

The Algorithms Beat

Nicholas Diakopoulos

Northwestern University, School of Communication

The Machine Bias series from ProPublica began in May 2016 as an effort to investigate algorithms in society¹. Perhaps most striking in the series was an investigation and analysis exposing the racial bias of recidivism risk assessment algorithms used in criminal justice decisions². These algorithms score individuals based on whether they are a low or high risk of reoffending. States and other municipalities variously use the scores for managing pre-trial detention, probation, parole, and sometimes even sentencing. Reporters at ProPublica filed a public records request for the scores from Broward County in Florida and then matched those scores to criminal histories to see whether an individual had actually recidivated (i.e. reoffended) within two years. Analysis of the data showed that black defendants tended to be assigned higher risk scores than white defendants, and were more likely to be incorrectly labeled as high risk when in fact after two years they hadn't actually been rearrested³.

Scoring in the criminal justice system is of course just one domain where algorithms are being deployed in society. The Machine Bias series has since covered everything from Facebook's ad targeting system, to geographically discriminatory auto insurance rates, and unfair pricing practices on Amazon.com. Algorithmic decision making is increasingly pervasive throughout both the public and private sectors. We see it in domains like credit and insurance risk scoring, employment systems, welfare management, educational and teacher rankings, and online media curation, among many others⁴. Operating at scale and often affecting large swaths of people, algorithms can make consequential and sometimes contestable calculation, ranking, classification, association, and filtering decisions. Algorithms, animated by piles of data, are a potent new way of wielding power in society.

As ProPublica's Machine Bias series attests, a new strand of computational and data journalism is emerging to investigate and hold accountable how power is exerted through algorithms. I call this *algorithmic accountability reporting*, a re-orientation of the traditional watchdog function of journalism towards the power wielded through algorithms⁵. Despite their ostensible objectivity,

¹ <https://www.propublica.org/series/machine-bias>

² Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner. Machine Bias. ProPublica. May, 2016.

<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

³ Jeff Larson, Surya Mattu, Lauren Kirchner and Julia Angwin. How We Analyzed the COMPAS Recidivism Algorithm. ProPublica. May, 2016. <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm/>

⁴ Cathy O'Neil. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Broadway Books. 2016; Frank Pasquale. The Black Box Society: The Secret Algorithms That Control Money and Information. Harvard University Press. 2015; Virginia Eubanks. Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. St. Martin's Press. New York. 2018.

⁵ N. Diakopoulos. Algorithmic Accountability: Journalistic Investigation of Computational Power Structures. 3 (3), *Digital Journalism*. 2015; The term "Algorithmic Accountability" was originally coined in: Nicholas Diakopoulos. Sex, Violence, and Autocomplete Algorithms. Slate. August, 2013.

http://www.slate.com/articles/technology/future_tense/2013/08/words_banned_from_bing_and_google_s_autocomp

algorithms can and do make mistakes and embed biases that warrant closer scrutiny. Slowly, a beat on algorithms is coalescing as journalistic skills come together with technical skills to provide the scrutiny that algorithms deserve.

In deciding what constitutes the beat, it's first helpful to define what's newsworthy about algorithms. An algorithm is an ordered set of steps followed in order to solve a particular problem or to accomplish a defined outcome—they make decisions. The crux of algorithmic power often boils down to computers' ability to make algorithmic decisions very quickly and at scale, potentially affecting large numbers of people. What makes an algorithm newsworthy is when it somehow makes a “bad” decision. This might involve an algorithm doing something it wasn't supposed to do, or perhaps not doing something it *was* supposed to do. For journalism, the public significance and consequences of a bad decision are key factors. What's the potential harm for an individual, or for society? Bad decisions might impact individuals directly, or in aggregate lead to issues like bias. Bad decisions can also be costly. Let's look at how various bad decisions can lead to news stories.

Angles on Algorithms

In my research I've identified four different driving forces for algorithmic accountability stories: discrimination and unfairness, errors or mistakes in predictions or classifications, legal or social norm violations, and misuse of algorithms by people either intentionally or inadvertently. Examples of specific algorithmic accountability stories will be illustrative here.

