



Workshop Discussion Notes: Algorithmic Accountability

The Social, Cultural & Ethical Dimensions of “Big Data”

March 17, 2014 - New York, NY

<http://www.datasociety.net/initiatives/2014-0317/>

This document was produced based on notes that were taken during the Algorithmic Accountability workshop as part of “The Social, Cultural, and Ethical Dimensions of ‘Big Data’”. Not all attendees were involved in every part of the conversation, nor does this document necessarily reflect the views and beliefs of individual attendees. All workshop participants were given workshop materials prior to the event to spark discussion. This primer can be found at: <http://www.datasociety.net/pubs/2014-0317/AlgorithmicAccountabilityPrimer.pdf>

Overview

This workshop centered on the accountability of algorithmic systems, but one of the key issues of the discussion revolved around how to think about an algorithm, and all that it stands for, particularly in a policy context. An algorithm itself may be a simple decision-tree, or an elaborate one, but the ways in which people are affected by algorithmic systems across different sectors are varied, and difficult to account for outside of obvious instances of social ills or adverse discrimination. Even when the group was able to identify specific instances of the consequences, risks, or benefits of algorithmic determinations, it was challenging to think about how different sectors, like healthcare, education, or law enforcement, use algorithms, and what mechanisms exist to hold them accountable to sector-based sets of criteria.

There was some debate on whether we are really talking about holding institutions accountable, rather than algorithms. Generally, the participants tried to grapple with the challenges of understanding what an algorithm is doing; how to interrogate it or challenge it; how to evaluate what values are implicit in either its operations or its outcomes; whom or what to hold accountable for its operations; and what the difference is between the operation of specialized algorithms and quantitative thinking more generally.

Themes and Discussion Topics

Interrogating a Moving Target

The initial workshop discussion began with several participants weighing in on what it means to hold an algorithm accountable. Some participants mentioned that the need for an audit may seem apparent when there are obvious harms or errors. But, others countered, what about the need for accountability outside obvious errors, such as cases of discrimination? Multiple threads of the discussion tried to weigh the risks and benefits of different types of accountability, and how to incentivize it. One example offered was the business sector, where trust is important to consumers. Is it possible to design algorithmic accountability in a way that actually turns itself into a competitive advantage?

Another part of the discussion turned on the way that “algorithms” might be a word like “technology” that functions as a stand-in for abstract or complicated notions that we don’t have a firm grasp on. Some participants mused that there is a chance that we might be putting too much emphasis on algorithms. After all, they are still operating in the context of specific databases, interfaces, use situations, users, business models, areas of society, etc. By focusing on algorithms only, we might actually be deferring accountability to algorithms and thus letting other actors and entities off the hook. Are we fetishizing algorithms?

Along those lines, the discussion turned on what kinds of implicit values end up in the construction of these algorithms. What happens when the authority of algorithms is challenged, as it happened in the case of Twitter trends? What sort of subtle influences are exerted when a search for “santorum” turns up information that purposefully redefines the name of a famous politician and associates it with another meaning?

Other participants asked if this sort of critical approach would work broadly. What if interrogation and reverse-engineering from the outside do not get us further? Some group members posited the question, shouldn’t we look at the effects of algorithmic systems and how they change relationships? Do we, in fact, have a *right* to interrogate those? Some participants also considered an approach that conceptualizes accountability in layers, such as a design layer, a reverse-engineering layer, and an output-measurement layer. Again, the group struggled with the fact that the technology will likely be a moving target, and thus would be difficult to regulate. There was also some brief discussion about the tensions between regulation on behalf of algorithmic accountability and first amendment concerns.

Mechanisms for Accountability

As an alternative to an interrogative or potential regulatory approach to algorithmic accountability, it was suggested that another approach would be to focus on the manifold practices involved in making an algorithm work in practice. For example, it could be asked of the operators: has there been any scrutiny and thinking about the

implications to begin with? Is there a process inside the company to look at that? Some participants mused about the possibilities for a public ombudsman. It was further suggested that challenging code alone might not be enough, given the complex and not always deterministic ways in which algorithms are developed and operated.

Against this backdrop, the group also discussed to what extent existing accountability mechanism might be able to deal with these challenges. For example, one participant asked how cases of contested algorithmic accountability relate to other cases that have already been accounted for under laws of product liability. Some participants also pointed out that protecting the right to challenge a product or service is important. As an example of such a mechanism, the Consumer Fraud Act (CFA) was mentioned. However, it was also suggested that the CFA “does not so much apply to governments” and that therefore an extended approach might be needed. Another idea concerned a Freedom of Information Act (FOIA) for algorithmic systems, although the Freedom of Information Processing Act (FOIPA) might be the better model.

