

Letter from the Special Issue Editors

As Artificial Intelligence (AI) and Machine Learning (ML) are increasingly being used to make consequential decisions that may critically impact individuals and many aspects of our society such as criminal justice, health care, education, and employment, there is a growing recognition that AI systems may perpetuate or, worse, exacerbate the unfairness of existing social systems. To tackle this problem, there has been a sharp recent focus on fair AI/ML research in the Machine Learning community. These efforts tend to focus on how to engineer fair AI models by (a) defining appropriate metrics of fairness and (b) designing new algorithms to mitigate bias and ensure fairness. It however frequently overlooks the messy, complex and ever changing contexts in which these systems are deployed. As AI fairness is a complex socio-technical issue which cannot be addressed by a purely technical solution, designing fair AI systems requires close collaboration across multiple disciplines to deeply integrate the social, historical, legal, and technical context and concerns in the design process.

In this special issue on **Interdisciplinary Perspectives on Fairness and Artificial Intelligence Systems**, leading researchers from engineering, social science and humanities present their work on ethical considerations in developing fair AI systems specifically, and responsible socio-technical systems in general. We sought high-quality contributions that integrate ideas from more than one field across the disciplines of technology, social science and the humanities. These papers will provide readers deep insights into the nature and complexity of AI biases, their manifestations in developing practical socio-technical systems and typical mitigation strategies. They also identify opportunities for the AI community to engage with the experts from the humanities.

Safiya U. Noble and *Sarah T. Roberts* from UCLA seek to expand the conversations about socio-technical systems beyond individual, moral and ethical concerns. They believe that the field of “ethical AI” must contend with how it affects and is affected by power structures that encode systems of sexism, racism, and class. They advocate the need for independent research institutes, such as the UCLA Center for Critical Internet Inquiry (C2i2) to promote investigations into the politics, economics, and impacts of technological systems.

Jennifer Keating from the University of Pittsburgh discusses the importance of incorporating ethical standards and responsible design features into the development of new technologies. Covid contact tracing was used as a case study to illustrate how rapidly developed tools can have unintended consequences if they are not carefully designed/monitored. She advocates that technologists should collaborate with experts from the humanities to integrate deeper cultural concerns and social/political context into technology development.

Lisa Singh and her co-authors from Georgetown University overview the challenges associated with using social media data and the ethical considerations they create. They frame the ethical dilemmas within the context of data privacy and algorithmic fairness and show how and when different ethical concerns arise.

Sebastian Schelter from University of Amsterdam and *Julia Stoyanovich* from NYU present a discussion on technical bias that arises in the data processing pipeline. A number of potential sources of bias during the preprocessing, model development and deployment phases of the ML development lifecycle are identified, with illustrative examples. They show how software support can help avoid these technical bias issues. The broader point of the work is to shed light on the challenge of bias introduced due to data engineering decisions, and to promote an emerging research direction on developing solutions to mitigate it.

James Foulds and *Shimei Pan* from UMBC aim to shed light on whether parity-based metrics are valid measures of AI fairness. They consider the arguments both for and against parity-based fairness definitions and provide a set of guidelines on their use in different contexts.

Finally, *Jared Sylvester* and *Edward Raff* from Booz Allen Hamilton argue that good-faith efforts toward implementing fairness in practical ML applications should be encouraged, even when the fairness interventions may not result from completely resolving thorny philosophical debates, as this represents progress over the (likely unfair) status quo. This viewpoint is discussed in several contexts: choosing the right fairness metric, solving trolley problems for self-driving cars, and selecting hyper-parameters for fair learning algorithms.

James Foulds and Shimei Pan
University of Maryland, Baltimore County