

Enabling Accountability of Algorithmic Media: Transparency as a Constructive and Critical Lens

Nicholas Diakopoulos

Abstract As the news media adopts opaque algorithmic components into the production of news information it raises the question of how to maintain an accountable media system. One practical mechanism that can help expose the journalistic process, algorithmic or otherwise, is transparency. Algorithmic transparency can help to enable media accountability but is in its infancy and must be studied to understand how it can be employed in a productive and meaningful way in light of concerns over user experience, costs, manipulation, and privacy or legal issues. This chapter explores the application of an algorithmic transparency model that enumerates a range of possible information to disclose about algorithms in use in the news media. It applies this model as both a constructive tool, for guiding transparency around a news bot, and as a critical tool for questioning and evaluating the disclosures around a computational news product and a journalistic investigation involving statistical inferences. These case studies demonstrate the utility of the transparency model but also expose areas for future research.

Abbreviations

CAR	Computer-assisted reporting
NPR	National Public Radio
RTDNA	Radio Television Digital News Association
SPJ	Society for Professional Journalists
SRF	Schweizer Radio und Fernsehen

N. Diakopoulos (✉)

Philip Merrill College of Journalism, University of Maryland, College Park,
Knight Hall, 7765 Alumni Drive, College Park, MD 20742, USA
e-mail: nad@umd.edu

1 Introduction

The news media is rapidly adopting algorithmic approaches to the underlying journalistic tasks of information collection, curation, verification, and dissemination. These algorithmic approaches are changing the production pipeline of news as well as the user experience of news consumption. Major news outlets like the Associated Press routinely use automated writing software to produce corporate earnings articles based on structured data [1], or deploy news bots that disseminate and expand the audience for news information [2]. Statistical models can now be found simulating, predicting, or visualizing news stories in domains like public health or elections.¹ Recommendation systems steer users towards popular news articles based on social activity on Twitter or Facebook, or on content consumption patterns [3]. Headlines written by editors are optimized through A/B tests, or even algorithmically generated from scratch [4]. Breaking news alerts are accelerated through story clustering and detection [5], or by developing statistical techniques that can alert journalists to potential emerging stories [6]. And social listening and curation tools are used to identify, verify, and rank social media content as it finds its way to the homepage [7, 8].

Questions about media accountability and ethics emerge as these algorithmic approaches become more entrenched in how the media system produces and mediates news information. While such technologies enable scale and an ostensibly objective and factual approach to editorial decision making, they also harbor biases in the information that they include, exclude, highlight, or make salient to journalists or end-users. Such biases can be consequential for the formation of publics and the fair and uniform provision of information [9]. At the same time, such black box systems can be difficult to comprehend due to the opacity in their automated decision-making capabilities [10, 11].

Recently media ethics in journalism has seen a shift away from objectivity as a sole focus. The key issue with objectivity, as media ethicist Stephen Ward puts it, is “If a news report involves (at least some) interpretation, how can it be objective?” [12]. Pervasive interpretation in news making problematizes objectivity, from deciding what is newsworthy, to choosing the words and frames of a story, to amplifying the salience of some elements over others in a visual design. The norm of objectivity is waning in the face of substantial growth of what some have termed “contextual journalism” [13]—that is news that advances analysis and interpretation via context. In its stead there is an increasing emphasis on transparency as the guiding norm [14], which is now listed prominently in several ethics codes such as

¹<http://www.nytimes.com/interactive/2016/upshot/presidential-polls-forecast.html>.

that of the Society for Professional Journalists (SPJ),² the Radio Television Digital News Association (RTDNA),³ and National Public Radio (NPR).⁴

There is certainly an aspect of algorithms that are objective: they execute the same set of procedures consistently each time they are run. And the corporate rhetoric surrounding algorithms often frames them as objective and devoid of human influence. But researchers have long criticized the embedded biases in algorithms, which can emerge through their human design or the data that feeds and trains them [15]. Much as human decision processes are subject to individual or group biases, algorithms also harbor bias. In some cases algorithms are much more explicit about bias (e.g. if embedded rules or definitions are at play), and other times algorithmic bias is deeply mediated via machine learning processes that learn biases from the training data they are fed. There is a parallel here: algorithms also “interpret” the information they process via the biases designed into or emergent in them, and thus the same skepticism of objectivity brought against traditional journalism is also applicable to algorithmic media. Transparency as a normative practice to facilitate media accountability can similarly be employed as way to verify the interpretation of an algorithmic component used in the newsmaking process.

It’s worth noting that transparency is not a silver bullet for media ethics; a range of other media accountability approaches like press councils, clear codes of conduct, and healthy media criticism are all useful and appropriate in different contexts [16]. Transparency is valuable for media accountability because it facilitates testing and validating the interpretations that underlie newswork, traditional or algorithmic. Deuze [17] defines transparency as the “ways in which people both inside and external to journalism are given a chance to monitor, check, criticize and even intervene in the journalistic process.” Transparency entails the disclosure of information about the news production process so that audiences can monitor that process, which reinforces the legitimacy of the media.

Transparency around newswork has traditionally involved strategies like linking to sources and documents used in reporting, signaling source expertise and qualifications to comment on a given topic, offering a window into the editorial decision making process, noting relationships with partners or funders, or disclosing and correcting errors or failures [18]. But with algorithms and data it’s less clear how to translate the ideal of transparency espoused in ethics guides into practical strategies for media organizations. The goal of this chapter is to begin to fill this gap in pragmatic understanding of how to engage in effective algorithmic transparency.

Drawing on prior research, this chapter will first outline a set of information that might be disclosed about the types of algorithms in use at news media organizations. It is important to understand the space of possible information that might be disclosed about an algorithm as part of a transparency effort. Then this model of algorithmic transparency will be illustrated using case studies in

²<http://www.spj.org/ethicscode.asp>.

