PS 2.4

DATA IS POWER! CONFRONTING DATA COLONIALISM, OWNERSHIP ISSUES AND HIDDEN BIASES

| BACKGROUND

Data is an essential resource, and just as important to health and development as any other, including energy, minerals, and raw materials. In the health sector, data is critical to analysis and advancement of all kinds. Data is increasingly important as more and more technologies and services, and innovations are built on it. Everything from generative AI to pharmaceutical trials to effective regulation is centered on data. Accordingly, the control of data is increasingly contested, whether because of commercial interests, public goods, or as the foundation of democratic societies. The growing contestation in all areas, including in population health and health care creates many challenges. include equity and ethical controversies such as data colonialism, ownership of data, and biases from inequities in data collection, management, use and application.

Digital coloniality refers to the unfair and unequal exercise of power over data and its benefits, where data can be seen as similar to any other extractive industry. As with gold and other precious ores, data is mined in developing countries and sent to rich countries, where it is refined into credit-bearing forms and used for career advancement, exploited economically, and controlled with legal and technical measures. In general, this means that the use of data benefits technology conglomerates and other organizations located in the global north, with few if any benefits accruing to the people whose bodies, lives, and experiences comprise the original resource. Moreover, with the development partners and international organizations being the most frequent users with high levels of demand and main financiers of information systems, concerns arise that this may crowd out domestic needs and utilization by and for these countries. The claim of local ownership rises as an issue, particularly in access to and application of data for health benefits. On the other hand, local capacity in data collection, management, and analyses remains a test to the local ownership issue.

Adding to this concern is the understanding of local context in cases of data science experts who are geographically or socio-demographically distant from the data subjects. Unintended (and sometimes intentional) bias against vulnerable populations and representation bias can arise from such situations. Additionally, the development of advanced data infrastructures such as machine learning and artificial intelligence may focus more on machine-readable data while overlooking the local forms and data on paper, which is still used in many low- and middle-income countries.

To address these expanding issues, strategies that consider domestic health needs and address representation and ownership problems are needed to improve the availability, accessibility, and usefulness of data local capacity in data collection and use to ensure the equitable and ethical application of quality data for health. Quality and balance of datasets, inclusion of gender dimensions and socio-cultural aspects, as well as dedicated public policies will need to be considered to tackle the biases, particularly against the vulnerable populations.

This session will confront the data colonialism, ownership, and hidden biases that come with the power of data and posit the reasons behind them. Upon better understanding the situations, speakers will also discuss actionable recommendations to minimize concerns in the data science world.

| OBJECTIVES

- To understand the shifting challenges and concerns regarding equity and ethical issues related to data use in technology and AI for health
- To explore solutions and resources needed to ensure that equity, ethics, and gender considerations are effectively incorporated into the design, implementation and oversight of data for health
- To examine the enabling environment and local capacity needed to ensure the availability of data and information systems to promote equitable access to health for all especially in vulnerable populations and low-income countries

Speaker / Panelist

Leo Anthony Celi

Clinical Research Director and Principal Research Scientist

MIT Laboratory of Computational Physiology (LCP)
United States of America

Dr. Celi is the principal investigator behind the Medical Information Mart for Intensive Care (MIMIC) and its offsprings, MIMIC-CXR, MIMIC-ED, MIMIC-ECHO, and MIMIC-ECG. With close to 100k users worldwide, an open codebase, and close to 10k publications in Google Scholar, the datasets have undoubtedly shaped the course of machine learning in healthcare in the United States and beyond. His group has written 3 open-access textbooks: "Secondary Analysis of Electronic Health Records" in 2016, "Global Health Informatics: Principles of eHealth and mHealth to Improve Quality of Care" in 2017, and "Leveraging Data Science for Global Health" in 2020. The first has been downloaded over 1.7 million times and translated into Mandarin, Spanish, Korean and Portuguese. The group has created two open online courses, "Global Health Informatics'' and "Collaborative Data Science for Healthcare". Finally, in partnership with hospitals, universities and professional societies across the globe, Dr. Celi and his team have organized over 50 datathons in 22 countries, bringing together students, clinicians, researchers, and engineers to leverage data routinely collected in the process of care.