

DESMOND ONAM

Nairobi, Kenya | +254 746 073 277 | desmondonam@gmail.com | [LinkedIn](#) | [Portfolio](#) | [GitHub](#)

DATA ENGINEERING | MACHINE LEARNING ENGINEERING | DATA ANALYTICS | AI ENGINEER

Machine Learning Data Engineer with 7+ years of experience developing and deploying ML pipelines, specializing in Large Language Models (LLMs) and Natural Language Processing applications. Proven expertise in Python, TensorFlow, PyTorch, and Scikit-learn with experience in multilingual AI systems and model optimization. Strong background in integrating ML models with cloud-based ETL data solutions and statistical modelling for predictive analytics. A knack in cloud solutions from IAM, CloudWatch, to EC2 deployment and monitoring of cloud based Architectural solutions.

- Python, TensorFlow, PyTorch.
- Pandas, Scikit-learn, Numpy
- Natural Language Processing
- Statistical Modelling & Predictive Analytics
- Generative AI
- Software Development
- Data Pipeline Management
- Java, Javascript, reactJS, Typescript
- LangChain, CrewAI
- Large Language Models (LLMs)
- Cloud Solution - AWS
- ML Pipeline Development & Deployment
- Database Management
- Cloud Computing - GCP
- Real-time Data Processing
- DevOps & CI/CD
- ETL Processes
- API – Restful API
- Multilingual AI Systems
- Performance Optimization
- Machine Learning
- Quality Assurance
- AI Agents
- Team Management
- Research
- Data Engineering
- VectorDB - Pinecone

KEY VALUE OFFERINGS

- **Data Engineering:** Engineered data pipelines for big data ingestion, feeding data into Amazon QuickSight from AWS Warehouse for advanced reporting and visualization using ETL pipeline and Machine learning models in EC2 instance.
- **Data Science:** Delivered outstanding Data Science and Analytics projects on time and guided interns in the department of Data Science as a Lead Data Scientist, resulting in 96% client satisfaction and best mentor at NextHikes.
- **Machine learning Engineering:** Optimized machine learning models by employing state-of-the-art models in MLOPS pipeline, using mlflow to refine models, fine-tune, choose the best performing ones and deploy them with docker/mlflow into AWS.
- **Large Language Models / Generative AI:** Led the development of creating a mental health chatbot using LLM with crew AI, by finetuning transformer models to be used, creating a data pipeline for retraining the model – Adopted by Bhutanese Government.
- **Pipeline Development:** Optimized model deployment pipelines that reduced time-to-market for AI solutions by 45%, using CI/CD practices and containerization with Docker and Kubernetes, creating MVP for ready market.

EXPERIENCE & ACHIEVEMENTS

HavarTech Solutions – |Georgia, USA – Remote | Full time

May, 2025 – Present

As an AI Data Engineer at HavarTech Solutions, I was instrumental in building, maintaining, and optimizing data pipelines that powered our core machine learning and AI applications. My work involved a hybrid of data engineering, machine learning, and cloud solutions engineering tasks, ensuring seamless and efficient data flow from ingestion to model deployment. I collaborated closely with data scientists to productionize models and with software engineers to integrate AI solutions into our products.

- Engineered and maintained a robust data infrastructure on AWS, utilizing services such as S3, Redshift, Glue, and EMR to handle large-scale data processing for machine learning model training and inference.
- Developed and deployed scalable ETL/ELT pipelines using Apache Spark and Databricks, reducing data processing time by 40% and ensuring timely delivery of clean, structured data for data science teams.
- Implemented CI/CD pipelines for MLOps using AWS CodePipeline and Jenkins, automating the deployment of machine learning models and data-related services, which improved deployment efficiency by 50%.
- Designed and managed data warehousing solutions using Redshift, optimizing query performance and access control for various business intelligence and analytics teams.
- Automated data ingestion from various sources (e.g., APIs, databases, streaming data) into a centralized data lake, ensuring data quality and availability for downstream applications.
- Collaborated with data scientists to productionize machine learning models, creating deployment scripts and services to integrate models into the existing product ecosystem, and monitoring their performance.
- Utilized Python for scripting and automation, and SQL for data manipulation and analysis, ensuring data integrity and consistency throughout the data lifecycle.