³http://www.rtdna.org/content/rtdna_code_of_ethics.

⁴<http://ethics.npr.org/>.

two capacities: constructive and critical. First, the model will be used as a way to structure and think constructively about algorithmic transparency in a news automation bot that we built called *AnecbotalNYT*. Then the model will be used as a critical lens to enable an evaluation of both editorial as well as news information products from Google and BuzzFeed News. Finally, the chapter will conclude with a discussion of challenges in applying algorithmic transparency in practice as well as opportunities for further research.

2 Transparency Model

This section presents a model of algorithmic transparency consisting of an enumeration of information factors that might be disclosed about algorithms in use in the news media. The full description of the model and how it was developed is presented in [19] but is briefly summarized here so that it's clear how the model is being applied to the case studies presented later in this chapter.

The model includes a set of pragmatic dimensions of information that might be disclosed as part of an algorithmic transparency effort. The dimensions of information disclosure were elicited through a series of focus groups that asked groups of 10–15 scholars and practitioners knowledgeable in algorithmic news media to brainstorm and discuss what might be disclosed around three case studies of algorithms used in news production. The cases were selected to address diverse aspects of how algorithms are used in news production processes and included automated news writing (e.g. data-driven writing software), algorithmic curation (e.g. moderation or recommendation), and simulation, prediction, and modeling in storytelling (e.g. election forecasts, visualized interactive models). Moderator-led focus groups spent an hour considering each of the cases and generated a host of candidate information factors that might be made transparent about the algorithms at the focus of that case. The transcripts for these focus groups were qualitatively analyzed through a process of iterative coding, affinity diagramming, typologizing and memoing [20] and indicated a set of layers across which information about an algorithm could be made transparent.

The four layers of information disclosure that were identified include: data, model, inference, and interface. Individual aspects that were discussed in the focus groups across the four layers articulated are shown in Table 1. The *data* layer is a key aspect where transparency is needed in algorithmic systems, particularly those that rely on machine learning. The information quality and validity of the data feeding algorithms was seen as paramount: garbage in, garbage out. Quality includes aspects of accuracy, uncertainty, timeliness, comprehensiveness, and provenance, as well as how data has been transformed, validated, cleaned, or edited. The *model* process and methodology for modeling were other aspects deemed worthy of disclosure. Details of the model might include the features, variables, weights, and type of model used. A model might also include heuristics, assumptions, rules, or constraints that might be useful and helpful to disclose. The output *inferences* from

Table 1 The transparency model used throughout this paper includes aspects of disclosable information across four layers: data, model, inference, interface

Layer	Factors
Data	<ul style="list-style-type: none">• Information quality<ul style="list-style-type: none">◦ Accuracy◦ Uncertainty (e.g. error margins)◦ Timeliness◦ Completeness• Sampling method• Definitions of variables• Provenance (e.g. sources, public or private)• Volume of training data used in machine learning• Assumptions of data collection• Inclusion of personally identifiable information
Model	<ul style="list-style-type: none">• Input variables and features• Target variable(s) for optimization• Feature weightings• Name or type of model• Software modeling tools used• Source code or pseudo-code• Ongoing human influence and updates• Explicitly embedded rules (e.g. thresholds)
Inference	<ul style="list-style-type: none">• Existence and types of inferences made• Benchmarks for accuracy• Error analysis (including e.g. remediation standards)• Confidence values or other uncertainty information
Interface	<ul style="list-style-type: none">• Algorithm “in use” notification• On/off• Tweakability of inputs, weights

an algorithmic process such as the classifications, predictions, or recommendations produced raised issues around errors, uncertainty, and accuracy. Transparency information might include what types of inferences are being made, as well as error analysis (and remediation standards for dealing with remaining errors), and disclosure of confidence values on inferences. Finally, at the interface layer we can talk about aspects of how algorithmic transparency information is integrated into the user experience, including through deeper interactivity, perturbation or sensitivity analysis, or even capabilities to override the algorithms (e.g. an on/off switch).

The results of the study also emphasize the deep entanglements of humans within algorithmic systems and the many human decisions that may need to be articulated as essential context in an algorithmic transparency effort. Such transparency around human involvement might include disclosing aspects of editorial goals or rationale for selection, inclusion, or exclusion of various inputs and outputs to an algorithm,

disclosing details of the configuration or parameterization of the system, and talking about adjustments to the system over time. In addition, knowing about the responsible actors, including the authors or designers of such systems could enhance the personal accountability and responsibility that individuals feel for such systems.

Barriers and challenges to effective algorithm transparency efforts were discussed. These included: concerns over disclosure of *proprietary* information that would damage competitive advantages, or leave a system open to *manipulation*; concerns with *overwhelming users* with too much information in the user interface that they might tune-out or not find relevant; business issues surrounding the *costs* of data preparation, documentation, or benchmark testing that would be entailed by a transparency effort; *privacy* considerations from disclosure of improperly anonymized data; and *legal* implications of admitting error or uncertainty in decision processes. Several of these issues, such as proprietary concerns and complexity in information presentation, are echoed in other research on the opacity of algorithms [21].

Subsequent sections in this chapter will present applications of this transparency model to specific instances of the use of algorithms in the news media. Different contexts and use cases may require the application of different subsets of facets enumerated in Table 1. The case studies take a user-centric approach in trying to align information with the decisions of end-users so that disclosures about algorithms maximize the possibility that users respond to that information [22]. Any information disclosure should have the potential to either impact an individual user's decision processes, or wider public understanding of aggregate system behavior. For that reason, during the application of the model to case studies, questions relating to the user decisions and the impacts of information disclosure that affect those decisions will be considered:

- What are the decisions the user or wider public would make based on the algorithm in question?
- What are the relevant bits of information about the algorithm that would help users or the public make those decisions more effectively? How might users respond to this information?
- What's the worst case outcome if users can't make such decisions effectively?

