Transparent to whom? No algorithmic accountability without a critical audience

Kemper, J.; Kolkman, D.

DOI
10.1080/1369118X.2018.1477967

Publication date
2019

Document Version
Final published version

Published in
Information, Communication & Society

INFORMATION, COMMUNICATION & SOCIETY
2019, VOL. 22, NO. 14, 2081–2096
https://doi.org/10.1080/1369118X.2018.1477967

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Transparent to whom? No algorithmic accountability without a critical audience

Jakko Kemper[a] and Daan Kolkman[b]

[a]Amsterdam School for Cultural Analysis, University of Amsterdam, Amsterdam, Netherlands; [b]Jheronimus Academy of Data Science, Technical University Eindhoven, Eindhoven, Netherlands

ABSTRACT

Big data and data science transform organizational decision-making. We increasingly defer decisions to algorithms because machines have earned a reputation of outperforming us. As algorithms become embedded within organizations, they become more influential and increasingly opaque. Those who create algorithms may make arbitrary decisions in all stages of the 'data value chain', yet these subjectivities are obscured from view. Algorithms come to reflect the biases of their creators, can reinforce established ways of thinking, and may favour some political orientations over others. This is a cause for concern and calls for more transparency in the development, implementation, and use of algorithms in public- and private-sector organizations. We argue that one elementary – yet key – question remains largely undiscussed. If transparency is a primary concern, then to whom should algorithms be transparent? We consider algorithms as socio-technical assemblages and conclude that without a *critical audience*, algorithms cannot be held accountable.

ARTICLE HISTORY
Received 4 October 2017
Accepted 11 May 2018

KEYWORDS
Data science; algorithms; transparency; algorithmic accountability; algorithmic decision-making; glitch studies

Society collects more data than ever before. Our databases contain emails, videos, audios, images, click streams, logs, posts, search queries, health records, and more (Sagiroglu & Sinanc, 2013). The abundance of available data and decreasing cost of computing capability leads to the digitization and automation of public- and private-sector decision-making. Application areas in government span from traffic management to public sector budgeting and food safety monitoring to cyber security (Janssen, Charalabidis, & Zuiderwijk, 2012). The private sector has also taken to the algorithm and found applications from e-commerce to logistics (Chen, Chiang, & Storey, 2012). Some algorithms, such as profiling systems, are used in either context (Hildebrandt, 2006). Examples of algorithms that the general public encounters include Google's PageRank algorithm that serves us with relevant search results, Spotify's weekly music recommendation algorithm, and dynamic pricing models that try to maximize the amount we pay for goods and services.

CONTACT Jakko Kemper j.kemper@uva.nl Amsterdam School for Cultural Analysis, University of Amsterdam, Spuistraat 134, Amsterdam 1012VB, Netherlands

The rapid development and dissemination of data science have set high expectations. Techniques such as deep learning and random forests can be used to develop highly accurate predictive models. Such algorithms are perceived to have considerable potential toward solving some of our society's most pressing issues such as mass migration and climate change (Floridi & Taddeo, 2016). Decisions that affect the lives of millions of people are increasingly underpinned with evidence that is created with algorithms. In some cases, such algorithms may carry more weight than human decision-makers, or have replaced human decision-making altogether. From an organizational perspective, big data and data science are perceived as techniques that can help reduce costs by scaling down bureaucracy and allowing organizations to make more effective decisions (Janssen & Kuk, 2016).

Algorithmic models

The concept 'algorithm' is subject to a number of interpretations, which we will not discuss at length here (see Kitchin, 2017 for an overview). In a narrow sense, an algorithm consists of a step-by-step procedure that processes numerical inputs to outputs (Stone, 1971). Colloquially, references to 'algorithms' may refer to a single algorithm, or a large collection of algorithms such as 'the Google algorithm' (Sandvig, Hamilton, Karahalios, & Langbort, 2014). Algorithms can be used for a variety of tasks, including information retrieval, image recognition, filtering, outlier detection, and recommendation. When one considers the sheer extent of everyday practices that are in some way modulated by the use of algorithms – from the trading-market to the realm of dating, from following the news to impression management – it might indeed not be so strange to speak of an 'algorithmic life' (Mazzotti, 2017).

