

# ZAKARIA KASSEMI

## Ingénieur d'Etat en Data Science

zakariakassemi65@gmail.com | +212 705 701 881 | [linkedin.com/in/zakaria-kassemi-2917b3289](https://linkedin.com/in/zakaria-kassemi-2917b3289) | [github.com/zakariakassemi2000](https://github.com/zakariakassemi2000) | [huggingface.co/kassemi](https://huggingface.co/kassemi)

### PROFIL PROFESSIONNEL

Ingénieur en Data Science spécialisé en IA & LLMs, RAG et Data Engineering. Je conçois des systèmes intelligents (pipelines data, assistants IA, solutions NLP/CV) avec un focus sur la performance, la scalabilité et l'impact métier

### EXPÉRIENCE PROFESSIONNELLE

#### Junior Data Scientist | WINCO Canada — Équipe RH

Déc 2025 – aujourd'hui

- Conçu et déployé 3 pipelines Python de traitement de données RH, alimentant les décisions de recrutement sur +4 000 profils candidats.
- Structuré des datasets multi-sources (APIs, bases professionnelles) pour la production de 8 rapports d'intelligence stratégique mensuelle.
- Automatisé les workflows de collecte et d'analyse via scripts reproductibles, réduisant 40 % le temps de traitement manuel.
- Développé un ATS (Applicant Tracking System) de A à Z : gestion des candidatures, parsing de CV et suivi des processus de recrutement pour +500 dossiers.
- Livré 100 % des jalons dans les délais sur 4 mois en autonomie complète, en environnement 100 % remote.

#### Développeur Web | Hôpital Molay Ahmed Ben Driss El Missouri

Avr 2024

- Conçu et livré en 1 mois une application complète de gestion des dons de sang : suivi donneurs/receveurs, stocks par groupe sanguin en temps réel.
- Intégré des tableaux de bord Chart.js permettant au personnel de piloter les tendances de dons — adoption immédiate par l'équipe médicale dès la mise en production.

#### Développeur Web Full Stack | ARIA SARL — Meknès

Août – Déc 2023

- Développé un CRM interne complet (clients, contrats, rendez-vous, export PDF, KPIs) remplaçant 100 % d'un processus papier pour une équipe de 10 utilisateurs.
- Réduit la charge administrative de 35 % grâce à l'automatisation des flux documentaires.

### PROJETS CLÉS

#### SHIFA AI — Plateforme Médicale Arabe (PFE) — *PyTorch · MONAI · FAISS · BM25 · LLaMA-3 · AraBERT · Whisper · Groq · Streamlit*

- Architecturé une plateforme médicale multimodale arabophone : 5 modules CNN (EfficientNet-B3 dermatologie, DenseNet-121 chest X-ray, ResNet-50+MONAI IRM cérébrale, InceptionV3 densité mammaire) avec Grad-CAM pour l'explainabilité, atteignant >90 % de précision moyenne sur l'ensemble des modules.
- Pipeline RAG hybride FAISS (dense) + BM25 (sparse) + RRF fusion + cross-encoder reranking couvrant 16 domaines médicaux, avec routing NLP automatique et latence de réponse <2 s.
- Déployé sur Streamlit Cloud avec interface RTL dark-mode full Arabic .

#### Agentic Swarm — Système Multi-Agents IA Hors-Ligne — *Transformers · BitsAndBytes · Accelerate · Gemma-4 · PySide6*

- Architecturé un système de 5 agents IA spécialisés en boucle infinie avec critique croisée, 100 % hors ligne
- Moteur d'inférence à gestion dynamique de VRAM
- Interface desktop native PySide6 glassmorphism avec intervention humaine temps réel et personnalisation par IA

#### Legal Assistant — Chatbot Juridique RAG — *RAG · LangChain · FAISS · LLM · Python*

- Développé un assistant juridique IA basé sur RAG fournissant des conseils conformes au droit marocain, couvrant 5+ corpus législatifs.
- Pipeline end-to-end : ingestion documentaire → vectorisation FAISS → récupération sémantique → génération de réponse contextualisée.

#### Brain Tumor Detection — Computer Vision Médicale — *CNN · TensorFlow · Python · OpenCV*

- Construit un système deep learning de détection automatique de tumeurs cérébrales sur images IRM, atteignant >95 % de précision en classification multi-classes (4 catégories).
- Pipeline complet : preprocessing, augmentation de données, entraînement, inférence et visualisation des résultats via heatmaps Grad-CAM.

## Hadith Text Completion — NLP Arabe — *Transformers · HuggingFace · Arabic NLP · Fine-tuning*

- Fine-tuné un modèle Transformer arabophone (AraBERT) pour la complétion contextuelle de hadiths à partir de texte partiel, avec amélioration de 15 % du BLEU score vs baseline.

## COMPÉTENCES TECHNIQUES

---

**Data Engineering** : Apache Spark, Hadoop (notions) · Kafka (streaming) · Airflow (orchestration) · ETL/ELT pipelines · Data Warehousing (Snowflake, BigQuery) · Data Modeling (Star Schema, Snowflake Schema)

**LLMs & Prompt Engineering** : GPT-4, Claude, Gemini Pro, LLaMA-3, Mistral · Fine-tuning LoRA/PEFT · Chain-of-Thought

**RAG & NLP** : FAISS, LangChain, BM25, RRF, Cross-encoder · AraBERT, SpaCy, NLTK, Whisper · HuggingFace

**Computer Vision** : PyTorch, TensorFlow, Keras · MONAI · OpenCV, YOLO, Detectron2 · Grad-CAM

**Machine Learning** : Scikit-learn, Pandas, NumPy · Power BI · Airflow, ETL (Talend, NiFi)

**Développement** : Python, FastAPI, Django · React, HTML5/CSS3 · Node.js, Laravel · REST API, .NET Core MVC

**Cloud & DevOps** : AWS, Docker, Kubernetes · GitHub Actions, CI/CD · n8n, Zapier · Streamlit Cloud

**Bases de données** : PostgreSQL, MySQL, Oracle, SQL Server · FAISS (vectoriel)

## FORMATION

---

**Ingénieur d'état en Data Science (Bac+5)** | SUPMTI — Rabat 2024 – 2025

**Bac+3 Génie Informatique — Mention Très Bien** | SUPMTI — Meknès 2021 – 2024

**Certifications Techniques I & II et Brevet Élémentaire en Génie Logiciel** | CIT — Kénitra 2019 – 2021

## LANGUES

---

**Arabe** : Langue maternelle

**Français** : C1 — courant professionnel

**Anglais** : C1 — courant professionnel

**Allemand** : A1 — débutant

## CERTIFICATIONS

---

Certifications disponibles sur Google Drive : [Lien](#)