Moreover, the barriers and challenges to transparency that the model articulates will be considered. For instance, the issues of gaming, manipulation, and circumvention as well as the consequences of those actions will be examined. Aspects of how different types of information disclosures affect the user experience, as well as the costs of transparency information production will be taken into account.

3 Case Studies

In this section the model introduced above will be applied in several case studies. First, a computational journalism project involving the creation of a news bot will be described, including the ways in which the model informed the editorial

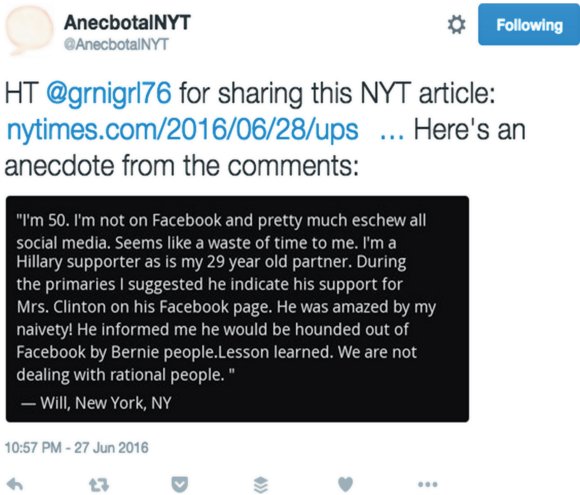
transparency of the algorithm used by the bot. Then, two case studies will be presented in which the model is used to develop critiques of a computational news product and a journalistic investigation.

3.1 Editorial Transparency in an Automated Newsbot

News bots are automated agents that participate in news and information dissemination, often on social media platforms like Twitter or in chat applications like Facebook Messenger [2]. Such bots can serve a range of journalistic tasks from curating and aggregating sources of information, to analyzing and processing data, and bridging different platforms to raise awareness or further dissemination. As newsrooms seek to develop and build more of these bots and other information appliances, it raises the question of how to express and provide editorial transparency of such tools.

Here we consider a Twitter news bot called AnecbotalNYT that we built to help surface and raise awareness for interesting personal experiences or anecdotes that relate to news articles published by the New York Times [23]. AnecbotalNYT works by listening to the public Twitter stream for anyone sharing a New York Times article link. It then collects all of the comments that are published on that article via the New York Times Community API, scores those comments according to several factors including length, reading level, and personal experience [8], and then ranks and selects a comment that is likely to contain a personal story or anecdote. Finally, the bot responds to the person who originally shared the article link with an image that contains the text of the comment. An example of the bot responding is shown in Fig. 1.

Fig. 1 An example of the output produced by the AnecbotalNYT Twitter bot



There are numerous editorial decisions that are embedded in the functioning of the bot, from where the data comes from, and how it's sampled, to how comments are filtered, scored, and ranked, as well as thresholds used to define reasonable constraints that improve the end-user experience of the bot such as filtering out short comments and ceasing to tweet if API limits are reached. In order to provide transparency for these various factors we were informed by the algorithmic transparency model presented above as we developed and published transparency information on the bot's Github page.⁵ As we considered what transparency information to include on the page we worked through the potential factors for disclosure enumerated in Table 1, asking whether there might be interest from potential end-users in having that information.

As a result, the Github page includes not only the full code base for the bot, as well as descriptions of the data, model, and inferences that the bot makes, but it also links to context such as research papers that informed the bot's design. Moreover, we explicitly parameterized many of the editorial rules the bot follows in an effort to make them explicit and salient. Instead of having to parse through dense code to see these editorial rules, they are embedded within configuration files which we felt would be easier to read and understand for non-technical users. Of course not everything could or should be parameterized but we identified what we thought were the most subjective editorial rules like the weights applied to the individual subscores in aggregating a final score, the threshold used to filter out comments that were short, or even how often the bot would potentially respond.

This approach to algorithmic transparency of the bot was motivated with two possible sets of users in mind: professional journalists who might adapt the open source code, and Twitter users who encounter the bot or had the bot respond to their tweets.

For professionals, the transparency was motivated by the desire to make sure that news organizations that might want to build on the open source code could explicitly see the design decisions embedded in the bot's algorithm and thus more clearly edit and produce the desired output according to their own editorial goals. The decision that such a news organization needs to make is whether it will adapt and use the open source code of the bot. The transparency information disclosed in the Github profile thus facilitates the decision making process by showing the code as well as showing how the bot makes decisions and can be flexibly re-parameterized for different editorial uses.

On the other hand, for Twitter end-users, the bot provides transparency via its Twitter profile, which explicitly links to its Github page. Users' curiosity may be aroused after interacting with the bot and thus they may be interested to explore what the bot is, who made it, and why it behaves the way it does. For instance, why did the bot choose a particular comment to share, or why didn't it respond to a different link the user had shared? The decision the user is implicitly making when they interact with the bot is whether to trust or rely on the information being presented by the

⁵<https://github.com/comp-journalism/Comment-Bot>.

bot. Is the comment that the bot shared real and is it a legitimate commentary made with respect to the news article? The transparency information becomes a tool for end-users who interact with the bot to gain additional information and an account of the bot's behavior, thus facilitating a more satisfying user experience.

3.2 Critical Application of the Algorithmic Transparency Model

The algorithmic transparency model presented above can be used not only as a way to guide the editorial transparency of news-oriented automation, like bots, but also as a way to inform the critical analysis of algorithms throughout society, including those that impact the media system which is the focus in this chapter. This evokes the notion of algorithmic accountability [11] which often involves different methods, such as reverse engineering or auditing, but which also encompasses the approach of critical commentary taken here. In this sub-section I apply the model in two separate cases that explore criticism of an algorithmic system with little to no transparency, and of an algorithmic methodology with a high-level of transparency. The two cases respectively focus on a news information product that Google created to curate and present issue-oriented quotes from presidential candidates directly within search results, and a piece of computational investigative journalism produced by BuzzFeed News in which simulation was used to identify potential fraud in professional Tennis.

