How Journalists Can Systematically Critique Algorithms

Daniel Trielli
dtrielli@u.northwestern.edu
Northwestern University

Nicholas Diakopoulos
nad@northwestern.edu
Northwestern University

ABSTRACT

As government and big tech algorithms become more pervasive and powerful, journalists have a role to keep those automated systems in check. In this paper, we will describe how algorithms can be systematically critiqued and covered by journalists. We outline how algorithms can be analyzed and covered even when they are opaque, and how journalists can do so even when they have no strong technical tools. The variety of threats that arise from algorithms can be broken down into the specific components of the algorithm itself. In essence, a systematic view that takes into account different types of possible harms created by different components of algorithms (i.e. design, input, calculation, output) allows for a method of critique that is attainable even by journalists that are less technically proficient.

ACM Reference Format:
Daniel Trielli and Nicholas Diakopoulos. 2019. How Journalists Can Systematically Critique Algorithms. In *Computation + Journalism Symposium 2020*. ACM, New York, NY, USA, 5 pages.

1 INTRODUCTION

Algorithms are becoming more pervasive, powerful, and varied. Whether they are used and deployed by private companies or by governments, they are affecting an increasingly large portion of society, and at many times, with unintended and harmful consequences [3, 18].

Since the role of journalists is to cover the issues that are impacting society, algorithms should then be part of the beat of news organizations [5, 6]. But journalists are not always technically skilled enough to cover complex computational systems. Furthermore, the variety of algorithms, their applications, and manifestations, might make it challenging for journalists to begin to grasp how to cover them.

Previous work on algorithmic accountability reporting has been valuable to diminish that knowledge gap [5, 7, 15, 19, 24]. Using examples and listing attributes of algorithms to guide journalists in their investigations, these contributions assist journalists who are interested in covering automated decision systems.

C+J '20, March 20–21, 2020, Boston, MA
.

In this paper, we hope to advance and deepen this work by proposing a systematic approach that is applicable to algorithms in general. Our contribution is a framework that takes into consideration each element of the algorithm, and which lays out a systematic approach that journalists can take when critiquing an algorithmic system. We delineate this approach in the following sections. First we untangle the variety of threats that arise from algorithms, in each of its components (input, calculation, and output). Then we explore scenarios of varied access to information for each of those components, and how journalists and researchers can tackle those gaps of information. These conceptual frameworks outline how algorithms can be critiqued, analyzed, and covered even when they are opaque or when journalists lack access to complex technical methods. We conclude with a discussion of how different scenarios of access to data elicit approaches that range from the technical to the conceptual.

2 ALGORITHMIC COMPONENTS AND THEIR POTENTIAL HARMS

In order to fully explore the threats arising from algorithms, we first break down each of its components: input, calculation, and output. For that, we use a standard technical representation of an algorithm, as shown on Figure 1. That representation (and the focus of this paper) covers the internal technical system that is part of larger algorithmic assemblages [26] that involve social actors as well. However, as we will see, understanding that larger context is also important for journalists who are investigating the internal process of input, calculation, and output.

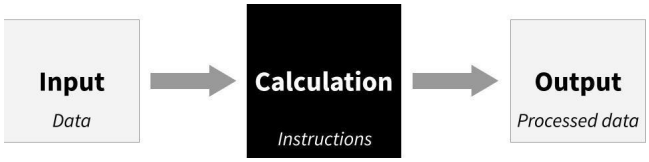

Figure 1: A representation of an algorithm

What makes this division useful is that, as we will see, each one of these components have specific threats and implications for society, and they manifest in different ways. Therefore, journalists who are investigating algorithms might ask specific questions in each step. We explain the rationale for

these questions in the next subsections, but their summary is in Table 1.

Table 1: Questions at each component of the algorithm

Component	Questions
Input	Is the algorithm using the appropriate data for analysis? Is it avoiding data that it should not be using?
Calculation	Is the algorithm making the expected calculation with the data? What are the threats of false positives and false negatives?
Output	Are the data or decisions generated by the algorithm useful and understandable by its operators?
Design	Is the application of the algorithm valid?

The input

One of the key aspects of the development of algorithms is determining what kind of data will be used as their raw material. No matter how correct the calculations are, if they are done with inappropriate data, they will lead to faulty conclusions and/or bias. An input can be inappropriate if it does not measure what the algorithm is aiming to predict or classify. But it can also improperly use data that is already biased or that reinforces biases in society.

