香港浸會大學 HONG KONG BAPTIST UNIVERSITY

DEPARTMENT OF COMPUTER SCIENCE 計算機科學系

DEPARTMENT OF COMPUTER SCIENCE

SEMINAR

2025 SERIES

Explanatory Debiasing: Mitigating Representation Bias Through Human-AI Interaction

DATE & TIME

5 AUG 2025 (TUE) 3:00 - 4:00 PM

VENUE

DLB637, 6/F, DAVID C. LAM BUILDING, SHAW CAMPUS

MR. ADITYA BHATTACHARYA

Lead Researcher
Explainable Artificial Intelligence
KU Leuven

ABSTRACT

Representation bias remains a persistent challenge in the development of fair and reliable AI systems, often leading to suboptimal performance on underrepresented data segments. While various technical strategies exist to mitigate this bias, their effectiveness is frequently limited by a lack of domain-specific insights during the debiasing process. This talk presents a novel approach to bridging this gap by introducing a set of generic design guidelines aimed at effectively integrating domain experts into the representation debiasing workflow.

We demonstrate the practical application of these guidelines through a healthcare-focused case study, evaluated via a mixed-methods user study involving 35 healthcare professionals. Our findings reveal that structured involvement of domain experts not only reduces representation bias but also maintains model accuracy. The talk will conclude with actionable recommendations for AI developers to design more inclusive and robust debiasing systems, grounded in our proposed framework.

SPEAKER'S BIOGRAPHY

REGISTER NOW