

ISIDORE ZONGO

Data Scientist · Ingénieur en intelligence artificielle

 zongisidore10@gmail.com

 +33 7 54 18 23 89

 Brest, mobile France entière

 Disponible immédiatement

 LinkedIn

 GitHub

 Portfolio

Ingénieur diplômé en intelligence artificielle et big data, avec trois ans d'expérience de data scientist en alternance chez EDF. Je conçois et mets en production des modèles de machine learning et de deep learning (prévision de séries temporelles, classification) et des applications d'intelligence artificielle générative (LLM, pipeline RAG), du besoin métier jusqu'à la mesure de leur performance.

EXPÉRIENCE

Data Scientist (alternance)

EDF SEI, Systèmes Énergétiques Insulaires

 oct. 2023 - sept. 2026

 Brest

Prévoir la consommation électrique d'une île pour décider quels groupes électrogènes démarrer

- Mis en production une agrégation d'experts qui bat tous les modèles individuels de 30 min à 24 h et fait jeu égal à 48 h : **erreur moyenne en pourcentage (MAPE) de 3,32 % à 30 min et 5,11 % à 24 h**
- Réduit la part des pics de consommation sous-estimés **de 72 % à 34 %** (GAM quantile, loi Gamma)
- Comparé six modèles (Ridge, GAM sous R, TAM d'EDF R&D, XGBoost, LightGBM, GRU retenu face au LSTM), optimisés sous Optuna et suivis sous MLflow, sur un protocole sans fuite d'information (découpage chronologique, test mesuré une seule fois)
- Construit 45 variables à partir de six ans de mesures au pas de 5 min (658 000 points extraits par l'API interne d'EDF), croisées avec la météo et le calendrier
- Développé l'application qui donne ces prévisions aux exploitants pour planifier le démarrage des groupes (moins de fioul, de CO₂ et d'usure)
- Présenté résultats et limites aux exploitants et à la R&D d'EDF ; documenté la démarche dans un rapport et un article

Prévoir la production renouvelable, superviser et piloter

- Développé et comparé des modèles de prévision de la production photovoltaïque selon la météo, la nébulosité, l'inclinaison, l'orientation et la taille des panneaux
- Prévu la production d'hydroliennes à partir des marées
- Détecté des anomalies sur les groupes électrogènes diesel, par des méthodes supervisées et non supervisées
- Construit la supervision temps réel de cinq centrales (photovoltaïque, diesel, batteries) avec InfluxDB, Grafana et Home Assistant
- Conçu le pilotage intelligent des équipements d'une maison connectée en fonction de la production renouvelable, pour optimiser la consommation

FORMATION

Diplôme d'ingénieur

Intelligence artificielle & big data

ISEN Yncréa Ouest, Brest

 2021 - 2026

Classe préparatoire scientifique

École Polytechnique d'Agadir, Maroc

 2020 - 2021

COMPÉTENCES

Langages : Python SQL R

Machine learning : scikit-learn XGBoost LightGBM Optuna

Classification Clustering Détection d'anomalies

Séries temporelles : Prévision GAM

Deep learning : PyTorch TensorFlow LSTM / GRU

Transformers

Vision par ordinateur : CNN Transfer learning YOLO OpenCV

IA générative : LLM RAG Agents LangChain LangGraph

Ollama

MLOps et cloud : MLflow Docker FastAPI Git CI/CD AWS

Données : Pandas PostgreSQL PySpark Dataiku Databricks

Supervision et visualisation : InfluxDB Grafana Power BI

CERTIFICATIONS

Machine Learning Specialization

DeepLearning.AI & Stanford

2025

Deep Learning Essentials with Keras

IBM

2026

LANGUES

Français (langue maternelle)

Anglais (B2, TOEIC 845)

PROJETS

IA GÉNÉRATIVE ET MODÈLES DE LANGAGE

Assistant RAG sur documents confidentiels

📅 2026

Python, LangChain, sentence-transformers, ChromaDB, Ollama, FastAPI, Docker

- Conçu un assistant qui répond aux questions sur des documents confidentiels, entièrement en local : aucune donnée ne quitte la machine
- Construit un pipeline RAG modulaire : ingestion de PDF, TXT et Markdown, découpage, embeddings, recherche vectorielle dans ChromaDB, génération par un LLM local (Llama 3.2 via Ollama)
- Exposé le service en API REST : réponses qui citent le document et la page, réindexation sans doublon, tests automatisés (pytest), conteneurisation sous Docker

Mini GPT codé de zéro

📅 2026, en cours

PyTorch · GitHub [🔗](#)