One example of that is the fact that algorithms can discriminate based on race even when they do not specifically use race data. In countries that have racially segregated neighborhoods, such as the United States, postal codes can be strong predictors of race [18].

For this element of the algorithm, the questions journalists should ask are: Is the algorithm using the appropriate data for analysis? Is it avoiding data that it should not be using?

The algorithmic calculation

The central element of an algorithmic system is the calculation itself; the instructions that are designed to process the data according to the algorithm's objective. These can vary from simple weighted scoring systems to complex amalgamations of different algorithms into one larger system.

A basic threat to calculations are mistakes such as typos in the code. Software bugs are pervasive and diverse and caused financial losses of $1.7 trillion in 2017 [23].

In more complex algorithms, such as predictive systems that use machine learning to detect a feature or classify data into categories, the concept of errors is more nuanced. Since prediction is based on statistical inferences, there is always a degree to which the algorithm will make incorrect predictions. There are specific measurements that account for these incorrect predictions, such as false positives (type I error) and false negatives (type II). The accuracy and precision of an algorithm are calculated based on these values. In such complex systems, the threat is not only that the calculation is wrong per se; but what is, first, the *magnitude* and, second, the *direction* of the error.

The magnitude question is usually the purview of validation and testing which determines whether algorithms are considered accurate enough to be adopted. Reports from those tests can be useful for investigations.

The direction of the error is more nuanced. False positives and false negatives have different implications, depending on the application of the algorithm [7]. For example, in an algorithm that determines which restaurant gets a health inspection, a false positive might send an inspector to a clean restaurant, and a false negative does not flag a dirty restaurant for inspection. In a false positive, relatively little harm was done: the inspectors and restaurant lost some time, and this strained the resources of an inspecting agency. But in a false negative, a dangerous restaurant is still operating and putting customers at risk.

This trade-off can be more nuanced when taking into account moral values of the society in which the algorithm is operating. Consider an example of an algorithm that determines if a criminal should go to prison or be released on bail. A false positive can flag a low-risk person as dangerous and keep that person incarcerated. A false negative can release a high-risk criminal. Depending on your view of incarceration, one is more damaging than the other.

For this step, journalists should then ask: Is the algorithm making the expected calculation with the data? What are the threats of false positives and false negatives?

The output

As in any information system, the value of information generated by an algorithm is whether that information is actionable and valuable in decision-making processes. On the output side of an algorithm, there are two main threats: lack of usefulness and lack of understandability.

An example of lack of usefulness currently involves predictive policing, a system that uses historical crime data to predict when and where a crime would occur. Aside from the biases involved in the input data for such algorithms [14], there are issues as to what would be done with the data output by the algorithm.

The output of predictive policing systems can be at the same time harmful and not novel. On one hand, since these systems activate more police activity, they also generate more crime data, which in turn generates a feedback loop of policing [8]. On the other hand, there have been instances in

which police officers have questioned the usefulness of the output of predictive policing. The adoption of the algorithm can clash with the culture and craft of policing [20].

The other threat is understandability of the data or the actions generated by the algorithm. For any system, its operators must be able to understand and react to its outputs. In order to understand the types of harm from misreading algorithms, journalists should be aware of what the operators of the system are seeing and how they can react.

When two Boeing 737 Max airliners crashed in October 2018 and March 2019, killing 336 people, one of the issues was understability of an algorithm. The new airliners had a software called the Maneuvering Characteristics Augmentation System (MCAS), that compensated an inherent imbalance in the design of the airplane that caused it to tilt back under certain conditions. MCAS detects when the angle of attack of an airplane is too steep and corrects it, adjusting the stabilizers to lower the nose of the plane and pushing the pilot's yoke down. But pilots did not receive any training on this algorithm, and were unable to identify what its output (the lowering of the nose and the yoke moving down) meant [10].

In this step, journalists should ask: Are the data or actions generated by the algorithm useful and understandable?

