Data, Responsibly

Data-driven algorithmic decision making promises to improve people's lives, accelerate scientific discovery and innovation, and bring about positive societal change. Yet, if not used responsibly, this same technology can reinforce inequity, limit accountability and infringe on the privacy of individuals: irreproducible results can influence global economic policy; algorithmic changes in search engines can sway elections and incite violence; models based on biased data can legitimize and amplify discrimination in the criminal justice system; algorithmic hiring practices can silently reinforce diversity issues and potentially violate the law; privacy and security violations can erode the trust of users and expose companies to legal and financial consequences**.**

In this talk I will discuss our recent work on establishing a foundational new role for database technology, in which managing data in accordance with ethical and moral norms, and legal and policy considerations becomes a core system requirement. I will define properties of responsible data management, which include fairness, transparency, and data protection. I will highlight some of our recent technical advances, and will discuss the over-all framework in which these responsibility properties are managed and enforced through all stages of the data lifecycle. The broader goal of our project is to help usher in a new phase of data science, in which the technology considers not only the accuracy of the model but also ensures that the data on which it depends respect the relevant laws, societal norms, and impacts on humans. Additional information about our project is available at DataResponsibly.com.