For the sake of clarity, we will use the term 'algorithmic models' to refer to a particular subset of algorithms that is used to inform or make decisions. Because developing a formal definition of algorithmic models is beyond the scope of this paper, we adopt the following working definition by drawing on the work of several authors (Frigg & Hartmann, 2012; Gross & Strand, 2000; Haag & Kaupenjohann, 2001; Minsky, 1965): an algorithmic model is a formal representation of an object that an observer can use to answer questions about that object. Within this definition, 'observer' refers to a human or machine decision-maker.

Algorithmic models are imbued with the promise of bringing 'reliability and objectivity to otherwise uncertain procedures' (Mazzotti, 2017) and are associated with ideas of technocratic governance (Janssen & Kuk, 2016). This view stands in contrast with recent studies that show algorithmic models are value-laden and can introduce inadvertent biases. For instance, because of biases in input data, algorithmic models can learn to adopt similar discriminatory attitudes based on words associated with a particular gender or social group (Barocas & Selbst, 2016; Caliskan, Bryson, & Narayanan, 2017) – similarly, algorithmic models have been demonstrated to potentially reproduce racialized and sexualized repertoires, for example in the fields of car insurance and platform-mediated service work (Angwin, 2016; van Doorn, 2017).

Algorithmic accountability

Safeguarding the quality of such algorithmic model-informed decision-making requires scrutiny of data quality and all the subsequent steps in the so-called data value chain

(Miller & Mork, 2013) Such quality assurance has always been part of effective organizational decision-making. However, the recent surge in the use of big data and the increasing intricacy of algorithms have dramatically changed the complexity of such quality assurance (Peng et al., 2016). Moreover, the speed at which new data science techniques, tools, and libraries are developed and released is unprecedented. The resulting pervasive collection of data-centric innovations has been subjected to limited academic scrutiny, especially when compared to the dissemination of earlier statistical techniques (Gandomi & Haider, 2015).

The increased prominence of algorithmic models in organizational decision-making and the speed at which new data science techniques are developed and adopted, have been case for concern and has led many to call for increased transparency (Hildebrandt, 2012; Janssen & Kuk, 2016; Pasquale, 2015). Transparency can be understood as the 'understandability of a specific model' (Lepri, Oliver, Letouzé, Pentland, & Vinck, 2017, p. 9) and is seen as a requisite for algorithmic accountability. The transparency ideal has found its way into open standards for government. Some suggest that publication of datasets and other open-access schemes can bring about gains in transparency, accountability, and fairness (Lathrop & Ruma, 2010). At the same time, the limitations of transparency have been the subject of debate. Annany and Crawford (2016) suggest that transparency cannot be a characteristic of an algorithmic model. Rather the opacity of algorithms should be considered with a sensitivity for the contexts of their use; transparency is performed by socio-technical assemblages of algorithm and people.

This paper contributes to the discussion on transparency and algorithmic accountability by problematizing the notion of transparency and approaches it from a glitch studies perspective. We demonstrate that transparency of algorithms can only be attained by virtue of an interested critical audience. Even then, there are pronounced limits on the degree of transparency that can be attained. These considerations are particularly relevant in light of recent attempts of regulators – such as the GDPR – to develop guidelines for the use of algorithmic models more generally, and the right to explanation in particular (Council of Europe, 2017).

This paper is organized as follows. We begin with a succinct overview of the literature on transparency and responsible data science, highlighting the call for transparency that animates this discourse. We then introduce the field of glitch studies and explore it in order to illustrate the problematic role that temporality, criticality, and complexity play in attaining effective apprehensions of digital systems. These considerations will be tied to a wider discussion concerning the value of transparency. We conclude that those looking to improve the transparency of algorithmic models should look beyond open-access alone. If transparency is to bring about algorithmic accountability, we cannot ignore the role of a *critical audience*.

Guidelines for transparency

Although the scale of the challenges pertaining to the use of algorithmic models is unprecedented, the challenge in itself is not new. Examples of rogue algorithms and improper use of algorithm-produced evidence in organizational decision-making are manifold. The seminal example of the former is the Google Flu Trends algorithmic model that inadvertently predicted more than double the proportion of doctor visits for influenza-like illness (Lazer, Kennedy, King, & Vespignani, 2014). Earlier high-profile cases include the Red River flood incident in the United States, in which misinterpretation of model outputs led government to wrongly assume that the dikes were high enough (Pielke, 1999) and, more recently, the mistakes in the United Kingdom West-Coast mainline bidding process, which led to a cancellation of the franchise (Department for Transport, 2012).