The overall idea of the algorithm

Since algorithms are socio-technical system, an investigating journalist must also take into account the social circumstances in which an algorithm is embedded and whether or not its original idea is ethical. In other words, if the calculation of the algorithm works perfectly without bugs; if the inputs are appropriate to the task of the algorithm; and if the outputs are clearly understandable and actionable by the operators; does that mean that the algorithm is free of any threats? Does a perfect technical chain result in a perfect system? Not if the system rests on faulty values and ethics. As a matter of fact, in those cases, effective systems actually enable large-scale wrongdoing.

A salient example here is facial recognition software and its applications. There are currently multiple facial recognition systems that are aptly using the input data (images of faces), have a high degree of accuracy, and are used by their operators as intended. However, those intentions – and the underlying assumptions – are what make facial recognition systems problematic. For instance, the government of China uses facial recognition systems to track members of the Uighur ethnic minority [17]. The issue is not just a misuse or misapplication; facial recognition systems are based on faulty race and gender categorizations, and help reinforce them, which led some researchers to equate this technology to plutonium: dangerous and with few legitimate uses [22].

Journalists who cover algorithms must therefore always be mindful not only of the individual components that make these systems do unpredictable or unintended things, but also reflect on predicted and expected outcomes that are harmful to society. The questions journalists should ask here are more abstract: Is the algorithm as a whole ethical? Does it enable what it should enable, according to society's values?

3 INVESTIGATING ALGORITHMS WITH DIFFERENT LEVELS OF ACCESS

Now that we have delineated the components of algorithms and their respective threats, we will explore how journalists can investigate algorithms based on their access to each of those items, ranging from wide access to technical details to general value-based descriptions of the automated decision systems. As we will see, the variety of data availability presents different opportunities and challenges for investigation. Table 2 shows possible approaches for each type of access, and the following subsections explore each scenario in more detail.

Table 2: Different levels of access in investigations.

Scenario	Approaches
Access to code	Inspection for bugs; inspection for inputs; inspection of outputs after running with simulated data
Access to both input and output	Reverse-engineering; auditing
Access to either input and output	Descriptive data analysis of inputs or outputs; comparison with other data sources
Access to supplementary information	Reporting on values of algorithm; interviewing users and people impacted by algorithm

Access to code

A case with the most transparency in algorithms is when the investigators have access to the underlying code that does the actual calculation and data manipulation. With that at hand, one can inspect the code to see if there are any bugs in logic or implementation, or run it using simulated data to see hypothetical outcomes. These approaches also allow one to see the output of the algorithm in its most natural form, so that the investigator can determine its interpretability. The trade-off to this level of access is that it requires specific technical abilities, namely to handle and read code.

Even among journalists that have the technical ability to inspect and run code, however, the hardest step of this type of investigation is to actually have access to it. There are a limited number of ways in which a journalist can have access

to code. There are cases in which the developers themselves might open the code for inspection, as in the case of Brazil's election system [25]. In other cases, when the system is developed or used by the government, journalists can try public information requests [7]. However, those types of requests usually restrict access to trade secrets, which is the case for algorithms that are developed by private companies, even when government is using them [7].

In 2017, ProPublica investigated complaints about mistaken convictions caused by a proprietary software used for DNA testing of crime scene evidence in the state of New York [13]. The source code had been requested by the defense attorneys of a suspect, had been evaluated by experts and was shown to be faulty. But it was still being withheld from the public. ProPublica subsequently filed a motion in the Southern District of New York to request access to that source code [12]. The material was eventually unsealed and shared publicly by ProPublica [11] in a GitHub repository[1].

Access to both input and output data

Even having access to only the input and output data of an algorithm is enough for powerful investigations. If the internal mechanism of an algorithm can be equated to a black box, the inputs and outputs can be seen as doors into it [5]. In those cases, journalists can begin to reverse-engineer the internal calculations of the system by doing data analysis that compare inputs to outputs.

That is what the New York Times' The Upshot did when they investigated Chicago's Strategic Subject List (SSL). The SSL is generated by an algorithm that predicts the likelihood of a person being involved in a shooting incident, either as an offender or a victim [4]. The algorithm itself is not publicly available, but the list, with 399,000 individuals, is released on the city's open data portal[2].

The list does not contain names, but it does have useful information for data analysis aside from the actual risk score assigned to that person: it contains age, sex, race/ethnicity, whether or not that person was arrested for drug or weapons offenses, location of latest arrest (if any), among other data.