Discrimination and Unfairness

Uncovering discrimination and unfairness is a common theme in algorithmic accountability reporting. The story from ProPublica that led this chapter is a striking example of how an algorithm can lead to systematic disparities in the treatment of different groups of people. Northpoint, the company that designed the risk assessment scores, argued the scores were equally accurate across races and were therefore fair. But their definition of fairness failed to take into account the disproportionate volume of mistakes that affected black people. Stories of discrimination and unfairness hinge on the definition of fairness applied, which may reflect different political suppositions⁶.

I have also worked on stories that uncover unfairness due to algorithmic systems—in particular looking at how Uber pricing dynamics may differentially affect neighborhoods in Washington, DC⁷. Based on initial observations of different waiting times and how those waiting times shifted based on Uber's surge pricing algorithm we hypothesized that different neighborhoods would have different levels of service quality (i.e. waiting time). By systematically sampling the

[lete_algorithms.html](#) and elaborated in: Nicholas Diakopoulos. Rage Against the Algorithms. The Atlantic. October, 2013. <https://www.theatlantic.com/technology/archive/2013/10/rage-against-the-algorithms/280255/>

⁶ Lepri, B. et al., 2017. Fair, Transparent, and Accountable Algorithmic Decision-making Processes. *Philosophy & Technology*, 84(3), pp.1–17.

⁷ J. Stark and N. Diakopoulos. Uber seems to offer better service in areas with more white people. That raises some tough questions. *Washington Post*. March, 2016. <https://www.washingtonpost.com/news/wonk/wp/2016/03/10/uber-seems-to-offer-better-service-in-areas-with-more-white-people-that-raises-some-tough-questions/>

waiting times in different census tracts over time we showed that census tracts with more people of color tend to have longer wait times for a car, even when controlling for other factors like income, poverty rate, and population density in the neighborhood. It's difficult to pin the unfair outcome directly to Uber's algorithm because other human factors also drive the system, such as the behavior and potential biases of Uber drivers. But the results do suggest that when considered as a whole, the system exhibits disparity associated with demographics.

Errors and Mistakes

Algorithms can also be newsworthy when they make specific errors or mistakes in their classification, prediction, or filtering decisions. Consider the case of platforms like Facebook and Google which use algorithmic filters to reduce exposure to harmful content like hate speech, violence, and pornography. This can be important for the protection of specific vulnerable populations, like children, especially in products like Google's YouTube Kids which are explicitly marketed as safe for children. Errors in the filtering algorithm for the app are newsworthy because they mean that sometimes children encounter inappropriate or violent content⁸. Classically, algorithms make two types of mistakes: false positives and false negatives. In the YouTube Kids scenario, a false positive would be a video mistakenly classified as inappropriate when actually it's totally fine for kids. A false negative is a video classified as appropriate when it's really not something you want kids watching.

Classification decisions impact individuals when they either increase or decrease the positive or negative treatment an individual receives. When an algorithm mistakenly selects an individual to receive free ice cream (increased positive treatment), you won't hear that individual complain (although when others find out, they might say it's unfair). Errors are generally newsworthy when they lead to increased negative treatment for a person, such as by exposing a child to an inappropriate video. Errors are also newsworthy when they lead to a decrease in positive treatment for an individual, such as when a person misses an opportunity. Just imagine a qualified buyer who never gets a special offer because an algorithm mistakenly excludes them. Finally, errors can be newsworthy when they cause a decrease in warranted negative attention. Consider a criminal risk assessment algorithm mistakenly labeling a high-risk individual as low-risk—a false negative. While that's great for the individual, this creates a greater risk to public safety by letting free an individual who goes on to commit a crime again.

Legal and Social Norm Violations

Predictive algorithms can sometimes test the boundaries of established legal or social norms, leading to other opportunities and angles for coverage. Consider for a moment the possibility of algorithmic defamation⁹. Defamation is defined as "a false statement of fact that exposes a person to hatred, ridicule or contempt, lowers him in the esteem of his peers, causes him to be shunned, or injures him in his business or trade"¹⁰. Over the last several years there have been

⁸ Sapna Maheshwari. On YouTube Kids, Startling Videos Slip Past Filters. New York Times. November, 2017. https://www.nytimes.com/2017/11/04/business/media/youtube-kids-paw-patrol.html?_r=0

⁹ Nicholas Diakopoulos. Algorithmic Defamation: The Case of the Shameless Autocomplete. Tow Center. August, 2013. <https://towcenter.org/algorithmic-defamation-the-case-of-the-shameless-autocomplete/>

¹⁰ Defamation. <http://www.dmlp.org/legal-guide/defamation>

numerous stories, and legal battles, over individuals who feel they've been defamed by Google's autocomplete algorithm. An autocompletion can link an individual's or company's name to everything from crime and fraud to bankruptcy or sexual conduct, which can then have consequences on reputation.