Such mishaps spark the interest of policy-makers and inspire the development of guidelines for the responsible development and use of algorithms. For instance, the United Kingdom government issued a review of all 'business critical models' that inform policy-making. The results of this review are laid out in a report conducted by the (Macpherson, 2013). The report stresses the importance of proportionate quality assurance and states that models should be fit for purpose. However, such guidelines on proper model use have been around for quite some time (see Dekker, Groenendijk, Sliggers, & Verboom, 1990), which suggests that guidelines alone are not necessarily effective in preventing algorithmic calamities, nor do such guidelines naturally aid in making algorithms accountable.

The recent attention for algorithmic accountability has set off a new surge in writings on guidelines for proper data management and responsible algorithmic model use. Although it cannot be seen as an integrated field of inquiry, the recent surge of these ideas is sometimes referred to as 'responsible data science' (Stoyanovich et al., 2017). In the following, we discuss some examples of guidelines that have recently been put forward and that champion transparency, and identify a shared weakness of such work.

FAIR guiding principles

Academia is one area where the rapid digitization of society has had a profound impact on the day-to-day work. Online platforms for collaborations, literature recommendation engines, prepublication websites, and real-time metrics are some examples of new tools that are impacting academic research. At the same time, the rapid increase of the amount of available data and the growing complexity of the methods used to store, analyse, and model that data present challenges for the academic research process (Wilkinson, 2016; Wilkinson et al., 2017). This has not gone unnoticed and led to the formulation of the 'FAIR guiding principles for scientific data management and stewardship'. Although these guiding principles revolve around data management in academia, their impact extends well beyond the realm of universities, funding bodies, and publishers, which is why we discuss them here.

The FAIR principles find their origin in a workshop held at the Lorentz Centre in Leiden, the Netherlands. The discussions demonstrated that there exists a wide consensus for the development of minimal guiding principles for the management of research data and resulted in the definition of the FAIR principles. The FAIR acronym refers to data management that is Findable, Accessible, Interoperable, and Reusable (Wilkinson, 2016, p. 4; Wilkinson et al., 2017, p. 3):

Findable – data should be identified using globally unique, resolvable, and persistent identifiers, and should include machine-actionable contextual information that can be indexed to support human and machine discovery of that data;

Accessible – identified data should be accessible, optimally by both humans and machines, using a clearly defined protocol and, if necessary, with clearly defined rules for authorization/authentication;

Interoperable – data become interoperable when it is machine-actionable, using shared vocabularies and/or ontologies, inside of a syntactically and semantically machine-accessible format;

Reusable – reusable data will first be compliant with the F, A, and I principles, but further, will be sufficiently well-described with, for example, contextual information, so it can be accurately linked or integrated, like-with-like, with other data sources. Moreover, there should be sufficiently rich provenance information so reused data can be properly cited.

These four principles of FAIR hold equally for both humans and machines (Rodríguez-Iglesias et al., 2016). Recent discussions of the FAIR principles suggest that transparency – which is the focus of this paper – is essential, and stakeholders discussed the importance of data accessibility and the need to contextualize data (National Institute of Food and Agriculture, 2017).

FACT

Although the FAIR principles have a usefulness outside of academia, their primary concern is to improve data management for scholarly endeavours. Others have put forward guiding principles for the storage, analysis, and modelling of data in a wider perspective. The FACT principles have been positioned as an antithesis to the widely cited four V's of Big Data. The acronym covers the Fairness, Accuracy, Confidentiality, and Transparency for data science. FACT-proponents argue that the challenges faced by data science are in themselves not new. In statistics and economics, the cognitive biases of people (e.g. confirmation bias) and analytical pitfals (e.g. selection bias) have been subject of inquiry for many decades and are relatively well understood (see Tversky & Kahneman, 1974). The FACT authors suggest that big data and data science, however, bring something new to the domain of statistics, which requires us to develop new guidelines (Aalst, Bichler, & Heinzl, 2017). The Transparency component of the FACT principles can be described as follows (Responsible Data Science Initiative, 2016):[1]

Transparency – Data science should provide transparency; how to clarify answers in such a way that they become indisputable? Data science can only be effective if people trust the results and are able to correctly infer and interpret the outcomes. Data science should thus not be viewed as a black box that magically transforms data into value. Many design choices need to be made in a typical 'data science pipeline' as shown in Figure 1. The journey from raw data to meaningful conclusions involves multiple steps and actors; thus accountability and comprehensibility are key for transparency.