With both the risk scores (the output) and the data that is used to calculate the risk score (the input), journalists were able to do a linear regression analysis and isolate the criteria that correlated more or less with the score [2]. Additionally, if the analysis were conducted with all but a small test sample of the dataset, the journalists could validate their findings with their own dataset, by predicting the scores of that remaining test sample.

The Upshot found that the algorithm assigned a higher score for individuals who were younger, and had been crime victims. Contrary to what the police department had stated, gang affiliation was a relatively low predictor of high risk score [2]. So, with reverse engineering, the reporters were able to contest a government narrative.

Access to either input or output data

There are cases in which journalists cannot reverse-engineer the algorithm because they only have one end of the data – either the output or the input. In this case, however, it is still possible to obtain valuable insights into the algorithms by using other types of data analysis.

When ProPublica investigated Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), an algorithm that assesses a criminal defendant's likelihood of becoming a recidivist, they obtained via public request the risk scores from 18,610 people [1]. In order to actually investigate whether the algorithm was biased, they first did a descriptive data analysis of the scores according to race of person rated, and found that Black defendants had a higher proportion of high risk of recidivism. Then they compared the scores with another dataset that they obtained, the public criminal records from the same area, to see the proportion of people who would go on to become recidivists, by race. They found that Black defendants had a higher proportion of false positives than White defendants. In other words, they did not need the inputs of the algorithm to tell the story they wanted to tell: they only needed the outputs, which they analyzed and compared with another dataset.

Access to supplementary and contextual information

There are cases in which no output or input data is available, either because the algorithm is secretive, proprietary, or still under development. While this lack of access makes it difficult to quantify an algorithm, investigations can also use supplementary information or descriptions of the systems.

Journalists might be able to request data dictionaries or instruction manuals, for instance, to get a glimpse as to what the operators of the algorithm would see. But even if that is unavailable, journalists can gather information using traditional reporting techniques. Interviewing people – not only the ones who are involved in the use of algorithms, but also people impacted by it – should always be an element of algorithmic accountability reporting. Those interviews might provide high-level insights into how they work.

Press releases and other institutional documents that aim to publicize the algorithm are also a trove of interesting information. While they might not provide deep technical information, they can give some understanding of what the developers think is important, or what do they aim to solve with their system.

Looking through documentation is what websites that cover search engines do [21]. They keep an eye out for every

[1] https://github.com/propublica/nyc-dna-software

[2] data.cityofchicago.org/Public-Safety/Strategic-Subject-List/4aki-r3np

guideline or statement that comes out Google or Bing. They parse out information and compare it to older documents, to find meaningful differences.

Another example of the use of contextual information is the New York Times' report on the use of facial recognition by the government of China [17]. While they do review some data that is generated by the algorithm, the bulk of the reporting is done through interviewing experts, giving context around why the government would be interested in these systems and the identities of the companies supplying them [17].

4 DISCUSSION

In this paper, we developed a framework that outlines a systematic approach that journalists can take when critiquing an algorithmic system. We do that by untangling the variety of threats each component of an algorithm (input, calculation, and output) creates, and by exploring scenarios of varied access to information and approaches that have been effectively used in them.

This framework, of course, is not definitive: the variety of algorithms and their applications and the creativity of investigators are hard to predict. But the contribution we hope to bring is a conceptual understanding of how algorithms can be understood in a way that informs their critique.

One notable aspect of this framework is that the more the investigation focuses on the central calculation of the algorithm, the more technical it is. The concerns are whether or not the system is doing what it promises. On the other end, questions about general design and idea of the algorithm are linked to the values that motivate it and whether those values match with the society in which they are embedded. In the intermediate, the appropriateness of inputs and uses of outputs relate to whether the algorithm is aptly designed for the social circumstances in which they are deployed.

The same tendency from technical to value-based can be traced when discussing the approaches to investigating algorithms. Access to the math raises questions about whether that math is correct. But if the algorithm, its inputs and outputs are opaque, the investigation is more based on values. Each of these approaches are appropriate in their own way.

What journalists actually do with their findings of algorithmic investigations, however, warrants more exploration. There have been interesting developments in trying to more clearly explain how algorithms work, using interactive elements and other data visualization techniques [9]. Also, more permanent news structures of watchdog journalism [16] are being established. Future work might focus on the tensions and needs of this type of journalism, as well as barriers to adoption not only by specialized news organizations, but by all journalists that will eventually come across algorithms on whatever beat may be their primary focus.

