Abubakar ALIYU
Data Scientist & ML Engineer
✉**: Email** | ☎:(+33) 605981637 | **Location:** 4 Rue Pasteur, Kremlin Bicêtre, Paris, France
Technical Blog | **LinkedIn** | **GitHub**

_____PROFESSIONAL SUMMARY_____

Data Scientist & ML Engineer pursuing an MSc in Data Science at EPITA with hands-on experience in deep learning, computer vision, and large language models. Proven track record in designing, training, and deploying production-ready AI systems backed by rigorous experimentation and evaluation. Passionate about transforming research innovations into reliable real-world applications through ethical, high-impact AI solutions. Self-starter and collaborative teammate seeking a 6-month internship or entry-level role in AI/ML engineering or research.

_____EDUCATION_____

EPITA School of Engineering and Computer Science — Paris, France
Master of Science, Data Science and Analytics (BAC +5)   **Sept. 2024 – Apr. 2026 (Expected)**
Thesis: Retrieval-Augmented Reasoning with Self-Correction for Biomedical QA (RARE + self-correction)

Ahmadu Bello University — Kaduna, Nigeria
Bachelor of Engineering, Computer Engineering (BAC +4 equivalent)   **Oct. 2015 – Feb. 2022**
Second Class Upper (14/20, Mention Bien)

_____WORK EXPERIENCE_____

AI Fellow — Centre for Journalism Innovation and Development (Contract, Nigeria)   **Jun. 2024 – Nov. 2024**
- Prototyped ChatJourno (chatjourno.com), a RAG-based fact-checking chatbot empowering West African journalists.
- Built scraping pipelines with BeautifulSoup and curated datasets from trusted news sources for retrieval-augmented generation.

Data Science Project Manager — Ibsonka Energy Resources (Freelance, Nigeria)   **Jan. 2024 – Jul. 2024**
- **Standardized healthcare data (**clinical records, staff accounts, pricing, drugs, and service catalogs) for migration to EMR across two hospitals.
- **Led implementation and trained over 100 staff**, cutting support issues by 70% via stakeholder and change management.

_____PERSONAL AND ACADEMIC PROJECTS_____

Retrieval-Augmented Reasoning with Self-Correction for Biomedical QA (Master's Thesis) | Thesis Poster

Project Lead | Literature Review, Finetuning, Quantization, LLMs Deployment, Redis, FAISS, PyTorch, HIPAA Compliance
- Conducted a literature review on retrieval-augmented reasoning, self-correction, and hallucination detection to identify gaps in biomedical QA.
- Built a multi-stage RARE pipeline (Llama-3.1-8B + SciBERT + Flan-T5) achieving 87% error detection accuracy and 0.82 semantic preservation F1, deployed via Hugging Face Inference Endpoints with HIPAA-compliant UI.

Diffusion Model Tuning (Stable Diffusion)| Git Repository

Diffusers, Transformers, Accelerate, PyTorch, Evaluation
- Conducted controlled experiments on schedulers, CFG scales, inference steps, and prompt engineering.
- Analyzed trade-offs between image quality, prompt adherence, and generation speed for photorealistic outputs.

Serverless AWS Data Pipeline (NYC Taxi Data) | Git Repository

AWS (Lambda, EMR, Glue, Athena, CloudFormation), PySpark
- Processed and transformed over 1.4 million records using a fully event-driven ETL architecture with Spark jobs for efficient big data analytics.
- Built a secure, auditable pipeline featuring fine-grained IAM permissions, encryption, and automated job orchestration with CloudFormation.

YOLOv8 Object Detection and Instance Segmentation | Git Repository

PyTorch, Ultralytics YOLOv8, Evaluation
- Built end-to-end pipelines for trash object detection and plant disease instance segmentation using YOLOv8.
- Achieved 85.9% mAP on trash detection and 36.6% mask mAP on plant disease segmentation through systematic tuning.

_____TECHNICAL & SOFT SKILLS_____

- **PROGRAMMING LANGUAGE:** Python(PyTorch, TensorFlow, Keras, Scikit-Learn, Hugging Face Transformers, OpenCV, Matplotlib), R, SQL
- **AI & DATA:** LLM finetuning (LoRA/QLoRA), CNNs, GANs, object detection, segmentation, Stable Diffusion, data generation and data pipelines
- **EXPERIMENTATION AND EVALUATION:** LLM-as-a-Judge, model benchmarking, prompt engineering, metric analysis (ROUGE, BLEU, BERTScore)
- **MLOPS & DATA ENGINEERING:** Docker, Airflow, Git, Hugging Face, AWS, Streamlit, Redis, CI/CD fundamentals (GitHub Actions), Streamlit dashboards, Tableau, Dataiku, Grafana
- **DATABASES:** PostgreSQL, MySQL, Redis, MongoDB
- **SOFT SKILLS:** Analytical thinking, Proactive, detail-oriented, adaptable, autonomous, and confident working with cross-functional teams and clients
- **LANGUAGES:** English (Bilingual), French (A2)

_____CERTIFICATIONS_____

- **Google Certified Data Analyst (2023) |** Certificate
- **Google Certified Project Manager(2023) |** Certificate