

# David Adarkwah

Data Engineer

[LinkedIn](#) | [GitHub](#) | [Email](#) | [+447467082048](#) | London, UK

## PROFILE SUMMARY

---

Data Engineer with 5+ years of experience building production LLM-powered tools, RAG workflows, specialized AI agents, and operating large-scale data pipelines and ETL infrastructure in production environments. At Bloomberg, owns the end-to-end climate disclosure ingestion platform for the Net-Zero Data Public Utility, building fault-tolerant Airflow-orchestrated pipelines processing 14M+ data points annually across multiple formats and storage systems. Specializes in distributed data processing with PySpark, canonical dataset design, data quality frameworks and building automation tools easily adopted by non-technical team members.

## PROFESSIONAL EXPERIENCE

---

### **Bloomberg LP | London, UK**

06/2023 - Present

*Senior Data Engineer, Net-Zero Data Public Utility*

- Designs, builds and owns a multi-provider climate disclosure ingestion pipeline processing 50,000+ filings and 14M+ data points annually from 8 regulatory and third-party sources across XBRL, JSON, CSV and XML formats into GCS, MongoDB and PostgreSQL, underpinning analytics, product metrics and client reporting across 12,000+ companies.
- Implemented fault-tolerant ingestion with exponential back-off retry and idempotent publishing, achieving zero data loss across both API-based and file-based acquisition paths over 12 months of production operation.
- Cut end-to-end pipeline execution from 8 minutes to under 2 minutes through parallel PySpark processing, Python multiprocessing for validation workloads and batched API publishing, reducing data-to-insight latency by 75%.
- Orchestrates all pipeline workflows with Airflow on GCP, managing task dependencies, retry policies and alerting via Slack and PagerDuty, achieving 99.5% pipeline reliability over 12 months.
- Designed a provider plugin architecture cutting new provider on-boarding from 4 weeks to 5 days, and built a CLI coverage tool automating disclosure completeness analysis across 8 providers and 6 reporting years, eliminating 3 weeks of manual cross-referencing per data cycle.
- Ensured data integrity and compliance through Pydantic schema enforcement and validation gates at every pipeline stage, provider-scoped authorization, versioned audit trails and automated metadata conflict detection across 12,000+ companies.
- Built and deployed 4 internal AI agents using Gemini and LangChain, including a Jira ticket automation agent saving 4 hrs/week, a JSON test data generator cutting QA prep by 2 hrs/cycle, and a climate data coverage analysis agent collectively reducing analyst operational overhead by 10+ hours per sprint.
- Designed and documented 4 LLM-powered workflows using Gemini and LangChain, covering data quality investigation, anomaly analysis, data summarization, and documentation generation, with reusable prompt templates that enabled non-technical team members to self-serve without engineering support.

### **Consolidated Bank Ghana | Accra, Ghana**

09/2021 - 04/2023

*Senior Data Scientist*

- Built and maintained 10+ SQL Server data marts powering core banking operations including the banking app and ledger updates, and engineered a Prefect-orchestrated backup COB extraction pipeline that eliminated vendor-dependent outages and saved the bank multiple days of cross-timezone triage per incident.
- Built an NLP-based incident triage and routing system processing 430+ incidents monthly with 96.5% routing accuracy, reducing response times from weeks to minutes and used daily by IT operations staff.
- Developed a real-time credit scoring system (89% accuracy) replacing a 5-day manual review process with 2-minute automated decisions, and built a recommendation engine serving 700,000+ banking customers that increased loan uptake by 15%.

*Lead Data Engineer (Consultant)*

- Built and deployed an end-to-end lending platform integrating live credit scoring, risk modelling and loan forecasting pipelines, disbursing loans to 1,500+ applicants annually.
- Established the Data and AI function from scratch, hiring and mentoring a team of 4 engineers and deploying production models on AWS using Docker, FastAPI and MLflow.

**Dataware Tech Ghana | Accra, Ghana**

01/2021 - 12/2022

*Lead Machine Learning Engineer (Consultant)*

- Owned the full data and ML lifecycle across 5+ production pipelines from problem framing through to deployment and monitoring, and led integration of AutoML and pipeline modules into the core BI platform as reusable assets across all clients.
- Partnered with 10+ business stakeholders across enterprise financial clients to translate workflow requirements into solution designs with measurable outcomes, running live working sessions to gather feedback and iterate on feature design.

**EDUCATION**

---

**MSc. Computer Science (Machine Learning Specialisation)**

08/2024 - Present

Georgia Institute of Technology

**BSc. Biomedical Engineering (First Class Honours)**

08/2014 - 06/2018

Kwame Nkrumah University of Science &amp; Technology

**TECHNICAL SKILLS**

---

**Data Engineering:** Python, SQL, PySpark, Airflow, Prefect, ETL/ELT Pipelines, Data Modelling, Pydantic, Great Expectations, Web Scraping (Selenium, BeautifulSoup), API Design

**AI / LLM:** RAG Architecture, LLM Agent Design, Prompt Engineering, AI Workflow Automation.

**Cloud & Storage:** GCP (BigQuery, GCS), AWS (EC2, S3, SageMaker), PostgreSQL, MongoDB, Redis

**Machine Learning:** XGBoost, PyTorch, Scikit-learn, NLP, Collaborative Filtering

**MLOps & Deployment:** MLflow, Docker, Kubernetes, FastAPI, GitHub Actions, Evidently AI

**KEY ACHIEVEMENTS**

---

**Zindi Hackathon Winner**

09/2021

*Best-performing insurance claim prediction model using XGBoost***Zindi Hackathon Runner-up**

05/2023

*Cryptojacking detection classifier, 3rd best F1 Score***Conference Speaker**

11/2021, 11/2022

*Africa Digital Skills Conference, training 30+ participants***RELEVANT CERTIFICATIONS**

---

AI Product Manager

Udacity, 2024

Data Scientist

Udacity, 2024

Business Intelligence &amp; Data Analyst

CFI, 2023

Machine Learning

Coursera, 2023

AWS Certified Cloud Practitioner

AWS, 2022

SQL

Udacity, 2024

AI Programming with Python

Udacity, 2020