Recent contributions to the FACT-principles stress that these principles should be applied not only to the use of algorithmic models, but are equally important in the design and development phase. The data scientist ought to work responsibly from the moment he or she receives data. This continues through the phases of data wrangling, modelling, and deployment and should persist when algorithmic model results are interpreted and the

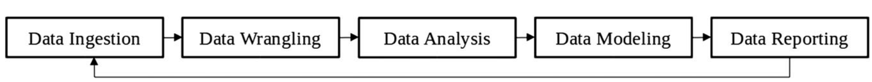

Figure 1. The data science pipeline. Adapted from Ojeda, Murphy, Bengfort, and Dasgupta (2014).

algorithmic model maintained – transparency remains a core guiding principle here (Stoyanovich et al., 2017).

Transparency of algorithms in context

The principles discussed above all attribute importance to the concept of transparency. The increased influence of algorithmic models on our daily lives has intensified this call for more transparency of the mechanics of those algorithmic models. The initial response of many authors has been to call for algorithmic transparency through the open-sourcing of algorithms (Goodman & Flaxman, 2016). However, such schemes may not necessarily have the required effect. It is possible for organizations to share all available documentation, procedures, and code, yet this will not constitute transparency if that information is not understood by the relevant audience (Heald, 2006).

As with any other technology, algorithms are embedded within existing social, political, and economic settings. To understand the impact of algorithmic models in particular, and quantification objects more generally, it is paramount to study how they become embedded in the networks of people and existing systems that make use of them, and the practices that facilitate this embedding (Espeland & Stevens, 2008). Algorithmic models cannot be separated from the practices in which they are designed, programmed, and used (Geiger, 2014). Qualitative approaches can help draw detailed and rich accounts of algorithmic model use (Kitchin, 2017).

The importance of this is illustrated by two examples. First, (Bodó et al., 2017) argue that, although it can be very challenging to separate algorithms from the data they are based on, and the decisions they inform, it is important to consider their context. Discriminatory algorithmic decisions are hardly hard-coded, and may be the emergent properties of the machine learning process, not identifiable from the review of code; full code transparency may actually aid the abuse of the algorithms by malevolent agents. In any case, there are very few algorithmic agents whose full code is available for review either by the public or by a closed group of experts. Even the most transparent companies in this domain (such as Reddit) keep parts of their code closed to avoid abuse.

Second, empirical work on the role of digital quantification objects in the financial sector has demonstrated the potential of such a practice-based approach. Social studies of finance describe the process through which people and digital quantification objects interact to arrive at 'calculative frames' (Hardie & Mackenzie, 2007) or 'market frames' (Beunza & Garud, 2007); world-views that exist as mental models and have a material manifestation in the form of text, spreadsheets and models. Millo and MacKenzie (2009) illustrate how the widespread adoption of the Black–Scholes–Merton options pricing model fundamentally impacted the dynamics of that market. In this fashion, they show that the world-view entrenched in algorithmic models can have performative capacity that may not be readily apparent to their users.

This means that, despite good intentions, guidelines or principles for fostering responsible data science will fail if they do not consider the context in which algorithms are used. In the following, we will contribute to the discussion about transparency by building on insights from the field of glitch studies, arguing for a more critical approach to the implementation of measures that increase a model's transparency – what is it that such measures truly achieve and what issues confront them?

Glitch

A brief introduction

Glitch theorists Hugh S. Manon and Daniel Temkin offer the following introduction to the scope of glitch studies:

Almost invariably, digital imagery greets its beholder in the guise of analog – as a smooth and seamless flow, rather than as discrete digital chunks. A glitch disrupts the data behind a digital representation in such a way that its simulation of analog can no longer remain covert. What otherwise would have been passively received – for instance a video feed, online photograph, or musical recording – now unexpectedly coughs up a tumorous blob of digital distortion. Whether its cause is intentional or accidental, a glitch flamboyantly undoes the communications platforms that we, as subjects of digital culture, both rely on and take for granted (Manon & Temkin, 2011, p. 1).