- Codé un Transformer de type GPT sans bibliothèque de Transformer, pour comprendre les modèles de langage de l'intérieur : attention causale multi-têtes, blocs feed-forward, connexions résiduelles, LayerNorm, génération autorégressive
- Mené des ablations : à paramètres égaux, 4 têtes font mieux qu'une (2,40 contre 2,51) ; sans résiduel ni LayerNorm, tout bloque (3,19)
- Entraîné le modèle sur 18 millions de caractères (34 romans de Jules Verne) : **perte de test ramenée de 2,51 à 2,19**, sous le plafond du meilleur bigramme (2,47) ; en cours : embeddings de position, abstention par fine-tuning supervisé, appel d'outil

Assistant conversationnel pour un hôtel (Layana Resort, Thaïlande)

📅 2025

Python, API OpenAI, Supabase (PostgreSQL)

- Développé la partie intelligence artificielle de l'assistant intégré à l'application de l'hôtel, pour le personnel et pour les clients
- Connecté un LLM (API OpenAI) à une base Supabase créée pour le projet, afin que chaque réponse s'appuie sur les données réelles de l'hôtel
- Donné au personnel les arrivées, les départs et les allergies des clients du jour, et aux clients les activités et leurs horaires

Galen, aide au diagnostic médical par imagerie

📅 2026

Python, PyTorch, DenseNet121, Claude (LLM multimodal)

- Conçu à deux un prototype déployé où l'utilisateur dépose une image médicale (radiographie, IRM, scanner) et obtient une analyse, avec le business plan associé (patients, hôpitaux, télémedecine)
- Intégré Claude, un LLM multimodal qui décrit l'image, signale les anomalies, rédige un compte rendu et répond aux questions
- Fine-tuné un DenseNet121 qui détecte la pneumonie sur radiographie thoracique avec une probabilité : **ROC-AUC de 0,98 sur le jeu de test**

SÉRIES TEMPORELLES

Transformers pour séries temporelles : étude et reproduction de TimeXer

📅 2026

Python, PyTorch, XGBoost, LSTM, GRU

- Étudié l'état de l'art des Transformers pour séries temporelles (PatchTST, iTransformer, TimeXer) et l'architecture TimeXer (NeurIPS 2024), qui relie la série cible aux variables externes par attention croisée
- Implémenté TimeXer de zéro en PyTorch et reproduit le dépôt officiel sur ETTh1 (17 420 observations horaires, 4 horizons), puis appliqué le modèle à la prédiction de présence dans un bâtiment intelligent (météo, calendrier, réservations)
- Comparé six modèles : **XGBoost en tête (R^2 de 0,888)** ; le LSTM, meilleur réseau de neurones (R^2 de 0,528), devance GRU, RNN et TimeXer

VISION PAR ORDINATEUR

Reconnaissance d'actions humaines dans des vidéos

📅 2026

Python, PyTorch, R3D-18, OpenCV

- Construit un système qui reconnaît 101 actions humaines dans des vidéos, sur un jeu de 13 451 vidéos
- Appliqué le transfer learning à un réseau convolutif 3D (R3D-18, pré-entraîné sur Kinetics-400) : 51 000 paramètres entraînés sur 33 millions
- Atteint **une précision de 93,4 %**

Classification d'images, détection d'objets et analyse de sentiments

📅 2025

Python, TensorFlow, MobileNetV2, EfficientNet-B0, YOLO

- Classé des images CIFAR-10 et Fashion-MNIST avec des réseaux convolutifs (CNN), des perceptrons multicouches et du transfer learning (MobileNetV2, EfficientNet-B0)
- Détecté des objets dans des images avec YOLO ; classé le sentiment de tweets en trois classes (positif, neutre, négatif)
- Reconnu les chiffres manuscrits de MNIST avec **plus de 98 % de réussite**

STATISTIQUE ET OPTIMISATION

Analyse de données fonctionnelles

📅 2025

Python, scikit-fda, scikit-learn, TensorFlow

- Classé et prédit des données fonctionnelles, des spectres entiers plutôt que des valeurs isolées (jeux Octane et Moisture)
- Mis en œuvre le k-NN fonctionnel et la régression logistique fonctionnelle
- Comparé ces méthodes au deep learning (MLP, CNN, LSTM, GRU) et au k-NN classique

Recherche opérationnelle et théorie des jeux

📅 2024

Python, NumPy, Matplotlib

- Codé de zéro un algorithme génétique pour le voyageur de commerce (sélection, croisement, mutation), en programmation orientée objet
- Comparé l'algorithme au plus proche voisin, avec visualisation en temps réel des routes ; mis en œuvre Dijkstra et une recherche heuristique
- Développé un joueur de Puissance 4 qui cherche le meilleur coup (théorie des jeux)