

Anna Glass  
Professor Frazier  
DATA 150-01 Evolving Solutions  
24 August 2020

The request for humbler data science implies that the contemporary goals of big data, such as the rapid distribution of aid money or the use of digital credit scores, are unattainable. Instead of humbler data science, I call for a more considerate data science. Big data promises a rapid means of distributing aid money, but developers must first consider two factors; local contexts and their susceptibility to data misrepresentation, and the impact of machine error.

As applied to international aid, the ultimate goal of data is the fair distribution of money in a rapid and effective manner. For this reason, it is essential to consider the relationship between efficiency and efficacy in big data. If an algorithm can quickly determine which subsets of a population require aid, it has increased the overall efficacy of an international organization's intervention efforts. However, if this evaluation of data is inaccurate, this negates the positive impact of rapid action with unfairness and improper allocation of scarce resources. Blumenstock argues that, "The flaws in the new approaches are not well understood. There is a risk that such tools will be deployed before they have been adequately tested," which highlights the fear that, although many methods of data collection are imperfect, big data is volatile. The criminal justice system in the United States has already demonstrated the adverse effects of inaccurate machine learning. The machine learning algorithms implemented in the United States are largely trained with European faces, which reduces their ability to distinguish between faces with darker complexions and inserts racial bias into their system of finding suspects through facial recognition. Robert Williams of Michigan was wrongfully arrested for felony larceny with the only piece of evidence being inaccurate facial recognition. Machine error in both of these contexts can be devastating, and data science must evaluate the impact of this machine error before implementing software on the local or international scale.

Joshua Blumenstock proposes three tenets of big data collection that could shift big data from unpredictability to a viable method for providing aid: validate, customize, and deepen collaboration. Blumenstock argues, "Next-generation solutions must be designed and produced by people who understand the problems and context – not just those who understand the algorithms." Machine learning algorithms have the ability to detect shifts in wealth, but they often fail to account for cultural or seasonal shifts in phone usage. To confront this, data scientists must collaborate with local experts to reevaluate algorithms that would account for these shifts in society. Algorithmic transparency is a double-edged sword, with one side being trust and the other cynicism. A cynic would argue that algorithmic transparency would lead to the informed population gaming the system for their own benefit, which often places those in poverty at a disadvantage. However, a lack of transparency would require deep trust that the powerful few that control the algorithm would prevent bias and unfair practices within the

algorithm, as “the power to derive value from the data tends to be concentrated in the hands of a few.” The applications of data science in the realm of humanitarian aid are largely uncharted compared to its uses in industry or technology. Blumenstock’s suggestions for humbler data science have a common thread— bring the “humanity” back into algorithms.

The use of big data for development faces several obstacles, including the ability of entities to corrupt or misrepresent data, the lack of regulations surrounding software implementation, and the ever-changing nature of socioeconomic and cultural contexts. However, a more considerate data science can combat these impediments to scientific and humanitarian progress.