The field of glitch studies, in other words, emphasizes digital errors and failures in order to bring to the fore normative modes of engagement with digital technology. Departing from the Heideggerian assertion that when a tool malfunctions, one is forced to apprehend it in a different way, glitch-based artworks generally propose that errors and hiccups in a system can engender the necessary distance required for critical reflection; 'the shock comes because when we work with the machine we are contained by it. A glitch ruptures this immersive environment, undercutting the sovereignty of the digital by revealing its pervasiveness' (Manon & Temkin, 2011, p. 7).

As demonstrated, this pervasive quality of the digital has been manifested through the expansion and intensification of algorithmic mediation. As individuals, organizations and governments alike rely on algorithmic models in order to achieve tasks or to get a grip on the world, daily life is increasingly underpinned by the 'subterranean, ongoing operation of [algorithmic] assemblages which have not yet been resolved, and may never resolve; assemblages beyond human mastery, yet in which humans are implicated and entangled' (Cloninger, 2013, pp. 25–26). We can understand glitch studies as a field that is invested in on the one hand mapping and elucidating these assemblages, but on the other hand in revealing the extent to which the digital necessarily remains opaque. In the following paragraphs, we will draw on three dominant motifs within the field of glitch studies that prove particularly pertinent to the case at hand. These motifs will be supplemented with a wider discussion of algorithms and transparency[2]. In light of the call for transparency that we have been outlining, the essential question that unites the following considerations is this: what value is there to transparency when that which is rendered transparent still troubles or escapes comprehension (specialist or otherwise)?

Glitch and speed

Within the field of glitch studies, glitches are generally apprehended in terms of the deceleration of speed or the hampering of a smooth flow of information transmission; glitches

> interfere with a particular software or device to an extent that they cannot be ignored by the user (altering media aesthetics, modifying the scope of software's operations logic), but (…) do not lead to a complete failure of a system / machine understood as a tool. (Contreras-Koterbay & Mirocha, 2016, p. 97)

Glitches, in this sense, are more about delay than about total malfunction. When a glitch arises, it can impede on the functionality of a system and cause a suspension in achieving the user's intended goal. This makes glithces a privileged site for inquiring into the speed at which digital operations proceed; fluctuations in velocity rip the user from his or her immersion and thus shape a space for critical reflection (Manon & Temkin, 2011, p. 7).

Glitches, however, can also signify speed on a more overarching and conceptual level. Rosa Menkman's *The Collapse of PAL* (2010) is an audiovisual performance that employs glitches in order to raise questions about digital culture and its economy of obsolescence. The performance invokes Walter Benjamin's famous figure of the Angel of History (inspired by Paul Klee's 1920 painting *Angelus Novus*) and recasts it in a digital sheen, reflecting on the role of the historical within the digital. Through its reliance on glitches, the performance signifies the attrition of technology and its perpetual displacement by newer technologies or applications. Menkman's performance captures the 'twinning of ennui and excitement' (Chun, 2016, p. 73) that animates digital culture; a general logic that inspires both users and creators to above all desire the arrival of new versions, new updates, new innovations (Chun, 2016). This marks a digitalized echo of the economic rationale of creative destruction, 'with the emphasis on "creative" and almost no serious reflection on destruction' (Liu, 2004, p. 322). *The Collapse of PAL,* whose glitches serve as a eulogy for one of many technologies that have been rendered obsolete by the incessant mantra of digital innovation (Rhizome, 2012), helps us understand the general temporality of digital culture. What the field of glitch studies thus discloses is that there is a twofold logic of speed that underpins our engagement with digital technology; not only do these technologies achieve tasks at a rate unfathomable to human cognition, but digital culture's perpetual drive towards the new also means that whatever system we are engaged with today might be supplanted by a newer system tomorrow.

What does all this mean for the valuation of transparency? First of all, through its insistence on materiality (Halter, 2010), glitch studies demonstrate that the speed and efficiency of a digital model can be impacted by its material basis. The idea of considering the material properties of models may seem odd because we tend to think of materials as physical substances such as wood and stone. However, the rapid growth of the digital has sparked a debate on the material status of digital objects. Conflicting perspectives on this issue exist (Betancourt, 2017; Faulkner & Runde, 2013; Knoespe & Zhu, 2008; Leonardi, 2010; Suchman, 2007), but the intimate connection between immaterial digital objects and their material bearers cannot be overlooked when considering a model's functionality (Kallinikos, Aaltonen & Marton, 2013). Consider, for example, the possibility of material bearers of digital objects to crash, to be unresponsive or to encourage lock-ins. A model may be transparent, but this renders its relation to a material bearer no less precarious. As such, as glitch studies reminds us, the temporal efficiency of a model cannot be entirely decoupled from the material properties of its bearer(s).

