow

Données et MLOps : Airflow, MLflow, Docker, GitHub Actions, FastAPI, PostgreSQL, MongoDB, Spark

Visualisation et applications : Dash, Streamlit, Tableau

Cloud et environnements : GCP, AWS, Linux, Bash

Langues

Français : Langue maternelle **Anglais** : Courant