ABRAHAM AJIBADE

Chicago, Illinois | 859-693-3366 | abraham0ajibade@gmail.com | LinkedIn | GitHub | Website

PROFESSIONAL SUMMARY

Results-driven Artificial Intelligence and Machine Learning Engineer with 5+ years experience in driving the conception, design and deployment of artificial intelligence and machine learning solutions across healthcare, finance, retail and real estate. Expertise spans Generative AI, Agentic AI, Deep Learning, and Classical Machine Learning applications, utilizing core frameworks like PyTorch, LangChain/LangGraph and Scikit-learn/XGBoost. Proven success in bridging the gap between development and production by implementing robust MLOps and LLMOps pipelines, integrating appropriate deployment tools across cloud and/or on-premises environments. Strong team collaborator who communicates complex solutions clearly, aligns stakeholders/non-technical audiences, and drives initiatives with an agile methodology based approach.

TECHNICAL SKILLS

Programming Languages

- Python, Groovy, R, SQL, Linux

Machine Learning and Deep Learning

- ***Frameworks & Libraries***: PyTorch, Scikit-Learn, PySpark MLlib, XGBoost, Huggingface Transformers, Open Neural Network Exchange (ONNX),
- ***Domains***: Statistical Analysis, Supervised and Unsupervised Learning, Deep Learning, Natural Language Processing, Transformer Models, Computer Vision, Hyperparameter Tuning

Generative AI

- ***Frameworks & Libraries***: LangChain, LangGraph, SemanticKernel, vLLM, LiteLLM, Pydantic AI, CLIP, BLIP, FastMCP
- ***Domains***: Large Language Models, Prompt Engineering (Zero-shot, Few-shot), Retrieval Augmented Generation (RAG), Text-to-Speech modeling, Speech-to-Text modeling, Model Context Protocol (MCP), Multi-Agent Systems, Vector Databases, LLM Inference Optimization

Deployment and Monitoring

- ***Production Deployment***: FastAPI, Docker, Triton Inference Server AWS SageMaker, Azure ML, Databricks
- ***Monitoring***: Arize AI, AWS Sagemaker Model Monitor

Data Engineering

- ***Big Data***: ETL/ELT, Spark, Databricks, AWS EMR/Glue, Airflow, Epic Systems

Cloud Computing & Infrastructure

- ***Cloud Platforms***: AWS, Azure, GCP, Oracle Cloud
- ***Infrastructure as Code***: Terraform, Ansible

Standard Practices

- ***Code Quality***: pytest, unittest, pre-commit (ruff, sqlfluff), devcontainers
- ***Version Control****:* Git, GitHub, BitBucket, Azure DevOps
- ***Continuous Integration and Delivery***: GitHub Actions, Azure Pipelines, Jenkins, Drone CI

WORK EXPERIENCE

Machine Learning Engineer **December 2024-Present**

Northwestern Medicine | Chicago, Illinois

- Architected and led the end-to-end deployment of a high-throughput incident classification system (6,500+ daily reports), containerizing a fine-tuned Mixtral 7x8B model via Triton Inference Server on Azure VMSS, eliminating over 90% of the manual review workload.
- Developed and deployed a prototype HIPAA-conscious NLP pipeline to distill complex clinical summaries and physician notes into patient-friendly narratives at a 6th-grade reading level. Leveraged OpenAI's GPT-5 with sophisticated prompt engineering and few-shot learning to ensure medical accuracy while stripping out dense terminology; implemented a FastAPI wrapper to provide instantaneous, empathetic health insights, significantly improving health literacy and patient follow-through.
- Designed and led the deployment of a distributed, event-driven multi-agent system for real-time emergency room triage, utilizing an asynchronous FastAPI and Azure Service Bus orchestration layer to achieve sub-second patient severity classification with 99.9% uptime during high-concurrency surges.
- Established a department-wide model deployment framework using ONNX + Triton, resolving cross-backend compatibility for 4+ model families, which reduced deployment time by 25% and unified inference performance.
- Spearheaded the DevSecOps transformation within Azure DevOps, integrating security to automate code and secret auditing, resulting in the remediation of 150+ critical vulnerabilities pre-production.
- Engineered the scalable data orchestration layer on Databricks (PySpark/SQL), automating secure ingestion and HIPAA-compliant storage of patient data, thus enabling real-time model feedback loops and auditability.
- Productionized a medical document summarization microservice, leveraging LLMs and Pydantic AI for automated validation and PII redaction to convert complex reports into 6th-grade readability for clinical staff.
- Collaborated with a team of data scientists to optimize feature engineering pipelines, implementing parallel PySpark workflows that slashed feature generation latency by 92% and directly enabled faster model training and real-time inference.

