Deepfakes, Disinformation, & Democracy: Strategies to Support Election Integrity in the Age of AI

Wednesday, October 21st
12:00 - 1:00pm

In light of the community concerns regarding COVID-19, this talk will be given as a webinar. The link will be provided on October, 20th to those that have registered by 5:00 pm on Monday, October 19th at uccs.ucdavis.edu.

Elections are increasingly susceptible to AI-enabled threats—from the spread of malicious deepfakes intended to sew discord to the surreptitious manipulation of public opinion on social media through disinformation. This presentation provides examples of the use of deepfakes and disinformation to influence elections and the legislative and industry actions that have been taken to mitigate their harmful effects. The presentation concludes with recommendations for appropriate technology and policy strategies that can be implemented by the public and private sectors to better support election integrity in the age of AI.

Brandie Nonnecke, PhD is Founding Director of the CITRIS Policy Lab, headquartered at UC Berkeley. Brandie has expertise in information and communication technology (ICT) policy and internet governance. She studies human rights at the intersection of law, policy, and emerging technologies with her current work focusing on fairness, accountability, and appropriate governance mechanisms for AI. She is a Technology and Human Rights Fellow at the Carr Center for Human Rights Policy at the Harvard Kennedy School. She served as a fellow at the Aspen Institute’s Tech Policy Hub and at the World Economic Forum on the Council on the Future of the Digital Economy and Society. She was selected as a 2018 RightsCon Young Leader in Human Rights in Tech and received the 2019 Emerging Scholar Award at the 15th Intl. Common Ground Conference on Technology, Knowledge, and Society. Her research has been featured in Wired, NPR, BBC News, MIT Technology Review, Buzzfeed News, among others. Her research publications, op-eds, and presentations are available at nonnecke.com.