31. The Algorithms Beat: Angles and Methods for Investigation

Nicholas Diakopoulos

Abstract

A beat on algorithms is coalescing as journalistic skills come together with technical skills to provide the scrutiny that algorithms deserve.

Keywords: algorithms, algorithmic accountability, computational journalism, investigative journalism, algorithm studies, freedom of information (FOI)

The “Machine Bias” series from *ProPublica* began in May 2016 as an effort to investigate algorithms in society.¹ Perhaps most striking in the series was an investigation and analysis exposing the racial bias of recidivism risk assessment algorithms used in criminal justice decisions (Angwin et al., 2016). 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 pretrial 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 actual 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 labelled as high risk when in fact after two years they hadn’t actually been rearrested (Larson et al., 2016).

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

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

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 others (Eubanks, 2018; O'Neil, 2016; Pasquale, 2015). Operating at scale and often impacting 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 algorithms (Diakopoulos, 2015).² Despite their ostensible objectivity, 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.

There are, of course, a variety of forms of algorithmic accountability that may take place in diverse forums beyond journalism, such as in political, legal, academic, activist or artistic contexts (Brain & Mattu, n.d.; Bucher, 2018).³ But my focus in this chapter is squarely on algorithmic accountability reporting as an independent journalistic endeavour that contributes to accountability by mobilizing public pressure. This can be seen as complementary to other avenues that may ultimately also contribute to accountability, such as by developing regulations and legal standards, creating audit institutions in civil society, elaborating effective transparency policies, exhibiting reflexive art shows, and publishing academic critiques.

2 The term *algorithmic accountability* was originally coined in: Diakopoulos, N. (2013, August 2). Sex, violence, and autocomplete algorithms. *Slate Magazine*. <https://slate.com/technology/2013/08/words-banned-from-bing-and-googles-autocomplete-algorithms.html>; and elaborated in: Diakopoulos, N. (2013, October 3). Rage against the algorithms. *The Atlantic*. <https://www.theatlantic.com/technology/archive/2013/10/rage-against-the-algorithms/280255/>

3 For an activist/artistic frame, see: Brain, T., & Mattu, S. (n.d.). *Algorithmic disobedience*. <https://samatt.github.io/algorithmic-disobedience/#/>. For an academic treatment examining algorithmic power, see: Bucher, T. (2018). *If . . . then: Algorithmic power and politics*. Oxford University Press. A broader selection of the academic scholarship on critical algorithm studies can be found here: <https://socialmediacollective.org/reading-lists/critical-algorithm-studies/>

In deciding what constitutes the beat in journalism, it is first helpful to define what is newsworthy about algorithms. Technically speaking, an algorithm is a sequence of steps followed in order to solve a particular problem or to accomplish a defined outcome. In terms of information processes, the outcomes of algorithms are typically decisions. The crux of algorithmic power often boils down to computers' ability to make such decisions very quickly and at scale, potentially affecting large numbers of people. In practice, algorithmic accountability is not just about the technical side of algorithms, however—algorithms should be understood as composites of technology woven together with people such as designers, operators, owners and maintainers in complex sociotechnical systems (Ananny, 2015; Seaver, 2017). Algorithmic accountability is about understanding how those people exercise power within and through the system, and are ultimately responsible for the system's decisions. Oftentimes what makes an algorithm newsworthy is when it somehow makes a “bad” decision. This might involve an algorithm doing something it was not 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 is the potential harm for an individual, or for society? Bad decisions might impact individuals directly, or in aggregate may reinforce issues like structural bias. Bad decisions can also be costly. Let's look at how various bad decisions can lead to news stories.

Angles on Algorithms

In observing the algorithms beat developed over the last several years in journalism, as well as through my own investigations of algorithms, I have identified at least four driving forces that appear to underlie many algorithmic accountability stories: (a) discrimination and unfairness, (b) errors or mistakes in predictions or classifications, (c) legal or social norm violations, and (d) misuse of algorithms by people either intentionally or inadvertently. I provide illustrative examples of each of these in the following subsections.

Discrimination and Unfairness. Uncovering discrimination and unfairness is a common theme in algorithmic accountability reporting. The story from *ProPublica* that opened this chapter is a striking example of how an algorithm can lead to systematic disparities in the treatment of different groups of people. Northpointe, the company that designed the risk assessment scores (since renamed Equivant), 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 suppositions (Lepri et al., 2018).