3.2.1 Google's Issue Guide

On February 1st, 2016 Google launched a feature on its search results that highlighted US presidential candidates' views on various political issues like abortion, immigration, or taxes. If a user searched for "Hillary Clinton abortion" their search results would include an infobox generated by Google at the top of the page in a prominent position. If the candidate had provided a statement on the issue it was presented first, and then the user could browse through quotes by the presidential candidate extracted from news articles around the political issue. A view of search results showing all of the issues and with a specific issue (Economy and Jobs) expanded is shown in Fig. 2. Our original analysis and assessment of the bias in the tool was published in journalistic form by Slate [24].

Google does not position itself in the market as a media company but rather as a technology company. A common argument that platforms like Google, Twitter, and Facebook use is that since they do not produce original content, they should not be considered media companies [25]. However, in this case, Google is taking on a similar role to that of a news organization in terms making editorial decisions about the selection and salience of information that is available when users search for

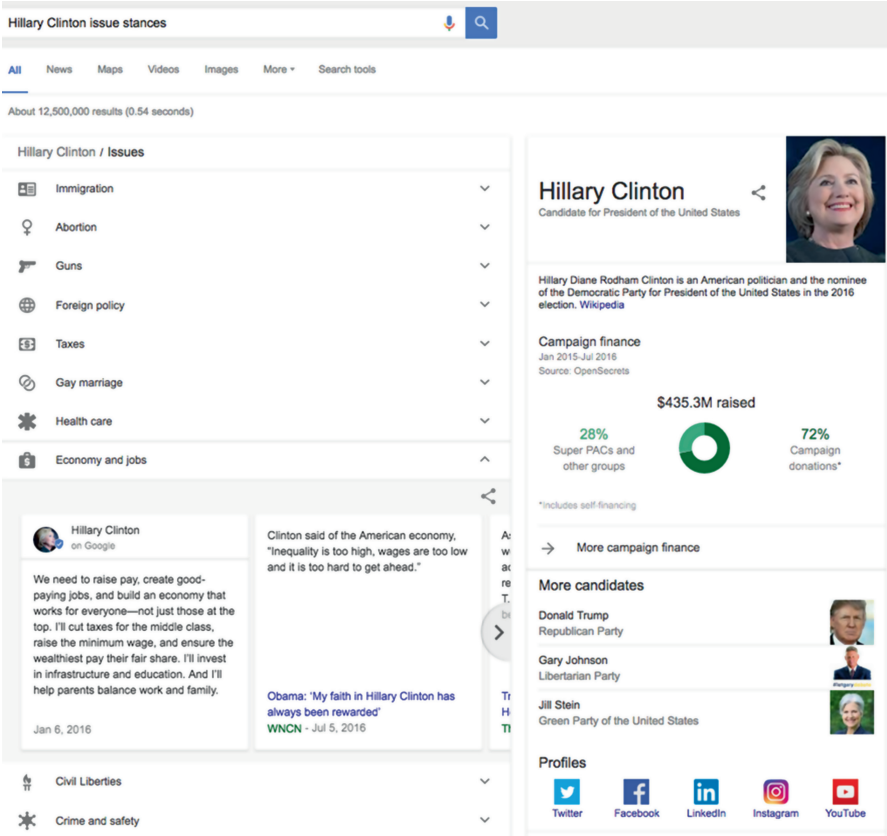


Fig. 2 The Google issue guide shown when searching for Hillary Clinton

candidate information. This matters because search engines are powerful arbiters of human attention that have the potential to shape the electorate’s attention and voting propensities [26] and are seen as trustworthy by a majority (73% according to a 2012 Pew study) of users [27].

To examine the biases that are visible in this new Google feature we gathered data including all of the news articles and quotes that were selected for the infoboxes on April 1, 2016 and again on April 22nd and May 13th, 2016. Our approach towards analyzing the data collected and asking questions based on our observations was informed by the algorithmic transparency model above.

The first aspect of the tool that we observed was that the ordering of the issues presented by the guide is not alphabetical, which might be the simplest logical algorithm for sorting the list of issues. The ordering bias of the issues is consequential because items listed higher up on search results pages get more attention from searchers [28]. We then considered whether that ranking might come from issue importance polls conducted by Gallup and Pew but found no correlation.

Because it wasn't clear what input variables and features were being used to rank the issues, this informed a question that we posed to the Google representative that we spoke to about our investigation. According to that representative the issues are ranked using "several factors" which include user interest in the various issues as well as the ways that other organizations rank issues (e.g. Pew Vote Smart, and OnTheIssues.org were mentioned explicitly). Thus we had limited success in getting enough detail in the information that Google disclosed. Because the Google spokesperson mentioned user interest as a factor in the ranking we attempted to recreate the issue ordering using Google Trends data, which indicates the frequency of searches for specific terms over time. However we were unable to recreate the issue ordering using simple searches for the issue terms on Google Trends.

Another aspect of the investigation was whether there was evidence of human influence in the ranking of issues. We collected data three times over the course of about 6 weeks to understand if there were significant shifts, but found the list instead to be relatively static. This suggests (but of course does not confirm) that any data-driven techniques used to order the issues are being passed through a human editorial or review process. The only change that we observed in the 6 weeks of observation was that a new top-level issue was added: gay marriage. But in examining the search volume of "gay marriage" on Google trends it was clear that there was not a data-driven peak or spike in interest in the topic. Before the change, quotes referring to same-sex marriage were including under the "civil liberties" issue. Only in the data collected on May 13th were those quotes moved to a separate issue. This suggests that human editors decided at some point that this issue was important or distinct enough to warrant its own top-level category.

