**Training Curriculum: Data Science for Clinicians and Researchers**

**Duration:** 5 Days

**Target Audience:** Clinicians, Health Researchers, Data Scientists

**Format:** Lectures, Keynotes, Hands-on Sessions, Group Discussions

**Delivery Mode:** In-person, instructor-led, interactive group work

**Day 1: Introduction to Data Science in Health Research**

**Learning Objectives:**

* Understand the role of data science in health research.
* Learn about the basic tools used in data science.
* Get introduced to AI/ML applications in healthcare.
* Define a real-world health problem for group research.

**Topics & Activities:**

1. **Workshop Launch & Overview**

* ENEZA Project Introduction
* Workshop Goals & Objectives
* Keynote Address: The Role of Data Science in Health Research
* Introductions & Expectations – Participants & Facilitators

1. **Lecture: Technical Tools for Data Science**

* Overview of data analysis tools used in health research.
* Hands-on: Basic Data Analysis in R
* Introduction to R and its applications.
* Running basic descriptive statistics.
* Exploring data visualization techniques.

1. **Group Work: Identifying AI/ML Use Cases in Health**

* Break into small groups (5 per group).
* Define a real-world health problem in Kenya.
* Discuss data acquisition strategies and challenges.

1. **Group Presentations**

* Each group presents a 5-minute summary of their problem definition and data acquisition plan.

**Day 2: AI/ML Foundations & Data Management**

**Learning Objectives:**

* + Learn the fundamentals of AI/ML in healthcare.
  + Understand how to manage health research data effectively.
  + Apply AI/ML methods to real-world datasets.

**Topics & Activities:**

1. **Recap & Introduction to AI/ML**

* AI/ML Basics for Healthcare
* Types of ML models (Supervised, Unsupervised, Reinforcement Learning).
* AI techniques for structured & unstructured data.
* Use cases: Predictive modeling, disease diagnosis, and medical imaging.
* Keynote: AI/ML in Digital Health – Philip Oyier, JKUAT
* Q&A Session – Interactive discussion on AI/ML in clinical settings.

1. **Lecture: Data Management in Health Research**

* Data Collection & Preprocessing
* Handling missing data.
* Data integration from multiple sources.
* Ensuring data quality and consistency.
* Ethical & Privacy Considerations
* Health data regulations (HIPAA, GDPR).
* De-identification and data anonymization techniques.

1. **Group Work: Applying AI/ML Methods**

* Groups update their projects by integrating AI/ML methods learned in the morning.
* Discuss data preprocessing & cleaning techniques.
* Facilitators provide guidance & feedback.

**Day 3: FAIR Data, Open Science & Research Applications**

**Learning Objectives:**

* Learn the FAIR Data Principles (Findable, Accessible, Interoperable, Reusable).
* Explore open science resources for data sharing.
* Analyze a data science research paper.

**Topics & Activities:**

1. **Recap & Lecture on Open Science & FAIR Data**

* Importance of FAIR data in health research.
* Open-source tools & repositories (e.g., Zenodo, Dryad, Figshare).

1. **Case Study: AI in Antimicrobial Resistance**

**Paper Review & Discussion**

"Using Machine Learning to Predict Antimicrobial Resistance" – Sakagianni et al. (2023).

Key takeaways: AI applications, model selection, dataset challenges.

1. **Keynote: Ethical, Legal, and Social Issues (ELSI)**

* Bias & Harms in AI for Health
* Addressing algorithmic bias and fairness.

1. **Group Work: Refining Project Discussions**

* Groups integrate FAIR data principles into their project designs.
* Identify open datasets relevant to their problems.
* Update project plans based on insights gained.

**Day 4: Data Science in Clinical Research & Cloud Computing**

**Learning Objectives:**

* Learn how data science is applied in clinical trials.
* Understand cloud-based data management for healthcare.
* Explore real-world health datasets.

**Topics & Activities:**

1. **Recap & Case Study: AI in Clinical Research**

* Discussion: Data Science in Diabetes Research
* How open-source health datasets can be used for research.

1. **Keynote: Data Science & Clinical Trials**

* The role of data science in modern clinical trials.

1. **Lecture: Cloud-Based Data Management**

* Data Storage & Processing on the Cloud
* AWS, Google Cloud, and Azure for healthcare data.
* Managing big data & scalability.
* Hands-on: Setting up cloud data pipelines.

1. **Panel Discussion: AI for Health in Africa**

**Panelists:** Dr. Aisha Walcott-Bryant, Dr. William Ogallo, Dr. Sekou Remy (Google Research Africa)

**Moderator:** Geoffrey Siwo, University of Michigan

**Topics:**

* The future of AI in African healthcare.
* Challenges in AI adoption in low-resource settings.

**Day 5: Future Trends & Project Presentations**

**Learning Objectives:**

* Present and receive feedback on mini-projects.

**Topics & Activities:**

* Group Project Presentations
* Each group presents their final project.
* Peer Review & Expert Feedback.
* Closing Session
* Certificates & Recognition
* Final Remarks
* Group Photo & Networking

**Assessment & Certification**

* Participation in group discussions & hands-on sessions
* Completion of a group mini-project
* Final presentation on Day 5
* Certificate of completion awarded