Even more pertinent in relation to matters of transparency is the speed at which algorithmic models perform their tasks. The reasons that many of the algorithmic models under scrutiny here have been developed is that they achieve a goal at a much faster rate than would a human actor or even an entire team of professionals. The downside to this is that a (regular) critical and deliberative assessment of the model would in a sense defeat the purpose; the time required to map the entire functionality of an algorithmic model generally greatly exceeds the time such algorithmic models require to achieve their tasks. Philosopher Franco 'Bifo' Berardi has written at length about the constitutive disconnect between the performance of algorithmic models and our own cognitive faculties: 'The technical composition of the world has changed, but the modalities of cognitive appropriation and elaboration cannot adapt to this change in a linear way' (Berardi, 2015, p. 44) – the great tragedy here is that while cyberspace is infinite, 'cybertime' is not (Berardi, 2015). Algorithmic models may expand, evolve, mutate and perform their operations at ever greater speed, but our cognitive resources remain necessarily limited. Even when an organization chooses to disclose the algorithmic models they use and what data they are based on, a typical audience will not have the required time or expertise to critically assess the implications of that algorithmic model.

A further challenge surrounding the transparency of algorithmic models lies in their dynamic nature and the dynamic nature of digital culture in general. Algorithmic models are rarely static; even when they are not refined, extended, or updated, they can be reactive to their inputs (Miyazaki, 2012). Algorithmic models may be adapted as more data become available, or can be adjusted to benefit from the latest data science breakthroughs. Especially since new data science methods are developed at a high pace and because open-source implementations of these methods are released in rapid succession, there is a high turnover in terms of the algorithmic models being used. In this context, the idea of an algorithmic model as a stable object that can be submitted to critical inspection may sit uncomfortably with practice; an algorithmic model that is in use today, might not be next week. This falls in line with the critique against the fetishization of the new weighed from the perspective of glitch studies; what is the value of transparency when the model under scrutiny is already being (or has already been) refined, adapted, updated, and expanded? In sum, both the fact that the output of an algorithmic model is generated at a far greater speed than the time required to assess its operations and the fact that the contemporary technological ecosphere quickly relegates individual versions to the realm of obsolescence raise questions about the value and usefulness of transparency.

Glitch and critical audiences

The problem that faces many theorizations of glitches is that they presuppose an inherent critical quality to the glitch; the glitch is often a priori figured as emancipatory, revelatory, and empowering, wresting the user from the spell of the digital. The truth is, however, that there is no given situation in which the encounter of a glitch invariably leads to a more critical and informed engagement with the system at hand. Michael Betancourt is thus right in arguing that glitches in and of themselves do not guarantee a critical engagement – in fact, the digital's axiomatic status generally ensures that a glitch is deemed ultimately insignificant (Betancourt, 2017, p. 100). Betancourt introduces the notion of an active and critical audience into the equation, designating the requirement that a glitch not only needs to be aesthetically registered, but also deemed meaningful in order for any critical engagement to emerge. The way in which our engagement with digital technologies takes shape generally forecloses such an engagement, as minor medial failures tend at best to be attributed to the functionality of the particular device we are using and seldom inspire a more general reflection on the naturalization and ubiquitization of the digital.

We can expand Betancourt's work on glitches and active audiences in relation to transparency by on the one hand considering the notion of context-dependence and on the other hand by expanding his idea of criticality to indicate not simply the nature of an aesthetic judgment, but also the subsequent critical engagement (or lack thereof) with a glitch or a digital system. As discussed, Betancourt's work is based on the assertion that in glitch studies the role of the audience in making critical judgments is often elided. A similar logic can be discerned when it comes to the marker of transparency (and similar monikers like 'open-access' or 'open-source'). Let us offer an example taken from UK government. The 2050 Calculator, an energy and emissions model, was designed to be open-source, meaning that anyone could explore its details. The developers of the 2050 Calculator, however, noted that very few people bothered to look into the documentation. More importantly, they felt that by open-sourcing the model, people were less inclined to contest its outcomes; the model was thus more credible, but evoked less credibility work (Kolkman, in press). This is an example of how a signifier of transparency ('open-source') can incite people a priori to place their trust in a system – transparency here leads to a less critical attitude, but not necessarily to a better product.

