LIXUS: AI-Driven Pseudo Multi-Omics data for AI-Integration for Disease Prediction

This internship will focus on closing the gap between real and simulated multi-source health data leveraging AI-driven approaches to improve a disease prediction. The intern will work on integrating simulated electronic health records (EHR), genetic, biochemical, imaging, and lifestyle data to enhance predictive modeling. We will try to:

- Generate baseline synthetic multi-omics data using statistical distributions.
- Introduce realistic variation using AI-driven generative models (GANs, VAEs).
- Incorporate temporal modeling for disease progression.
- Test and validate predictions using real-world biobank data.

Objective: This project aims to generate artificial multi-omics data to develop and test AI models for disease prediction. By integrating diverse datasets—genetics, biochemical markers, transcriptomics, proteomics, electronic health records (EHR), lifestyle factors, iDXA, and imaging data—we will construct a synthetic yet biologically realistic dataset for training machine learning models.
The project will begin by simulating multi-omics data using statistical methods and generative AI models to reflect real-world patient profiles. Each data type will be synthesized to maintain realistic correlations, mimicking patterns observed in real populations. Next, a machine learning pipeline will be designed to predict disease risk using the generated dataset. The AI framework will integrate multi-modal inputs, enabling personalized risk assessment for cardiometabolic diseases. By the end of the project, we will assess the model's performance and refine the synthetic dataset for future applications.

Project Type: Research

Internship Batch:

- **Batch 1:** May 11 to July 10, suitable for Education City students, i.e., CMUQ, TAMUQ and HBKU students

Expected Outcomes:

1. A well-structured artificial dataset that integrates multiple biological and clinical data types.
2. A working AI-based disease prediction model trained on synthetic data.
3. Insights into how this technology can enhance precision medicine research.

Learning Opportunities:
This project provides hands-on experience in AI-driven multi-omics integration, data generation, and predictive modeling, making it ideal for students interested in computational biology and precision health.

Expected Learning Outcomes

- Exposure to multi-omics data integration.
- Hands-on experience with AI-driven predictive modeling in health research.
- Understanding of risk stratification and polygenic risk score.
- Training in coding (Python/R), machine learning

Preferred Skills

- Basic programming (Python/R).
- Interest in bioinformatics, AI in precision health.

Mentors

Name: Dr. Halima Bensmail email: hbensmail@hbku.edu.qa