I have also worked on stories that uncover unfairness due to algorithmic systems—in particular looking at how Uber pricing dynamics may differentially affect neighbourhoods in Washington, DC (Stark & Diakopoulos, 2016). 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 neighbourhoods would have different levels of service quality (i.e., waiting time). By systematically sampling the waiting times in different census tracts over time, we showed that census tracts with more people of colour tend to have longer wait times for a car, even when controlling for other factors like income, poverty rate and population density in the neighbourhood. It is difficult to pin the unfair outcome directly to Uber's technical algorithm because other human factors also drive the system, such as the behaviour 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 (such as 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 content (Maheshwari, 2017). 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 is 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 labelling 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 setting free an individual who might go 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 defamation (Diakopoulos, 2013; Lewis et al., 2019). 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.”⁴ Over the last several years there have been numerous stories, and legal battles, over individuals who feel they have been defamed by Google's autocomplete algorithm. An autocomplete can link an individual's or a company's name to everything from crime and fraud to bankruptcy or sexual conduct, which can then have consequences for reputation. Algorithms can also be newsworthy when they encroach on social norms like privacy. For instance, *Gizmodo* has extensively covered the “People You May Know” (PYMK) algorithm on Facebook, which suggests potential “friends” on the platform that are sometimes inappropriate or undesired (Hill, 2017b). In one story, reporters identified a case where PYMK outed the real identity of a sex worker to her clients (Hill, 2017a). 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. Since algorithms necessarily rely on a quantified version of reality that only incorporates what is measurable as data they can miss a lot of the social and legal context that would otherwise be essential in rendering an accurate decision. By understanding what a particular algorithm actually quantifies about the world—how it “sees” things—journalists can inform

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

critique by illuminating the missing bits that would support a decision in the richness of its full context.

Human Misuse. Algorithmic decisions are often embedded in larger decision-making processes that involve a constellation of people and algorithms woven together in a sociotechnical system. Despite the inaccessibility of some of their sensitive technical components, the sociotechnical nature of algorithms opens up new opportunities for investigating the relationships that users, designers, owners and other stakeholders may have to the overall system (Trielli & Diakopoulos, 2017). 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. But Broward County was misusing the risk score designed and calibrated for men by using it for women as well (Larson, 2016).

How to Investigate an Algorithm

There are various routes to the investigation of algorithmic power and no single approach will always be appropriate. But there is a growing stable of methods to choose from, including everything from highly technical reverse engineering and code-inspection techniques, to auditing using automated or crowdsourced data collection, or even low-tech approaches to prod and critique based on algorithmic reactions (Diakopoulos, 2017, 2019).⁵ Each story may require a different approach depending on the angle and the specific context, including what degree of access to the algorithm, its data and its code is available. For instance, an exposé on systematic discrimination may lean heavily on an audit method using data collected online, whereas a code review may be necessary to verify the correct implementation of

5 For more a more complete treatment of methodological options, see: Diakopoulos, N. (2019). *Automating the news: How algorithms are rewriting the media*. Harvard University Press; see also: Diakopoulos, N. (2017). Enabling accountability of algorithmic media: Transparency as a constructive and critical lens. In T. Cerquitelli, D. Quercia, & F. Pasquale (Eds.), *Transparent data mining for big and small data* (pp. 25–43). Springer International Publishing. https://doi.org/10.1007/978-3-319-54024-5_2

an intended policy (Lecher, 2018). 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 as important as ever. I can't go into depth on all of these methods in this short chapter, but here I want to at least elaborate a bit more on how journalists can investigate algorithms using auditing.

Auditing techniques have been used for decades to study social bias in systems like housing markets and have recently been adapted for studying algorithms (Gaddis, 2017; Sandvig et al., 2014). The basic idea 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 (Diakopoulos, 2015). 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. When algorithms can be accessed via APIs or online web pages output data can be collected automatically (Valentino-DeVries et al., 2012). For personalized algorithms, auditing techniques have also been married to crowdsourcing in order to gather data from a range of people who may each have a unique "view" of the algorithm. AlgorithmWatch 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 plug-in (as discussed further by Christina Elmer in her chapter in this book).⁶ *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 can then solicit tips from users who think their results are worrisome or surprising (Hill & Mattu, 2018).