Omdena – | Palo Alto, USA**Dec 2024 – Present****• Product Manager – Mental Health and Well Being - Bhutan - May 2025 - Present**

Managed project on Mental health and well-being to develop a product, (MVP) for mental health and wellbeing in Bhutan through the application of NLP, LLM and Generative AI for Bhutan government thereby enhancing the Gross National Happiness Index.

• Data Science Trainer – Top Talent – Tanzania, Nepal. | Part time - March 2025- May 2025

Trained 150+ Data Science for top talent, to innovation challenges in Omdena, equipping professionals with Tech Level Skills to run a project end-to-end, using a practical project-based learning and optimizing the LMS by having auto grading reducing grading time by 70%. Instituted project-based curriculum adopted to train in different countries with Omdena, embracing modern technological Data Science skills, achieving a success rate of 97% students transitioning to Innovation challenges post the course.

• MyDay(Guardian Space) Data Engineer – Pipeline Fusion | Full time – Dec 2024 – March 2025

As a data Engineer, I led a successfully architecture of the data pipeline used in the AI-driven Disaster Management System presented to United Nations by implementation of scalable data pipelines using cloud-native technologies and serverless computing platforms, by extracting data from different sources – APIs and Sentinel-2, transforming the data by identification and labeling to different disasters – Flood, fire and earthquakes then loading the data in AWS for further Machine learning models to identify a disaster and alert.

• Omdena - Nepal - Data Science Trainer (Climate Change)– Nepal | Part time June 2024 – September 2024

As a Data Science trainer, with 120 Nepalese students, I led a project-based training of Data Science, Machine Learning and Data Engineering concepts to students, grading the assignment and led the project-based learning to identify the causes of Chronic Obstructive Pulmonary Disease (COPD) in Nepal, especially in Kathmandu province, providing a system to be used to identify the disease early and prevent condition leading to COPD.

• Omdena Academy – Machine Learning Data Engineer Trainer - Advanced| Volunteer| January 2024 – May 2024

As an Expert trainer in Machine Learning Data Engineering, I taught professional expert the tech-stack in Machine Learning – Data Cleaning and preparation, feature Engineering, model building and fine tuning, MLOPS – with mlflow, data Engineering – ETL with tool like Kafka, Airflow, Prefect and Docker and how to monitor pipelines, resolve pipeline failure – Leading to successful certification of 50 Machine Learning Data Engineers.

Ajua | Nairobi, Kenya**Jul 2024 – Nov 2024****Associate Machine Learning Data Engineer**

Served as an associate data scientist, responsible for engineering machine learning models, building dashboards, and optimizing data pipelines for customer experience by analyzing the engagement and satisfaction of customers in different companies. Collaborated with Customer Success Engineers and Managers to address and resolve customer experience challenges, report on customer experience and maintain the ETL pipelines for real time reporting of customer experience in embedded dashboards to client sites. As a reporting analytics, the dashboard I created helped to view the 3D of customers – Net Promoter Score (NPS), Customer Satisfaction Score (CSAT), Customer Effort Score (CES), Customer Lifetime Value (CLV) and Customer Churn Rate giving clients a 3D view of their customers. I crafted surveys, analyzed the survey responses, managed platform and retrained the old models to new ones, improved the pipeline and migrated system to AWS for scaling.

Nexthikes | Noida, Uttar Pradesh, India**Jun 2023 – Jun 2024****Lead Data Scientist**

As a project lead, I was in charge of discussing with the clients the needs of customers, translate them to technical problems, and led the team of data scientists and software engineer by creating and assigning tasks, leading standups, providing guidance on areas where there was skill gap and ensured projects are run with sprints (Agile) to completion within time and achieved a 96% client satisfaction and increased revenue.