Another algorithmic (or semi-algorithmic) process presumed to be a part of the issue guide is how the quotes are selected and presented for each candidate. We confirmed the presence of an algorithmic process to select these quotes with the Google representative who said, "The quotes are algorithmically aggregated from news articles on Google Search." This transparency about the system thus checked boxes both about an algorithmic process "in use", as well as the provenance of the input data feeding the quote extraction (i.e. news articles returned via Google Search). The quote selection process operates by scanning news articles to identify quotes made by a candidate and then associates each quote found with one of the issues that are being tracked.

What was journalistically interesting about these illuminations was both the focus on quotes to begin with, which weights the importance of what candidates say versus what they do, as well as the biases that we observed in the rate at which quotes were found when comparing Clinton and Trump. It turns out that as of May 14th, 2016 Clinton had 299 quotes across all issues whereas Trump had only 160. Such a discrepancy is difficult to explain from an external perspective as bias could be entering the system in several ways. For instance, the discrepancy in the number of quotes for each candidate might be due to Google not indexing the same number of articles for each candidate, or because the candidates simply aren't being covered equally by the mainstream media. Informed by the transparency model we considered whether there might be an input sample bias that could be causing the

output bias that we observed. To assess this we searched Google News for each candidate to enumerate all of the news articles that Google was indexing over the course of 1 month. We found 710 articles about Hillary Clinton, and 941 about Donald Trump, indicating that Trump is not underrepresented in the world of news that Google indexes. Other analyses have subsequently found a similar pattern in which Trump garners more media attention than Clinton.⁶

In this particular case we were interested in investigating the biases that were observable in the Google Issue Guide. The transparency model was informative for guiding certain questions that could be directed towards the company, such as confirming the use of algorithms, and eliciting information about the input variables and data. We felt that answers to these questions could inform a broad set of decisions that end-users might make based on the guide, such as whether to trust the information presented there, and whether the algorithm was treating the candidates fairly. By publishing this information we surmised that end-users might respond by discounting the tool since it does not clearly explain the biases that we observed. The algorithmic transparency model also guided the investigation by suggesting questions about input sampling bias that might be leading to the output biases observed, as well as orienting the investigation to look for signatures of human influence. While not yet at the stage of a “cookbook” that can guide the investigation or critique of an algorithm, we still found the model to be helpful in thinking through the *types* of information that we might be able to elicit or observe about the algorithm.

3.2.2 BuzzFeed News Investigation

Original analysis of data in a journalistic context is traditionally covered under the general rubric of Computer-Assisted Reporting (CAR) and has been practiced for decades [29]. It is not uncommon for news organizations that engage in such work to publish methodology articles or even whitepapers [30] that describe the process of analysis in a way that might be familiar to scientists who must be transparent about their data and methods for the sake of reproducibility. Outlets such as National Public Radio (NPR) and Schweizer Radio und Fernsehen (SRF) espouse the publication of open-source code and the use of programmatic data analysis and transformation in their data-journalism projects in order to ensure replicability.^{7,8} More recently, increasingly sophisticated uses of computing that go beyond standard statistical analysis, such as through advanced modeling, simulation, or prediction have been used in journalistic contexts. News organizations such as the New York Times, FiveThirtyEight, BuzzFeed and many others⁹ maintain repositories on

⁶<http://thedataface.com/trump-media-analysis/>.

⁷<http://blog.apps.npr.org/2014/09/02/reusable-data-processing.html>.

⁸<https://github.com/grssnbchr/rddj-reproducibility-workflow>.

⁹<https://github.com/silva-shih/open-journalism>.

Github, where they open-source their data and code in many (but not all) data- or computationally-driven news stories. In this section I will utilize the algorithmic transparency model as a mechanism to inform a critique of a particular case of a news organization being transparent in their use of simulation within a news investigation. The model is used to expose gaps in the methodological transparency offered in the project.

The focus of this case study is an article entitled “The Tennis Racket” published by BuzzFeed News [31]. This story investigates the broader issue of fraud in professional tennis, specifically as it relates to the possibility that players can easily throw a match (i.e. lose intentionally) and thus cause people betting against them to win. The sport is plagued by a lack of dedicated attention to both identifying and vigorously pursuing instances of suspected fraud, thus making it an interesting area for public interest investigation. The BuzzFeed investigation was particularly innovative in its use of simulation to identify professional players whose win-loss records in conjunction with betting odds statistics suggest that there was a suspiciously high chance they lost a game on purpose. By analyzing betting odds at both the beginning and ending of a match it’s possible to identify instances where there was a large swing (e.g. greater than 10% difference) which might be considered unlikely (if we are to believe the bookmakers’ initial odds calculation). If there are enough of these matches it casts suspicion on the player, but the article is clear to state that this statistical evidence is only suggestive and that “Identifying whether someone has fixed a match is incredibly difficult without betting records, without access to telephone records, without access to financial transactions.” Thus, the names of suspicious players are not included in the article, or in the other transparency disclosures of the project.

An innovative aspect of the approach towards editorial transparency in this project is that transparency disclosures occur in the interface at three levels (T1, T2, T3) of increasing sophistication and detail and are targeted at different types of end-users. Linked from the original story is an article (T1) that can only be described as “Buzzfeed-y” in the sense that it includes comical animated gifs as well as crystallized sub-heads and short accessible paragraphs that describe information about the methods and data used, as well as what was found as a result [32]. From there, a reader can click deeper on the “detailed methodology” (T2) which is displayed on a Github page¹⁰ that describes six methodological details such as the data acquisition and preparation as well as other calculations. This detailed methodology then links to the final, and most detailed transparency disclosure (T3): the source code used to run the analysis which is presented as a Jupyter Notebook with python code and additional annotations describing what each section of code does in terms of the final analysis.¹¹ Each level of disclosure adds additional details and nuance to the understanding of the analysis and thus allows different

¹⁰<https://github.com/BuzzFeedNews/2016-01-tennis-betting-analysis>.