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 are seeing in an algorithm's behaviour. There is also the challenge of knowing and defining what is expected of an algorithm, and how those expectations may vary across contexts and according to different global moral, social, cultural and legal standards and norms. For instance, different expectations for fairness may come into play for a criminal risk assessment algorithm in comparison to an algorithm that charges people different prices for an

6 <https://algorithmwatch.org/filterblase-geplatzt-kaum-raum-fuer-personalisierung-bei-google-suchen-zur-bundestagswahl-2017/> (German language)

airline seat. In order to identify a newsworthy mistake or bias you must first define what normal or unbiased should look like. Sometimes that definition comes from a data-driven baseline, such as in our audits of news sources in Google search results during the 2016 US elections (Diakopoulos et al., 2018). The issue of legal access to information about algorithms also crops up and is, of course, heavily contingent on the jurisdiction (Bhandari & Goodman, 2017). In the United States, freedom of information (FOI) laws govern the public's access to government documents, but the response from different agencies for documents relating to algorithms is uneven at best (see Brauneis & Goodman, 2018; Diakopoulos, 2016; Fink, 2017). Legal reforms may be in order so that public access to information about algorithms is more easily facilitated. And if information deficits, difficult-to-articulate expectations and uncertain legal access are not challenging enough, just remember that algorithms can also be quite capricious. Today's version of the algorithm may already be different than yesterday's: As one example, Google typically changes its search algorithm 500–600 times a year. Depending on the stakes of the potential changes, algorithms may need to 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 have developed a resource called Algorithm Tips, which curates relevant methods, examples and educational resources, and hosts a database of algorithms for potential investigation (first covering algorithms in the US federal government and then expanded to cover more jurisdictions globally).⁷ If you are looking for resources to learn more and help to get a project off the ground, that could be one starting point (Trielli et al., 2017). Secondly, focus on the outcomes and impacts of algorithms rather than trying to explain the exact mechanism of 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 in other accountability forums. Finally, much of the published algorithmic accountability reporting I have cited here is done in teams, and with

7 <http://algorithmtips.org/>

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 analysis of the returned documents, and writing results clearly and compellingly, while often also relying on 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, or with external collaborators, as long as there is clear communication, awareness and leadership. In this way, methods specialists can partner with different domain experts to understand algorithmic power across a larger variety of social domains.

Works Cited

- Ananny, M. (2015). Toward an ethics of algorithms. *Science, Technology & Human Values*, 41(1), 93–117.
- Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). Machine bias. *ProPublica*. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- Bhandari, E., & Goodman, R. (2017). Data journalism and the computer fraud and abuse act: Tips for moving forward in an uncertain landscape. *Computation+Journalism Symposium*. <https://www.aclu.org/other/data-journalism-and-computer-fraud-and-abuse-act-tips-moving-forward-uncertain-landscape>
- Brain, T., & Mattu, S. (n.d.). Algorithmic disobedience. <https://samatt.github.io/algorithmic-disobedience/>
- Brauneis, R., & Goodman, E. P. (2018). Algorithmic transparency for the smart city. *Yale Journal of Law & Technology*, 20, 103–176.
- Bucher, T. (2018). *If... then: Algorithmic power and politics*. Oxford University Press.
- Diakopoulos, N. (2013, August 6). Algorithmic defamation: The case of the shameless autocomplete. Tow Center for Journalism. <https://towcenter.org/algorithmic-defamation-the-case-of-the-shameless-autocomplete>
- Diakopoulos, N. (2015). Algorithmic accountability: Journalistic investigation of computational power structures. *Digital Journalism*, 3(3), 398–415. <https://doi.org/10.1080/21670811.2014.976411>
- Diakopoulos, N. (2016, May 24). We need to know the algorithms the government uses to make important decisions about us. *The Conversation*. <http://theconversation.com/we-need-to-know-the-algorithms-the-government-uses-to-make-important-decisions-about-us-57869>