Algorithms can also be newsworthy when they encroach on social norms like privacy. For instance, Gizmodo has been covering the "People You May Know" (PYMK) algorithm on Facebook, which suggests potential "friends" on the platform that are sometimes inappropriate or undesired¹¹. In one story, reporters identified a case where PYMK outed the real identity of a sex worker to her clients¹². This is problematic not only because of the potential stigma attached to sex work, but also out of fear of clients who could become stalkers.

Defamation and privacy violations are only two possible story angles here. Journalists should be on the lookout for a range of other legal or social norm violations that algorithms may create in various social contexts.

Human Misuse

Algorithmic decisions are often embedded in larger decision-making processes that involve people and algorithms, so-called sociotechnical systems¹³. If algorithms are misused by the people in the sociotechnical ensemble this may also be newsworthy. The designers of algorithms can sometimes anticipate and articulate guidelines for a reasonable set of use contexts for a system, and so if people ignore these in practice it can lead to a story of negligence or misuse. The risk assessment story from ProPublica provides a salient example. Northpointe had in fact created two versions and calibrations of the tool, one for men and one for women. Statistical models need to be trained on data reflective of the population where they will be used and gender is an important factor in recidivism prediction. Broward County was misusing the risk score designed and calibrated for men by using it for women as well¹⁴.

How to Investigate an Algorithm

There are various routes to the investigation of algorithmic power. Some stories may require methods that build on social science audit techniques, while other threads can be exposed through poking and prodding of algorithmic reactions. Traditional journalistic sourcing to talk to company insiders such as designers, developers, and data scientists, as well as to file public records requests and find impacted individuals are all as important as ever. There are more

¹¹ Kashmir Hill. How Facebook Figures Out Everyone You've Ever Met. Gizmodo. November, 2017. <https://gizmodo.com/how-facebook-figures-out-everyone-youve-ever-met-1819822691>

¹² Kashmir Hill. How Facebook Outs Sex Workers. Gizmodo. October, 2017. <https://gizmodo.com/how-facebook-outs-sex-workers-1818861596>

¹³ Seaver, N., 2017. Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society*, 4(2);

¹⁴ Machine Bias with Jeff Larson. Data Stories Podcast. October, 2016. <http://datastori.es/85-machine-bias-with-jeff-larson/>

methods than I can possibly cover in this short chapter¹⁵, but I want to at least talk briefly about how journalists can use auditing to investigate algorithms.

Auditing techniques have been used for decades to study social bias in systems like housing markets¹⁶. Algorithms can be studied with similar methods¹⁷. The premise is that if the inputs to algorithms are varied in enough different ways, and the outputs are monitored, then inputs and outputs can be correlated to build a theory for how the algorithm may be functioning. If we have some expected outcome that the algorithm violates for a given input this can help tabulate errors and see if errors are biased in systematic ways. For personalized algorithms, auditing techniques have been married to crowdsourcing in order to gather data from a range of people who may each have a unique “view” of the algorithm. Algorithm Watch in Germany has used this technique effectively to study the personalization of Google Search results, collecting almost 6 million search results from more than 4,000 users who shared data via a browser plugin¹⁸. Gizmodo has used a variant of this technique to help investigate Facebook’s PYMK. Users download a piece of software to their computer that periodically tracks PYMK results locally to the user’s computer, maintaining their privacy. Reporters then solicit tips from users who think their results are worrisome or surprising¹⁹.