REFERENCES

[1] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. Machine Bias: there's software used across the country to predict future criminals. And it's biased against blacks. ProPublica 2016.
[2] Jeff Asher and Rob Arthur. 2017. Inside the algorithm that tries to predict gun violence in Chicago. *The New York Times* 13 (2017).
[3] Meredith Broussard. 2018. *Artificial unintelligence: how computers misunderstand the world.* MIT Press.
[4] Monica Davey. 2016. Chicago police try to predict who may shoot or be shot. *The New York Times* (2016).
[5] Nicholas Diakopoulos. 2014. Algorithmic accountability reporting: On the investigation of black boxes. Tow Center for Digital Journalism.
[6] Nicholas Diakopoulos. 2018. *The data journalism handbook 2.* European Journalism Centre and Google News Initiative, Chapter 6.
[7] Nicholas Diakopoulos. 2019. *Automating the News: How Algorithms Are Rewriting the Media.* Harvard University Press.
[8] Danielle Ensign, Sorelle A Friedler, Scott Neville, Carlos Scheidegger, and Suresh Venkatasubramanian. 2017. Runaway feedback loops in predictive policing. *arXiv preprint arXiv:1706.09847* (2017).
[9] Karen Hao and Jonathan Stray. 2019. Can You Make AI Fairer Than a Judge? MIT Technology Review.
[10] Phillip Johnston and Rozi Harris. 2019. The Boeing 737 MAX saga: lessons for software organizations. *Software Quality Professional* 21, 3 (2019), 4–12.
[11] Lauren Kirchner. 2017. Federal Judge Unseals New York Crime Lab's Software for Analyzing DNA Evidence. ProPublica.
[12] Lauren Kirchner. 2017. ProPublica Seeks Source Code for New York City's Disputed DNA Software. ProPublica.
[13] Lauren Kirchner. 2017. Thousands of Criminal Cases in New York Relied on Disputed DNA Testing Techniques. ProPublica.
[14] Kristian Lum and William Isaac. 2016. To predict and serve? *Significance* 13, 5 (2016), 14–19.
[15] Francesco Marconi and Rajiv Daldrup, Tilland Pant. 2019. Acing the Algorithmic Beat, Journalism's next Frontier. Nieman Lab.
[16] The Markup. 2019. About The Markup. The Markup. https://themarkup.org/about.html
[17] Paul Mozur. 2019. One Month, 500,000 Face Scans: How China Is Using AI to Profile a Minority. *The New York Times* 14 (2019).
[18] Cathy O'Neil. 2016. *Weapons of math destruction: How big data increases inequality and threatens democracy.* Broadway Books.
[19] Ismael Peña-López et al. 2018. *Algorithmic Accountability Policy Toolkit.* Technical Report. AI Now Institute.
[20] Jerry H Ratcliffe, Ralph B Taylor, and Ryan Fisher. 2019. Conflicts and congruencies between predictive policing and the patrol officer's craft. *Policing and Society* (2019), 1–17.
[21] SearchEngineLand. n.d.. About Search Engine Land. Search Engine Land. https://searchengineland.com/about
[22] Luke Stark. 2019. Facial recognition is the plutonium of AI. *XRDS: Crossroads, The ACM Magazine for Students* 25, 3 (2019), 50–55.
[23] Tricentis. 2018. Tricentis Software Fail Watch Finds 3.6 Billion People Affected and $1.7 Trillion Revenue Lost by Software Failures Last Year. Tricentis.
[24] Daniel Trielli, Jennifer A Stark, and Nicholas Diakopoulos. 2017. Algorithm Tips: A Resource for Algorithmic Accountability in Government. In *Computation + Journalism Symposium.* Evanston, IL, USA.
[25] TSE. 2017. Testes Públicos de Segurança do Sistema Eletrônico de Votação 2017. Superior Electoral Tribunal. Tribunal Superior Eleitoral. http://www.tse.jus.br/eleicoes/eleicoes-2018/testes-publicos-de-seguranca-do-sistema-eletronico-de-votacao
[26] Rodrigo Zamith. 2019. Algorithms and Journalism. In *Oxford Research Encyclopedia of Communication.* Oxford University Press.