





## DISTINGUISHED LECTURE

## Navigate the Crossroad of Statistics, Generative AI and Genomic Health



## Professor Xihong LIN

Professor of Statistics, Professor of Biostatistics TH Chan School of Public Health Harvard University

COPSS Presidents' Award
APHA Mortimer Spiegelman Award
NCI Outstanding Investigator Award
Elected member, US National Academy of Sciences
Elected member, US National Academy of Medicine
Elected fellow, American Statistical Association
Elected fellow, Institute of Mathematical Statistics
Elected fellow, International Statistical Institute

Date: 9 July 2025 (Wednesday) Time: 2:30 pm — 3:30 pm

Venue: Room 103, Y C Liang Hall,

The Chinese University of Hong Kong

## **Abstract**

Scalable and robust statistical methods empowered by generative AI offer unprecedent potentials for trustworthy science as they quantify uncertainty, enhance interpretability, and accelerate scientific discovery. In this talk, I will discuss the challenges and opportunities as we navigate the crossroad of statistics, generative AI, and genomic health science. I will discuss robust and powerful statistical analysis by leveraging synthetic data generated by generative AI models, such as diffusion models and transformer, while ensuring valid statistical inference when generative AI models are misspecified. I will illustrate key points using the analysis of large scale biobanks, whole genome sequencing data, and electronic health records, and demonstrate the power of scientific discovery by integrating statistics and generative AI using synthetic data. I will also discuss how to conduct scalable and interpretable large-scale whole genome sequencing (WGS) data, and illustrate the WGS analysis ecosystem using the TOPMed WGS samples of 200,000, the UK biobank of 500,000 subjects in the cloud platform RAP and as well the All of Us data of 400,000 subjects in the NIH cloud platform AnVIL.

\*\*\* All are welcome \*\*\*