¹¹<https://github.com/BuzzFeedNews/2016-01-tennis-betting-analysis/blob/master/notebooks/tennis-analysis.ipynb>.

stakeholders to access the granularity of information that is most relevant to their interests and needs. A truly dedicated journalist, or a professional tennis investigator might be motivated to use T3 information to rerun the analysis code for the sake of reproducibility, or to try it out with a different set of public or private data. But someone who is merely curious about the provenance and sample of the data could access the T2 description without the additional cognitive complexity of understanding code in T3. This approach to transparency disclosure follows a multi-level “pyramid” structure involving progressively denser and more detailed transparency information [19]. Importantly, it mitigates one of the primary concerns around implementing algorithmic transparency—that it is too difficult for end-users to make sense of transparency disclosures around algorithms.

The transparency disclosures across T1, T2, and T3 hit on many of the possible dimensions of disclosable information enumerated in the algorithmic transparency model across layers of data, model, and inference. But the model also suggests gaps where there could have been disclosures elaborated more precisely.

At the data layer, provenance is indicated (i.e. OddsPortal.com), as well as the purposive sampling strategy used to focus on certain types of matches (ATP and Grand Slam). At the same time, the sample was underspecified because the bookmakers from which the data was sampled were not named. All that is disclosed is that the data came from, “seven large, independent bookmakers whose odds are available on OddsPortal.com,” but there were over 80 different bookmakers listed on OddsPortal.com when I looked in September 2016. If it was important to keep these bookmakers anonymous the disclosure could have included a rationale for that decision. This lack of sampling detail makes it more difficult to reproduce the analysis by re-collecting the data independently. Furthermore there is lack of specificity in the timeframe of data collected. We’re told that the tennis matches analyzed occurred, “between 2009 and mid-September 2015,” though the disclosure easily could have specified the exact dates so that their sample could be reproduced. Finally, there is at least one assumption around the data, which is that the bookmakers’ initial odds are somehow accurate. Since the model relies on a shift in those odds the accuracy of the initial value is particularly important.

The model specification included in T2 and T3 is quite extensive. We are linked to the source code in T3 permitting very detailed examination of how the model was created and simulated. The target variable of odds-movement is clearly identified and defined. And we’re explicitly told the specific thresholds used (i.e. 10%) to define what constituted a suspicious level in terms of odds-movement, which is further rationalized by citing “discussions with sports-betting investigators.” However, the expertise and identity of these investigators could have been disclosed and that transparency would have facilitated additional trust and potential to verify and replicate the findings. T2 indicates other thresholds and sensitivity levels for the model as well, such as that the odds movement calculation only had to reach the 10% level for one out of the seven bookmakers being analyzed in order to be further considered.

At the inference layer of the project it is clear from the transparency disclosures what is being calculated (i.e. odds-movement) and how those calculations are then

used in a simulation process to infer suspicious activity on the part of players. Moreover, additional information about the confidence of inferences is given by using statistical significance testing including Bonferroni correction to reduce the possibility of false positives. The final report lists suspicious players (anonymously) labeled according to their statistical significance at difference standard levels including $p < 0.05$, $p < 0.01$, $p < 0.001$, as well as listed for the Bonferroni corrected level. This allows us to see the degree of statistical confidence in the inference and judge the possibility of false positives in the final set of 15 players. Conspicuously missing from T2 or T3 however is a disclosure indicating error analysis. The closest we get is a note in the T3 Jupyter notebook stating, “In some simulations an additional player received an estimated likelihood just barely under 0.05. To be conservative we are not including that player among our totals.” But this implies that there were multiple runs of the simulation procedure and that additional error analysis might have been tabulated, or even confidence intervals listed around each of the likelihood values reported.

In this case, the algorithmic transparency model was useful for identifying gaps in the information disclosures surrounding the methods used. This was a very solid project, done by seasoned reporters, and many of the dimensions of disclosure enumerated in the model were indeed covered. But there were some opportunities for additional disclosure around the data sample used, around the rationale for thresholds, and around the error bounds of the inferences. These criticisms elucidate opportunities to have bullet-proofed the disclosed methods even further. The implications for a broader public are limited since much of this criticism focuses on the lower levels (T2, T3) of more technical information disclosure. Mostly the story and its methods serve to spur additional investigation by sanctioned tennis investigators who have access to all of the necessary records in order to investigate suspicious players.

The journalists who worked on this story were thoughtful about not naming the players suggested by their methodology. Heidi Blake, the first author of the main article, works as the UK investigations editor for BuzzFeed News and is based in London, England. Since British law makes it considerably easier to sue for libel than in the US [33], one might speculate that BuzzFeed News thought the statistical evidence was not strong enough to stave off libel lawsuits in that jurisdiction. At the same time, the level of transparency in the project, particularly in T2 and T3 allowed for a team of three undergraduate students to ostensibly de-anonymize the results relatively quickly and easily.¹² The students re-scraped the odds information (though there was some uncertainty in identifying the seven used in the original story) with identities preserved, and then cross-referenced with the BuzzFeed data based on the other data fields available in order to associate a name with each of the 15 players identified in the original analysis. We can see this use of transparency information as undermining the intent of the journalists publishing the story, but we can also see it in a more constructive light. By stopping short of naming

¹²<https://medium.com/@rkaplan/finding-the-tennis-suspects-c2d9f198c33d#.q1axxecwd>.

names BuzzFeed news was able to avoid the risk of potential libel lawsuits. But by putting enough information and method out there so that others could replicate the work with their own non-anonymized data, this allowed a less legally averse outlet (students blogging about it) to publish the names. Presumably the same replication could be done by tennis investigators who would also be interested in the specific identities of players suspected of match fixing.