In addition to transparency, the group also discussed insurance as a way of mitigating harms caused by algorithms. As one participant pointed out, we have mandatory car insurance because it is dangerous to drive one. Why not have the same for algorithms? This would be insurance at the level of the individual user and thus different from insurance for “algorithmic malpractice”, which may or may not already be covered by corporate liability insurance. Some participants highlighted the question of what requirements a successful transparency policy would have to meet. For example, one participant claimed that the A-B-C hygiene-grading scheme for restaurants in New York City was effective because it was so simple. However, other participants doubted whether such an approach could do justice to the complexities of algorithms. As an alternative, it was suggested that transparency was needed not so much about the technical workings of an algorithms but about the claims their creators and operators make about them. Such an approach to algorithmic accountability would focus on accounting for the design, purpose, and aspiration behind the algorithm rather than the actual code. Any use of the algorithms could then be judged against these claims.

A recurring issue in during the discussion was whether a sector-based approach is needed. Specifically, it was highlighted different standards might be appropriate for businesses and government because we trust those institutions differently. One of the emergent tensions from these discussion threads was on the corporatization or privatization of the hypothetical public square, where casual information-sharing practices are subject to algorithmic manipulation, in the sense that information can be rendered visible or invisible based on how it appears in one’s newsfeeds. Some participants offered that this anxiety may be stymied by the seeming dichotomy between the idea that the Internet promises unlimited access to free information,

countered by the point that there are always rules or structures in place for information dissemination.

Why Do Algorithms Make Us Uncomfortable?

The discussion then delved into the idea of what makes people uncomfortable with algorithms, and thus might raise issues of accountability. Some members of the group started with obvious cases such as algorithms producing “incorrect” results and thus causing “harm”, for instance by wrongly identifying a suspect in a policing context. However, the group then quickly turned to what was *perceived* as the even tougher question: how to deal with “icky” outcomes, i.e. outcomes that may be considered “true” or “accurate”, but are still undesirable. For example, if it turned out that people who buy milk in the morning, bring their kids to school, and make a phone call at noon are more likely to commit a crime, how do we feel about that?

One concept that was suggested here was that of a “cultural algorithms”. This is the idea that even in the absence of technology, we have certain assumptions and association about groups of people. For example, sex apps on smartphones often tend to be associated with gay rather than straight men. But what if there is no comfort in accuracy? Can there be a case in which an algorithmic system is “too correct” that we would want to account for?

Another issue that some members of the group discussed concerned the question of how one would measure the output of an algorithm, such as an alleged discriminatory effect. This would require us to develop and apply a set of criteria against which an algorithm and its implications could be judged. It would be tricky, especially in the case of learning algorithms, as some participants suggested. It was also remarked that many of these problems existed already before people talked about algorithms, and that this has more to do with how we address ills that become ingrained in systems without necessarily having any ill intentions. How do we catch problematic areas when systems or technologies are being rolled out? In a different approach, participants also pointed out another issue that might call for accountability, namely, that inefficient algorithms might also do harm by wasting government resources, such as in the computation of taxes or social security benefits.

Are You a Risk, or at Risk?

The next thematic discussion was on specific case studies, such as predictive policing, to focus on some of the pros and cons of algorithmic calculations. It was suggested that predictive policing is a good case for thinking about the limitations of judging algorithms wholesale. Predictive policing may harm people by labeling them *as*

a risk, but at the same time it may also benefit people *at risk*. Predictive policing has been done before the advent of “big data”. For example, police have always had an idea of which cars to stop, or which corners to watch, which neighborhood to patrol. Some people think it’s legitimate, some people don’t. So what has changed? Some participants offered that the tension has to do with community involvement or voices. How do we deal with the discomfort of computing who to survey more closely, rather than relying on the gut practice of local members of the community? What standards could be invoked to justify a claim for or against this specific practice? One related pre-algorithm case to think about could be the practice of [dragnet policing](#). The Supreme Court has developed a four-prong test to evaluate dragnets. How would these four prongs apply to predictive policing?

The discussion then turned more toward journalistic approaches and practices. For example, some participants made comparisons between editorial practice in “traditional” media and in algorithmic systems. The group discussed to what extent we can say that “the New York Times” can be held to account. Some participants suggested that journalists can actually show where their stories come from and how they went about researching them. Such an “audit trail” might be difficult to sustain in the case of algorithms. Another issue concerned the tensions between calls for algorithmic accountability and the ubiquitous trade secret defense. Some participants asked if there is a need to rebalance this. Finally, participants pointed to various websites, which measure media bias in news reporting by diagnosing “liberal”, “conservative”, or “republican” tendencies. Could such a watchdog system work for algorithms, too? Fact-checking websites are hardly used by a broader public, it was argued.

Further Exploration

The main take-away of the workshop discussions is that algorithms are exceedingly difficult to challenge or diagnose. This is partly because finding any singular individual or organization to charge with accountability seems like a highly intimidating task. One of the main unanswerable queries that underlay the discussions centered on the question of, who would one approach to solve problems that become evident in algorithmic systems and information flows? What kind of approach is needed, given that the number of actors operating between users and engineers is multiplying, including data brokers, data center operators, user experience designers, etc.? How do concerns about algorithmic accountability reflect greater social ills, rather than a unique problem with specific technologies? How can we correct for flaws in algorithmic outcomes in a way that is systematic, and/or domain-specific? Are there ways to catch problems before they become ingrained in systems or technologies that are deployed?