Auditing algorithms is not for the faint of heart. Information deficits limit an auditor’s ability to sometimes even know where to start, what to ask for, how to interpret results, and how to explain the patterns they’re seeing in an algorithm’s behavior. There is also the challenge of knowing and defining what’s expected of an algorithm, and how those expectations may vary across contexts. In order to identify a newsworthy mistake or bias you must first define what normal or unbiased should look like. The issue of legal access to information about algorithms in use in government also crops up. In the U.S., Freedom of Information (FOI) laws govern the public’s access to documents, but the response from government agencies for documents relating to algorithms is uneven at best²⁰. Legal reforms may be in order so that public access to information about algorithms is more easily facilitated. If information deficits, difficult to articulate expectations, and uncertain legal access aren’t challenging enough, just remember that

¹⁵ For more a more complete treatment of methodological options see: N. Diakopoulos. *Automating the News: How Algorithms are Rewriting the Media*. Harvard University Press. 2019; see also: N. Diakopoulos. *Enabling Accountability of Algorithmic Media: Transparency as a Constructive and Critical Lens*. *Towards glass-box data mining for Big and Small Data*. Eds. Tania Cerquitelli, Daniele Quercia, Frank Pasquale. Springer. June, 2017.

¹⁶ Gaddis, S.M., 2017. *An Introduction to Audit Studies in the Social Sciences*. In S. M. Gaddis, ed. *Audit Studies Behind the Scenes with Theory, Method, and Nuance*.

¹⁷ Sandvig, C. et al., 2014. *Auditing algorithms: Research methods for detecting discrimination on Internet platforms*. In Presented at International Communication Association preconference on Data and Discrimination *Converting Critical Concerns into Productive Inquiry*.

¹⁸ See: <https://datenspende.algorithmwatch.org/en/index.html>

¹⁹ Kashmir Hill and Surya Mattu. *Keep Track Of Who Facebook Thinks You Know With This Nifty Tool*. Gizmodo. January, 2018. <https://gizmodo.com/keep-track-of-who-facebook-thinks-you-know-with-this-ni-1819422352>

²⁰ See for instance: Nicholas Diakopoulos. *We need to know the algorithms the government uses to make important decisions about us*. *The Conversation*. May, 2016. <https://theconversation.com/we-need-to-know-the-algorithms-the-government-uses-to-make-important-decisions-about-us-57869>; Fink, K., 2017. *Opening the government’s black boxes: freedom of information and algorithmic accountability*. 17(1), pp.1–19; and Brauneis, R. & Goodman, E.P., 20 *Yale Journal of Law & Technology*. 103. 2018.

algorithms can also be quite capricious. Today's version of the algorithm may already be different than yesterday's. Algorithms must be monitored over time in order to understand how they are changing and evolving.

Recommendations Moving Forward

To get started and make the most of algorithmic accountability reporting I would recommend three things. Firstly, we've developed a resource called Algorithm Tips (<http://algorithmtips.org/>), which provides not only pointers to useful methods and further examples, but also provides an updated database of algorithms in use in the U.S. Federal government that may be worthy of investigation. If you're looking for a resource to help get something off the ground, that could be your starting point²¹. Secondly, focus on the outcomes and impacts of algorithms rather than trying to explain the exact mechanism for their decision making. Identifying algorithmic discrimination (i.e., an output) oftentimes has more value to society as an initial step than explaining exactly how that discrimination came about. By focusing on outcomes, journalists can provide a first-order diagnostic and signal an alarm which other stakeholders can then dig into. Finally, much of the published algorithmic accountability reporting I've cited here is done in teams, and with good reason. Effective algorithmic accountability reporting demands all of the traditional skills journalists need in reporting and interviewing, domain knowledge of a beat, public records requests, and writing results clearly and compellingly, plus a host of new capabilities like scraping and cleaning data, designing audit studies, and using advanced statistical techniques. Expertise in these different areas can be distributed among a team, as long as there is clear communication, awareness and leadership. Methods specialists can partner with different domain experts to understand algorithmic power across a larger variety of social domains.

Abstract

Operating at scale and often affecting large groups of people, algorithms make consequential and sometimes contestable decisions in an increasing range of domains throughout the public and private sectors. A distinct beat in journalism is emerging to encompass the investigation of societal power exerted through such algorithms. This chapter examines this beat and offers four newsworthy angles for computational and data journalists looking to cover algorithms: discrimination and unfairness, errors and mistakes, social and legal norm violations, and human misuse. These are each illustrated with examples. Methodological options and challenges for investigating algorithms are then outlined, focusing on audit techniques. Finally, several concrete recommendations are offered to help the fledgling data journalist get started thinking about their own investigations on the algorithms beat.

Keywords

Algorithmic accountability, computational journalism, investigative journalism, auditing

²¹ D. Trielli, J. Stark and N. Diakopoulos. Algorithm Tips: A Resource for Algorithmic Accountability in Government. October, 2017.