

<https://onezero.medium.com/the-seductive-diversion-of-solving-bias-in-artificial-intelligence-890df5e5ef53>

# The Seductive Diversion of ‘Solving’ Bias in Artificial Intelligence

Trying to “fix” A.I. distracts from the more urgent questions about the technology

By Julia Powles and Helen Nissenbaum

Dec 7, 2018



Image courtesy of Gary Zamchick, strategic designer in residence at Cornell Tech.

The rise of Apple, Amazon, Alphabet, Microsoft, and Facebook as the world’s most valuable companies has been accompanied by two linked narratives about technology. One is about artificial intelligence—the golden promise and hard sell of these companies. A.I. is presented as a potent, pervasive, unstoppable force to solve our biggest problems, even though it’s essentially just about finding patterns in vast quantities of data. The second story is that A.I. has a problem: bias.

The tales of bias are legion: online ads that show men higher-paying jobs; delivery services that skip poor neighborhoods; facial recognition systems that fail people of color; recruitment tools that invisibly filter out women. A problematic self-righteousness surrounds these reports: Through quantification, of course we see the world we already inhabit. Yet each time, there is a sense of shock and awe and a detachment from affected communities in the discovery that systems driven by data about our world replicate and amplify racial, gender, and class inequality.

Serious thinkers in academia and business have swarmed to the A.I. bias problem, eager to tweak and improve the data and algorithms that drive artificial intelligence. They've latched onto fairness as the objective, obsessing over competing constructs of the term that can be rendered in measurable, mathematical form. If the hunt for a science of computational fairness was restricted to engineers, it would be one thing. But given our contemporary exaltation and deference to technologists, it has limited the entire imagination of ethics, law, and the media as well.

There are three problems with this focus on A.I. bias. The first is that addressing bias as a computational problem obscures its root causes. Bias is a social problem, and seeking to solve it within the logic of automation is always going to be inadequate.

Second, even apparent success in tackling bias can have perverse consequences. Take the example of a facial recognition system that works poorly on women of color because of the group's underrepresentation both in the training data and among system designers. Alleviating this problem by seeking to "equalize" representation merely co-opts designers in perfecting vast instruments of surveillance and classification.

When underlying systemic issues remain fundamentally untouched, the bias fighters simply render humans more machine readable, exposing minorities in particular to additional harms.

Third—and most dangerous and urgent of all—is the way in which the seductive controversy of A.I. bias, and the false allure of "solving" it, detracts from bigger, more pressing questions. Bias is real, but it's also a captivating diversion.

What has been remarkably underappreciated is the key interdependence of the twin stories of A.I. inevitability and A.I. bias. Against the corporate projection of an otherwise sunny horizon of unstoppable A.I. integration, recognizing and acknowledging bias can be seen as a strategic concession—one that subdues the scale of the challenge. Bias, like job losses and safety hazards, becomes part of the grand bargain of innovation.

The reality that bias is primarily a social problem and cannot be fully solved technically becomes a strength, rather than a weakness, for the inevitability narrative. It flips the script. It absorbs and regularizes the classification practices and underlying systems of inequality perpetuated by automation, allowing relative increases in "fairness" to be claimed as victories—even if all that is being done is to slice, dice, and redistribute the makeup of those negatively affected by actuarial decision-making.

In short, the preoccupation with narrow computational puzzles distracts us from the far more important issue of the colossal asymmetry between societal cost and private gain in the rollout of automated systems. It also denies us the possibility of asking: Should we be building these systems at all?

The endgame is always to “fix” A.I. systems, never to use a different system or no system at all.

In accepting the existing narratives about A.I., vast zones of contest and imagination are relinquished. What is achieved is resignation—the normalization of massive data capture, a one-way transfer to technology companies, and the application of automated, predictive solutions to each and every societal problem.

Given this broader political and economic context, it should not surprise us that many prominent voices sounding the alarm on bias do so with blessing and support from the likes of Facebook, Microsoft, Alphabet, Amazon, and Apple. These convenient critics spotlight important questions, but they also suck attention from longer-term challenges. The endgame is always to “fix” A.I. systems, never to use a different system or no system at all.

Once we recognize the inherently compromised nature of the A.I. bias debate, it reveals opportunities deserving of sustained policy attention. The first has to be the wholesale giveaway of societal data that undergirds A.I. system development. We are well overdue for a radical reappraisal over who controls the vast troves of data currently locked down by technology incumbents. Our governors and communities should act decisively to disincentivize and devalue data hoarding with creative policies, including carefully defined bans, levies, mandated data sharing, and community benefit policies, all backed up by the brass knuckles of the law. Smarter data policies would reenergize competition and innovation, both of which have unquestionably slowed with the concentrated market power of the tech giants. The greatest opportunities will flow to those who act most boldly.

The second great opportunity is to wrestle with fundamental existential questions and to build robust processes for resolving them. Which systems really deserve to be built? Which problems most need to be tackled? Who is best placed to build them? And who decides? We need genuine accountability mechanisms, external to companies and accessible to populations. Any A.I. system that is integrated into people’s lives must be capable of contest, account, and redress to citizens and representatives of the public interest. And there must always be the possibility to stop the use of automated systems with appreciable societal costs, just as there is with every other kind of technology.

Artificial intelligence evokes a mythical, objective omnipotence, but it is backed by real-world forces of money, power, and data. In service of these forces, we are being spun potent stories that drive toward widespread reliance on regressive, surveillance-based classification systems that enlist us all in an unprecedented societal experiment from which it is difficult to return. Now, more than ever, we need a robust, bold, imaginative response.