Data Science Technology and Governance Challenges During Crisis Response

Srinivasan Parthasarathy
Professor, Computer Science and Engineering and Biomedical Informatics
Director Data Mining Research Laboratory
Co-Director, Responsible Data Science, Translational Data Analytics Institute
The Ohio State University
Contact: srini@cse.ohio-state.edu

Position Statement: The modern economy is increasingly reliant on our ability to generate and store large tracts of data and realize actionable insights from this data. The processes by which these insights are discerned and subsequently shared are often complex -- requiring multiple transformative steps. These complex multi-step processes in turn can lead to several sources of risk at each step, for which existing governance, economic and regulatory frameworks are insufficient. Acting on the insights generated by such processes can lead to entirely new sources of risk. For instance, new sources of risk can range from fairness with respect to artificial intelligence (AI) based decision making to illegal uses of collected data, and from the amplification of bias against a particular community to the unethical use of data for which consent was never sought from end-users (producers) of the data.

If such risks are not mitigated then it in turns leads to an erosion of trust with respect to the entity in question that stores and analyzes this digital data -- whether government, multi-national company, or organization. This erosion of trust can have substantive scientific, societal and geo-political consequences. Exemplars abound -- one has only to open the headline pages of newspapers worldwide to see a multitude of examples While security, confidentiality and privacy are important considerations and have rightfully received increased attention -- we believe that there needs to be a growing convergence of ideas that spans disciplines including Philosophy, Law, Business, Complex Systems, Economics and Computer Science to emphasize the roles of ethics and trust in this discussion around *responsible data science.*

I posit that there is a need for a rethink around mechanisms to mitigate, contain and manage such risk in the modern big data age. Treating data in an ethical manner throughout its lifecycle of transformations from raw data to insights is essential. Moreover, understanding, characterizing and modeling the implications and importance of digital trust is of paramount importance in this context to help better quantify the cost of associated risks. From the research perspective my interests in these areas revolve around technological mechanisms to mitigate such bias or risk through explainable and auditable data science and AI algorithms and systems. I also have a strong interest in understanding the impact of (possibly biased) decision-making on underprivileged and under-represented populations.

Of particular interest here is the role of such decision-support tools both during and immediately following crises -- whether from natural hazards (e.g. hurricane induced flooding) and pandemics. The current pandemic crisis offers several illustrative examples with some of these challenges ranging from efficient data preprocessing to timely-access to pre-processed data; from the role of sociological interpretivism in both building and

effectively leveraging decision support tools in emergent crisis environments to modeling challenges in the presence of uncertainty and accessibility challenges and from understanding the impact on critical infrastructure as well as impact on underprivileged communities. Personally in my research I have seen some of these challenges come to the fore both regionally within the State of Ohio (Opioid crisis, Covid19 pandemic) as well as nationally and internationally -- in some cases coupled with natural disasters (see for example the dual impact of hurricane-induced flooding while dealing with the ongoing pandemic, on Charleston, SC) further exacerbating such challenges. In addition to enhancing technological solutions I have together with colleagues, examined some of these issues through the lens of corporate data governance. In this context I am interested in better understanding some of these issues from the perspective of public health, policy and government organizations and open data portal managers at various levels: from community to regional and from national to international.

On the educational front - both as a regular instructor on data science courses as well as in my role as co-director of the undergraduate data analytics major at OSU I would like to both ensure and enable the next generation of data scientists we train at OSU to think and act upon the critical issues underpinning responsible data science. As one of the co-director(s) of OSU's Translational Data Analytics (TDAI) community of practice on Responsible Data Science, I also facilitate discussions around this important topic - both within OSU and externally to better understand and effectively develop research and educational initiatives around this important topic. I look forward to engaging with other workshop participants on this topic.