4 Discussion

In this chapter I have presented three cases in which an algorithmic transparency model informed both constructive and critical approaches towards media accountability. In the first case the model was used to guide the disclosure of editorial information about a news bot that might be useful to other journalists as well as to end-users. In the second case, the model suggested questions that were used to orient an investigation of the biases embedded in a news product tied to Google search rankings. It guided certain questions that could be directed towards the company, such as confirming the use of algorithms, and eliciting information about the input variables and data. And in the final case, the model identified gaps in the methodological transparency disclosed in a computationally informed piece of investigative journalism, suggesting opportunities that could have strengthened the disclosure from a replicability standpoint. This diversity in cases is important as it shows the versatility of the algorithmic transparency model employed and its ability to inform a range of activities that facilitate media accountability: of algorithmically-driven news products as well as algorithmically informed news investigations that result in a more traditional story. The model was found to be helpful in thinking through the different types of information that might be either voluntarily disclosed or elicited through interactive questioning or observation of an algorithm.

Still, there are a number of open questions and challenges in regards to algorithmic transparency and media accountability. One criticism that has been leveled at algorithmic transparency is that by disclosing too much information this would lead to a cognitive complexity challenge that overwhelms end-users who then don't know what to make of that information [19]. The BuzzFeed case addressed this issue by presenting information in a hierarchy of increasing detail, allowing end-users to choose the fidelity of transparency information that fit with their needs. The critique presented in this chapter suggests there might be even more nuance and detail that could be layered into these disclosures. But there is a key bit of additional work to do here. Research still needs to be done with end-users to see if such an approach is effective from their point of view. Such user research might report the rate of access or use of transparency information across different levels, or it might ask end-users explicitly what they found valuable from the disclosure at a particular level. It's not that user feedback should strictly dictate what is disclosed, but it may impact the interface and presentation of the disclosure. More generally,

research also still needs to answer the question of the magnitude of end-user demand for algorithmic transparency, and the specific contexts and circumstances in which such disclosures do become important to a broad public.

Another criticism of increased algorithmic transparency is the cost of producing additional information disclosures. Such efforts should not be underestimated as they do create an additional time and resource burden on the creators of algorithmic systems, both in terms of the initial creation of that information but also in terms of maintenance. For one-off stories like the tennis investigation the ongoing cost is probably not that substantial. BuzzFeed would merely need to update the code or documentation if there is a major new revelation or if someone submits a bug fix (roughly parallel to the cost of posting a correction in traditional media). BuzzFeed needed to produce the code and comments in order to do the story anyway, so the additional cost of making it public is incremental. But for platforms or products, open sourcing code could lead to substantially more time and effort spent on maintaining that code base including “customer-service” related to answering questions from other users interested in adopting the code. Moreover, algorithmic systems are often dynamic and depending on how frequently the product changes this raises the question and challenge of keeping transparency disclosures up-to-date with the latest version of the product. If a product opts not to disclose detailed data, code, or methodology, the cost of producing adequate transparency documentation could be reasonable though. In one personal experience working with a data analytics startup on a transparency disclosure for a new product, it took roughly 5 h including interviewing an engineer along the lines of the algorithmic transparency model, and writing up the disclosure in accessible language that was then vetted by the product manager. Still, more work needs to be done, perhaps by workshopping the production of such disclosures with many more members of industry, so that the bottlenecks and resources needed in order to create and publish meaningful transparency information can be better understood.

Transparency information was made available on the Github platform in both the news bot as well as the tennis investigation case studies. Github is a popular service that allows for putting code and documentation into online repositories that can be kept private or made public for open source projects. This has several advantages including the ability to version manage the disclosure, track any edits or changes, fork into new projects, track social interest through repository starring activity, and render commingled code and descriptive notes in Jupyter Notebook format. At the same time there are constraints to Github, such as strict file size limits that make the platform unsuitable for depositing large data files. For open and extensible projects like news bots, or investigative methodologies, platforms like Github serve well at the medium to low-levels of transparency. But presentation of transparency information that way might make less sense in the case of something like the Google infoboxes. In that case a written disclosure in the form of a blog post might have been sufficient since code-level detail is not strictly necessary (or warranted) for a computational product. Products surface different types of concerns around transparency, such as the loss of trade secrecy, or the weakening of a system with respect to manipulation or gaming. But while Google may be averse to showing

the code and the exact methods in which candidate quotes are integrated into their search engine, there are other factors (e.g. variables of interest, rationale for ordering decisions) that could be explained in a written disclosure so that end users have a higher degree of reliance on the information they receive. An open question for research is to determine the most appropriate vehicle, format, or platform to publish transparency information given the intended end-users and their goals. In some cases it may make sense to integrate transparency information directly into the user interface, but in others having that information linked on a repository, or published in a blog post or whitepaper may be more appropriate.

In closing, I would reiterate that there is still much work to do in studying algorithmic transparency. This chapter has shown some of the nuances in transparency around product-oriented computation versus editorial-oriented computation. Additional research should examine such differences and connect these to concerns over issues of proprietary trade secrecy or concerns over manipulation. The model applied in this chapter should be seen as a starting point that was helpful for enumerating different types of information that might be disclosed about algorithms, both in the context of using algorithms to create media, as well as in the context of investigating and critiquing others' use of algorithms in the media.