- Authored and implemented the "Shift-Left" quality strategy across the team, configuring Devcontainers and pre-commit hooks to mandate local execution of tools (ruff, pytest, sqlfluff) prior to commit, which reduced failed remote CI/CD runs by over 40% and minimized unnecessary build runs and charges.
- Authored and implemented architecture design standards for scalable ML systems, driving the adoption of Terraform and standardized MLOps patterns across Azure AKS and Container Apps to improve deployment velocity and governance.

Python Developer | AI Engineer **October 2025-February 2026**

MyCartsOnline | Lagos, Nigeria (Contract)

- Designed and deployed a multimodal search platform combining BLIP-based auto-captioning with OpenCLIP (768D) embeddings to enable high-precision text-to-image and image-to-image retrieval from a product catalog.
- Implemented a Retrieval-Augmented Generation (RAG) layer backed by Qdrant, improving semantic matching and boosting search precision by 40% over the baseline ChromaDB implementation through optimized vector indexing and hybrid retrieval strategies.
- Engineered a high-throughput Airflow data ingestion pipeline that automated real-time catalog synchronization for 10K+ products at 60 items/sec.
- Designed a modular decision engine combining deterministic rules, probabilistic reasoning, and AI-driven inference to detect shortages, overstock patterns, and operational anomalies with high interpretability.
- Built a Model Context Protocol (MCP) server exposing inventory, product, and vendor search capabilities, enabling multi-agent systems to perform real-time lookups, supplier matching, and product discovery through a unified tool layer.
- Engineered MCP-integrated workflows for automated stock queries, reorder checks, anomaly detection, and supplier validation, reducing manual intervention and improving data reliability.
- Implemented secure, scalable MCP endpoints for database access, email notifications, and system-to-system integrations, enabling agents to trigger alerts, generate reports, and initiate corrective actions autonomously.
- Implemented automated model versioning and experiment tracking via the mlflow model registry, while also exporting models for REST API deployment and A/B testing workflows through the mlflow PyFunc wrappers.
- Created model serving infrastructure using FastAPI endpoints and PyFunc wrapped models,, supporting real-time inference, batch processing, and LLM agent integration via Model Context Protocol tools.

Data Scientist **May 2023-November 2024**

Blue Lambda Technologies | Atlanta, Georgia

- Architected and scaled end-to-end ML systems across three distinct domains (real estate, retail, finance), delivering PyTorch, XGBoost, and Scikit-Learn models to production with 95%+ uptime via AWS SageMaker for scalable batch inference and automated retraining.
- Collaborated on the development and deployment of an internal RAG platform using LangChain and Qdrant Vector Database with Gemma 3.4B, resolving up to 7,000 weekly employee queries and reducing support ticket volume by 42% with 88% first-pass accuracy.
- Owned the data engineering backbone for 15+ concurrent models; built and maintained robust ETL pipelines using PySpark on Databricks/AWS EMR, strictly enforcing data lineage and quality gates to ensure production-grade, validated datasets.
- Instituted proactive MLOps monitoring using Arize and MLflow to detect data and concept drift; integrated auto-alerts and rollback protocols, successfully sustaining <5% metrics degradation across models over 18-month lifecycles.
- Championed and enforced engineering standards by establishing reproducible ML environments using Docker and Docker-Compose configurations via Devcontainers.
- Instituted static analysis, secret scanning, and automated linting in CI/CD workflows, reducing technical debt by 60% in training and deployment pipelines.
- Designed and implemented scalable feature engineering workflows for structured and unstructured data, which improved model generalization across the three distinct business domains.
- Authored living documentation standards for data flows, model cards, and inference APIs, adopted as team-wide templates to accelerate cross-team collaboration and streamline compliance audits.

