

## Letter from the Special Issue Editor

The big data revolution and advancements in machine learning technologies have revolutionized decision making, advertising, medicine, and even election campaigns. Data-driven software now permeates virtually every aspect of human activity and has the ability to shape human behavior: it affects the products we view and purchase, the news articles we read, the social interactions we engage in, and, ultimately, the opinions we form. Yet, data is an imperfect medium, tainted by errors, omissions, and biases. As a result, discrimination shows up in many data-driven applications, such as advertisements, hotel bookings, image search, and vendor services. In this issue, we bring together an exciting collection of recent and ongoing work that focuses on the problems of fairness, diversity, and transparency in data-driven systems. This collection highlights the central role that the data management research community can play in detecting, informing, and mitigating the effects of bias, skew, and misuse of data, and aims to create bridges with work in related communities.

We start with “Nutritional Labels for Data and Models”, by Stoyanovich and Howe. This paper argues for informational and warning labels for data, akin to nutritional labels, that specify characteristics of data and how it should be consumed. These nutritional labels help humans determine the fitness of models and data, aiding the interpretability and transparency of decision-making processes.

The second paper, “Data Management for Causal Algorithmic Fairness”, by Salimi, Howe, and Suci, provides a brief overview of fairness definitions in the literature, and argues for the use of causal reasoning in defining and reasoning about fairness. The paper exposes a vision of the opportunities of applying data management techniques, such as integrity constraints, query rewriting, and database repair to enforcing fairness, detecting discrimination, and explaining bias.

In the third paper, “A Declarative Approach to Fairness in Relational Domains”, Farnadi, Babaki, and Getoor focus on notions of fairness that capture the relational structure of a domain, and propose a general framework for relational fairness. Fairness-aware probabilistic soft logic includes a language for specifying discrimination patterns, and an algorithm for performing inference under fairness constraints.

The next paper, “Fairness in Practice: A Survey on Equity in Urban Mobility”, by Yan and Howe, places its focus on practical societal implications of fairness in the domain of transportation. The paper presents the findings of equity studies in mobility systems, such as bike-sharing and ride-hailing systems, and reviews experimental methods and metrics.

Again motivated by the societal implications of fairness and diversity, Benabbou, Chakraborty, and Zick put their sights on the allocation of public resources. “Fairness and Diversity in Public Resource Allocation Problems” focuses on two real-world cases, the allocation of public housing in Singapore and public school admissions in Chicago, models them as constrained optimization problems, and analyzes the welfare loss in enforcing diversity.

We conclude with “Towards Responsible Data-driven Decision Making in Score-Based Systems”, by Asudeh, Jagadish, and Stoyanovich. The paper focuses on designing fair and stable rankings, and discusses how these technologies can assess and enhance the coverage of training sets in machine learning tasks.

Thank you to all the authors for their insightful contributions, which bring into focus new and exciting challenges, and identify opportunities for data management research to contribute tools and solutions towards critical societal issues. Thank you also to Haixun Wang for his valuable assistance in putting together the issue. I hope you enjoy this collection.

Alexandra Meliou  
University of Massachusetts, Amherst