- Diakopoulos, N. (2017) Enabling Accountability of Algorithmic Media: Transparency as a Constructive and Critical Lens. In T. Cerquitelli, D. Quercia, & F. Pasquale (Eds.), *Transparent data mining for Big and Small Data* (pp. 25–44). Springer.
- Diakopoulos, N. (2019). *Automating the News: How Algorithms are Rewriting the Media*. Harvard University Press.
- Diakopoulos, N., Trielli, D., Stark, J., & Mussenden, S. (2018). I vote for—How search informs our choice of candidate. In M. Moore & D. Tambini (Eds.), *Digital Dominance: The power of Google, Amazon, Facebook, and Apple* (pp. 320–341). Oxford University Press. https://www.academia.edu/37432634/I_Vote_For_How_Search_Informs_Our_Choice_of_Candidate
- Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.
- Fink, K. (2017). Opening the government's black boxes: Freedom of information and algorithmic accountability. *Digital Journalism*, 17(1). <https://doi.org/10.1080/1369118X.2017.1330418>
- Gaddis, S. M. (2017). An introduction to audit studies in the social sciences. In M. Gaddis (Ed.), *Audit studies: Behind the scenes with theory, method, and nuance* (pp. 3–44). Springer International Publishing.
- Gillespie, T., & Seaver, N. (2015, November 5). Critical algorithm studies: A reading list. *Social Media Collective*. <https://socialmediacollective.org/reading-lists/critical-algorithm-studies/>
- Hill, K. (2017a, October). How Facebook outs sex workers. *Gizmodo*. <https://gizmodo.com/how-facebook-outs-sex-workers-1818861596>
- Hill, K. (2017b, November). How Facebook figures out everyone you've ever met. *Gizmodo*. <https://gizmodo.com/how-facebook-figures-out-everyone-youve-ever-met-1819822691>
- Hill, K., & Mattu, S. (2018, January 10). Keep track of who Facebook thinks you know with this nifty tool. *Gizmodo*. <https://gizmodo.com/keep-track-of-who-facebook-thinks-you-know-with-this-ni-1819422352>
- Larson, J. (2016, October 20). Machine bias with Jeff Larson [Data Stories podcast]. <https://datastori.es/85-machine-bias-with-jeff-larson/>
- Larson, J., Mattu, S., Kirchner, L., & Angwin, J. (2016, May 23). How we analyzed the COMPAS recidivism algorithm. *ProPublica*. <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>
- Lecher, C. (2018, March 21). What happens when an algorithm cuts your health care. *The Verge*. <https://www.theverge.com/2018/3/21/17144260/healthcare-medicaid-algorithm-arkansas-cerebral-palsy>
- Lepri, B., Oliver, N., Letouzé, E., Pentland, A., & Vinck, P. (2018). Fair, transparent, and accountable algorithmic decision-making processes. *Philosophy & Technology*, 31(4), 611–627. <https://doi.org/10.1007/s13347-017-0279-x>

- Lewis, S. C., Sanders, A. K., & Carmody, C. (2019). Libel by algorithm? Automated journalism and the threat of legal liability. *Journalism and Mass Communication Quarterly*, 96(1), 60–81. <https://doi.org/10.1177/1077699018755983>
- Maheshwari, S. (2017, November 4). On Youtube Kids, startling videos slip past filters. *The New York Times*. https://www.nytimes.com/2017/11/04/business/media/youtube-kids-paw-patrol.html?_r=0
- O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway Books.
- Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.
- Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014, May 22). Auditing algorithms: Research methods for detecting discrimination on Internet platforms. International Communication Association preconference on Data and Discrimination Converting Critical Concerns into Productive Inquiry, Seattle, WA.
- Seaver, N. (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society*, 4(2). <https://doi.org/10.1177/2053951717738104>
- Stark, J., & Diakopoulos, N. (2016, March 10). Uber seems to offer better service in areas with more White people. That raises some tough questions. *The Washington Post*. <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/>
- Trielli, D., & Diakopoulos, N. (2017, May 30). How to report on algorithms even if you're not a data whiz. *Columbia Journalism Review*. https://www.cjr.org/tow_center/algorithms-reporting-algorithmtips.php
- Trielli, D., Stark, J., & Diakopoulos, N. (2017). *Algorithm tips: A resource for algorithmic accountability in Government*. Computation + Journalism Symposium.
- Valentino-DeVries, J., Singer-Vine, J., & Soltani, A. (2012, December 24). Websites vary prices, deals based on users' information. *The Wall Street Journal*. <https://www.wsj.com/articles/SB10001424127887323777204578189391813881534>

About the Author

Nicholas Diakopoulos is an Assistant Professor at Northwestern University School of Communication, where he directs the Computational Journalism Lab. He is the author of *Automating the News: How Algorithms are Rewriting the Media* (Harvard University Press, 2019).