References

1. Graefe, A.: Guide to Automated Journalism. Tow Center for Digital Journalism, New York, NY (2016)
2. Lokot, T., Diakopoulos, N.: News bots: automating news and information dissemination on Twitter. *Digit. J.* **4**, 682–699 (2016)
3. Spangher, A.: Building the Next New York Times Recommendation Engine. New York Times, New York, NY (2015)
4. Wang, S., Han, E.-H. (Sam), Rush, A.M.: Headliner: an integrated headline suggestion system. In: *Computation C Journalism Symposium*, Palo Alto, CA (2016)
5. Wang, S., Han, E.-H.: BreakFast: analyzing celerity of news. In: *International Conference on Machine Learning and Applications (ICMLA)*, Miami, FL (2015)
6. Magnusson, M., Finnäs, J., Wallentin, L.: Finding the news lead in the data haystack: automated local data journalism using crime data. In: *Computation + Journalism Symposium*, Palo Alto, CA (2016)
7. Thurman, N., Schifferes, S., Fletcher, R., et al.: Giving computers a nose for news. *Digit. J.* **4**(7), 743–744 (2016). doi:10.1080/21670811.2016.1149436
8. Park, D.G., Sachar, S., Diakopoulos, N., Elmqvist, N.: Supporting comment moderators in identifying high quality online news comments. In: *Proceedings of the Conference on Human Factors in Computing Systems (CHI)* (2016)
9. Gillespie, T.: The relevance of algorithms. In: *Media Technologies: Essays on Communication, Materiality, and Society*. The MIT Press, Cambridge, MA (2014)
10. Pasquale, F.: *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press, Cambridge, MA (2015)
11. Diakopoulos, N.: Algorithmic accountability: journalistic investigation of computational power structures. *Digit. J.* **3**, 398–415 (2015)
12. Ward, S.J.A.: *Radical Media Ethics: A Global Approach*. Wiley-Blackwell, Malden, MA (2015)

13. Fink, K., Schudson, M.: The rise of contextual journalism, 1950s–2000s. *Journalism*. **15**, 3–20 (2014)
14. McBride, K., Rosenstiel, T.: The new ethics of journalism: principles for the 21st century. In: *The New Ethics of Journalism: Principles for the 21st Century*. Sage, Thousand Oaks, CA (2013)
15. Friedman, B., Nissenbaum, H.: Bias in computer systems. *ACM Trans. Inf. Syst.* **14**, 330–347 (1996)
16. Fengler, S., Russ-Mohl, S.: The (behavioral) economics of media accountability. In: Fengler, S., Eberwein, T., Mazzoleni, G., Porlezza, C. (eds.) *Journalists and Media Accountability: An International Study of News People in the Digital Age*, pp. 213–230. Peter Lang, New York (2014)
17. Deuze, M.: What is journalism?: professional identity and ideology of journalists reconsidered. *Journalism*. **6**, 442–464 (2005). doi:10.1177/1464884905056815
18. Silverman, C.: Corrections and ethics. In: McBride, K., Rosenstiel, T. (eds.) *The New Ethics of Journalism—Principles for the 21st Century*, pp. 151–161. Sage, Thousand Oaks, CA (2013)
19. Diakopoulos, N., Koliska, M.: Algorithmic transparency in the news media. *Digit. J.* (2016). <http://nca.tandfonline.com/doi/abs/10.1080/21670811.2016.1208053>
20. Lofland, J., Lofland, L.: *Analyzing Social Settings: A Guide to Qualitative Observation and Analysis*, 3rd edn. Wadsworth Publishing Company, Belmont, CA (1995)
21. Burrell, J.: How the machine “thinks”: understanding opacity in machine learning algorithms. *Big Data Soc.* **3**(1), 1–12 (2016)
22. Fung, A., Graham, M., Weil, D.: *Full Disclosure: The Perils and Promise of Transparency*. Cambridge University Press, Cambridge (2009)
23. Stark, J., Diakopoulos, N.: Towards editorial transparency in computational journalism. In: *Computation + Journalism Symposium*, (2016)
24. Trielli, D., Mussende, S., Stark, J., Diakopoulos, N.: How the Google Issue Guide on Candidates is Biased. *Slate* (2016)
25. Napoli, P.M., Caplan, R.: When media companies insist they’re not media companies and why it matters for communications policy. In: *Telecommunications Policy Research Conference* (2016)
26. Epstein, R., Robertson, R.E.: The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E4512–E4521 (2015)
27. Purcell, K., Brenner, J., Rainie, L.: *Search Engine Use 2012*. Pew Internet and American Life Project, Washington, DC (2012)
28. Agichtein, E., Brill, E., Dumais, S., Ragno, R.: Learning user interaction models for predicting web search result preferences. In: *Proceedings of SIGIR 2006* (2006)
29. Meyer, P.: *Precision Journalism: A Reporter’s Introduction to Social Science Methods*, 4th edn. Rowman & Littlefield Publishers, Lanham, MD (2002)
30. Goschowski Jones, R., Ornstein, C.: *Matching Industry Payments to Medicare Prescribing Patterns: An Analysis* (2016). <https://static.propublica.org/projects/d4d/20160317-matching-industry-payments.pdf?22>
31. Blake, H., Templon, J.: The Tennis Racket. *BuzzFeed News* (2016). <https://www.buzzfeed.com/heidiblake/the-tennis-racket>
32. Templon, J.: How BuzzFeed News Used Betting Data to Investigate Match-Fixing in Tennis. *BuzzFeed News* (2016). <https://www.buzzfeed.com/johntemplon/how-we-used-data-to-investigate-match-fixing-in-tennis>
33. Shapiro, A.: On Libel and the Law, U.S. and U.K. Go Separate Ways. *National Public Radio (NPR)* (2015)

Transparent Data Mining for Big and Small Data

Cerquitelli, T.; Quercia, D.; Pasquale, F. (Eds.)

2017, XV, 215 p. 23 illus. in color., Hardcover

ISBN: 978-3-319-54023-8