**Henry Harvin | Noida, Uttar Pradesh, India
2024****Jul 2023 – Jul****Data Science Trainer**

Conducted hands-on training sessions covering key Data Science technologies, including Python, R, SQL, Apache Spark, and cloud-based data platforms. Guided data science students in completing real-world projects.

Key Contributions:

- **Collaborated with research teams on cutting-edge ML** methods and current Tech skills to craft a project-based learning curriculum and project instituted learning for students.
- **Designed and implemented scalable data pipelines and ETL workflows** to support real-world data science and machine learning projects while upskilling data science expertise of 215+ students.

- **Led development of 12 real-world data science projects** by guiding students in data extraction, transformation, and model deployment to solve complex business challenges.
- **Architected and deployed a comprehensive Masters in Data Science curriculum** by integrating cutting-edge data engineering, analytics, and cloud computing concepts, now adopted by College De Paris.
- **Organized industry-recognized certification programs** in collaboration with leading technology firms, enhancing students' credibility in data engineering and cloud computing.
- **Spearheaded the development of an online learning platform** with interactive coding exercises, Jupyter Notebooks integration, and video-based data engineering tutorials to support remote learning.

Omdena | Palo Alto, CA

Sep 2022 – Mar 2025

Junior Machine Learning Engineer | Machine Learning Data Engineer | Trainer

Built and deployed machine learning pipelines, incorporating Bayesian segmentation and statistical modeling techniques to enhance data-driven insights. Engineered scalable data pipelines to process and transform large datasets and enhanced data processing.

Key Contributions:

- **Designed and optimized ETL workflows** to automate data ingestion, transformation, and storage, reducing manual effort by 60% and minimizing errors.
- **Enhanced data processing infrastructure** by testing, optimizing, and implementing distributed computing frameworks such as Apache Spark and Kafka.
- **Implemented data wrangling and cleaning techniques** to improve data integrity, reducing preprocessing time and increasing model accuracy.

10 Academy | Santa Clara, CA

Apr 2022 – Aug 2022

Machine Learning Data Engineer Consultant/Tutor

Trained 37+ junior data engineers, machine learning Engineers and Web3 engineers in building products that underpin machine learning models, Data Engineering and designing a real-time data pipeline that fastened semi-structured data processing.

Key Contributions:

- **Instituted a project-based learning approach among machine learning and data engineering trainees** by enabling them to ingest data from multiple third-party APIs on real-world projects.
- **Built top talented, tech-savvy data engineers** through comprehensive and data-driven technological training that equipped them with machine learning, data engineering, and web3 skills for job readiness.
- **Liaised with data engineers** to expand and optimize data and pipeline architecture; taught students how to make well-reasoned business decisions fuelled by new data.
- **Introduced innovative teaching methodologies**, such as gamification and hands-on hackathons, fostering engagement and practical problem-solving skills among trainees.
- **Introduced a robust assessment framework** to track trainee progress and ensure alignment with learning objectives, resulting in a 20% increase in project completion rates.

EARLY CAREER HISTORY

Freelance Programmer – Upwork, Freelancer, Fiverr, TopTal | Freelancing | Nairobi, Kenya | Jul 2018 – Present

ICT Assistant | National Government Constituencies Development Fund | Nairobi, Kenya | Mar 2023 – Mar 2024

AI Machine Learning Engineer (Volunteer) | AICE Africa | Feb 2021 – Jun 2021

Trainer in Post Graduate advanced Certifications in Data Science and Machine Learning | SNA EduTech – Birchwood University | Sep 2022 – Aug 2024

Lead Data Science Trainer | Denaco Analytics (Part-time) | Nov 2023 – Apr 2024

Consultant Data Science Trainer | Zindua School (Part Time) | Jun 2024 | Dec 2024

EDUCATION & CERTIFICATION

Bachelor of Science in Mathematics and Computer Science | Jomo Kenyatta University of Agriculture and Technology | 2021

AWS Certifications | AWS | 2025

Machine Learning with Python | Start-Tech Academy | 2023

Machine Learning Engineering/ Data Engineering | 10 Academy | 2021