Graduate Research Fellow **January 2021-May 2023**

University of Kentucky | Lexington, Kentucky

Mark B. Tyler Research Industrial Hemp Lab

- Implemented optimization softwares (AIMS, R and Python) and prescriptive analytics on resource allocation problems for bourbon manufacturing firms, culminating in published research and improving operational efficiency by 22%.
- Conducted rigorous statistical modeling for academic research papers, utilizing R and Python to validate complex datasets and ensure 95% confidence intervals of predictions.
- Authored research findings based on predictive modeling techniques (market trends for industrial hemp firms).

Data Scientist **January 2019-November 2020**
Fiverr & UpWork | Victoria Island, Lagos, Nigeria

- Delivered custom ML solutions (Scikit-learn, XGBoost) for classification and regression across diverse client portfolios, achieving 15% average R^2 uplift and 22% precision gain by directly aligning model metrics to business-critical KPIs.
- Architected a real-time incremental data ingestion system using PySpark on AWS EMR, processing over 1.5M rows per run with sub-minute latency, reducing data-to-model latency by 87% and enabling live predictive scoring.
- Engineered production-grade feature pipelines, incorporating automated hyperparameter optimization and SHAP-based explainability to enable non-technical stakeholders to interpret key model drivers via interactive dashboards.
- Refactored legacy code (Jupyter notebooks) into modular, PEP8-compliant packages; instituted comprehensive pytest/unittest suites (unit/integration/E2E), achieving 99% test coverage and guaranteeing audit-ready client handoffs.
- Automated and scaled preprocessing workflows across multi-gigabyte datasets, implementing schema validation, drift detection, and fault-tolerant retries to ensure always-ready data delivery with zero manual intervention.
- Drove model governance and transparency by integrating SHAP values and model cards directly into production APIs, empowering clients to trust and act on predictions with verifiable confidence.

Research and Content Intern **January 2018-December 2018**
Gidimo | Lagos, Nigeria

- Optimized mobile learning content for a WAEC prep application serving 10,000+ students, ensuring 100% alignment with national education policy and quality standards.
- Leveraged user feedback analysis and research to drive a 15% improvement in content engagement, providing data-backed insights for iterative platform enhancements.
- Streamlined content delivery workflows and coordinated large-scale uploads to support organizational strategies for expanding educational access.
- Standardized quality assurance protocols for mobile content uploads, reducing content errors and improving the overall user interface experience for students.
- Coordinated cross-functional workflows between the research and technical teams to streamline the deployment of high-priority exam prep features.

RECENT PROJECTS

Retail Chat Agent Backend – Multimodal Product Discovery

- Developed and deployed an AI-powered chat assistant for multimodal retail product discovery, enabling users to search via natural-language descriptions or photo uploads. Integrated OpenCLIP (for image embeddings), BLIP (for image captioning), text embeddings, and vector database searches; served via FastAPI and AWS infrastructure to deliver fast, relevant matches and streamline the shopper experience.

Retail Chat Agent MCP Server – Multifunction Tools Support

- Designed and deployed a Model Context Protocol (MCP) server to provide real-time product, vendor, and inventory search capabilities for multimodal retail agents. Implemented modular tool endpoints for vector search, database queries, and workflow automation, enabling chat-based and autonomous agents to perform high-precision product discovery, validate inventory, and trigger downstream operational actions. Integrated seamlessly with the multimodal search backend to unify text, image, and metadata retrieval across retail catalogs.

Human Resources Policy RAG Chatbot

- Architected and deployed a specialized RAG (Retrieval-Augmented Generation) chatbot, using FastAPI and Qdrant to provide context-grounded answers on human resources policy-related questions within the organization. The entire system was engineered for data privacy by ensuring local language model integration and storing proprietary documentation exclusively within a self-hosted Vector Database, eliminating reliance on external APIs.

Retail Chat Agent ETL – Automated Multimodal Data Pipeline

- Engineered and deployed a production-grade ETL pipeline to automate the ingestion and vectorization of large-scale e-commerce catalogs. Built with Apache Airflow and Python, the system orchestrates web scraping and API calls, AWS S3 image hosting, and metadata storage in PostgreSQL. Integrated OpenAI CLIP and BLIP to generate synchronized text and image embeddings, populating a Qdrant vector database to power high-performance semantic search for downstream AI chat agents.

EDUCATION

Master of Science in Agricultural Economics **2021-2023**
University of Kentucky | Lexington, Kentucky

Bachelor of Science in Agricultural Economics **2012-2017**
University of Benin | Benin-City, Nigeria