CENTRE FOR STUDIES IN SOCIAL SCIENCES, CALCUTTA

R-1 BAISHNABGHATA PATULI, KOLKATA 700094

ECONOMICS STUDY GROUP SEMINAR

Haimanti Bhattacharya

shall present

The Good, the Bad and the Ugly: Effects of AI Quality Information on Detecting Text-Based Lies

Haimanti Bhattacharya
The University of Utah,
Salt Lake City, USA

Subhasish Dugar
The University of Utah,
Salt Lake City, USA

Sanchaita Hazra
The University of Utah,
Salt Lake City, USA

Bodhisattwa Majumder
Allen Institute for AI,
Seattle, USA

Abstract

We experimentally investigate whether AI advisors of varying efficacy can help people distinguish between truth and lies in written text. We design AI advisors with low, medium, and high efficacy and either reveal or conceal their effectiveness from subjects. We utilize transcripts of conversations from the TV show "To Tell The Truth" to mitigate the influence of motivated reasoning and simulate social media exchanges marred by lies between parties with conflicting agendas on a topic with an objective truth. We discover that while subjects marginally outperform chance in detecting truth, expectations regarding AI efficacy greatly impact their reliance on AI advisors. As the efficacy of low and medium-quality AI advisors falls short of their expectations and remains undisclosed, subjects' overreliance on them causes the truth-detection rate to descend below their intrinsic ability. Upon disclosing AI efficacy, subjects reduce their reliance, which improves truth detection. The high-quality AI advisor whose efficacy matches the subjects' beliefs enhances truth detection, regardless of whether its efficacy is disclosed. These findings highlight the risks of undisclosed AI efficacy, which can exacerbate misinformation and underscore the urgent need for transparent AI policies.

Date: 6th September 2024 (Friday)
Time: 3 P.M.

Venue: CSSSC Room Number 314 (second floor), Patuli Campus
All are welcome to attend

Dr. Sattwik Santra
(Co-Convener, Seminar Sub-Committee)