Another example is found in the practice of peer reviewing, which is a practice through which a model commonly gets and maintains its credibility. In short, this typically entails an external organization coming in to review a model to assess its quality (the model is thus rendered transparent to the organization in question). Model professionals see the external review as a good means to assess the quality of a model and the Treasury's (Macpherson, 2013) report on government analytic models regards it as one of the most thorough means to quality assurance. In instances where there are strong connections between several models, external reviews are not without challenges, however. External reviews of a particular model are typically conducted by organizations that use a similar type of model for a similar purpose. For instance, one of the Pensim2 model's reviews[3] was conducted by the United States Congressional Budget Office, an organization that also uses a dynamic microsimulation. Because these two algorithmic models share their technical foundations, the US Congressional Budget Office was well positioned to review the Pensim2 model. At the same time, however, it seems unlikely that two parties that use the same underlying modelling paradigm would question the use of that paradigm. Again, a degree of transparency does not automatically ensure a better model.

Another problem concerns the shortage of expertise. There is a significant lack of people who know how to effectively formulate algorithms, let alone people who are willing to review their use in models (Miller, 2014; Rae & Singleton, 2015). Furthermore, even the developers of models themselves – even though the entire model is transparent to them – may not have a sound understanding of the entire model. Especially for models that are built up over the years and that are made up from many lines of code, this may present a difficulty. Developers of Pensim2, for instance, reported that they felt more familiar with some parts of that algorithmic model and less familiar with other parts. Moreover, some parts of the algorithmic model may even be completely unknown to them. Consider the following excerpt from an interview conducted with a developer of the Pensim2 model (Kolkman, in press):

> (…) How the regression equations are derived, I don't know a great deal about how they are derived. I've seen a lot of the spreadsheets, the workbooks and know how they are work ing,

> taking whatever variable and applying this parameter, but then I have no idea how those parameters were developed in the first place. I know they are the outcome of some minor logistic regression, but I don't know much more than that.

The thing to note here is that even model professionals who interact with the algorithmic model on a daily basis may not understand the model in its entirety. An algorithmic model may thus be entirely transparent, but if it is so complex or distributed that even people who work with it daily do not entirely understand it, how can we presuppose outsiders to thoroughly assess its qualities?

Just as a glitch does not automatically lead to a more critical attitude toward the digital, so too does the signifier of transparency not ensure a more critical evaluation of a model (in some cases it may even facilitate the averse, as the 2050 Calculator example shows). The qualification of transparency is in itself not enough to assure better quality; other factors (expertise, willingness, the model's relative complexity, objectivity, etcetera) determine whether such transparency will in fact prove beneficial.

Glitch, complexity, and irreducibility

In the previous paragraphs, we already briefly hinted at the complex nature of the models under scrutiny and this dovetails smoothly into the third glitch-informed dimension that problematizes the notion of transparency; the complexity and irreducibility of the algorithmic. Glitches have always been associated with the chaotic and unpredictable aspects of the computational. Good glitch art, conclude Manon and Temkin in their seminal article *Notes on Glitch* (2011), maintains 'a sense of the wilderness within the computer' (Manon & Temkin, 2011, p. 13). Similarly, Sean Cubitt argues that the glitch 'indicates another subject in the medium, the ghost in the machine, the inhuman in our communications' (2017, p. 20). Whether a glitch is ultimately deemed significant or insignificant, it still generally comes to us unexpectedly and in this sense indicates a degree of autonomy within the system or model. As such, glitches communicate the complex nature of technology, informed by procedures inscrutable to the human eye. In the case of algorithmic models, for example, specific collections of algorithms are often embedded within entire algorithmic ecologies. If access is gained, algorithmic models, as Seaver (2013) notes, are thus rarely reconstructible in a straightforward manner. Within code, algorithms are usually woven together with hundreds of other algorithms to create algorithmic systems. It is the workings of these algorithmic systems that critical inquirers are mostly interested in, not the specific algorithms (many of which are quite benign and procedural). However, disentangling these ecologies often proves nigh impossible due to their topological complexity.

Moreover, the consequences and affordances of algorithmic models cannot exhaustively be gleamed from the code behind a model, even if it is made entirely transparent. Here, Betti Marenko's recent work on glitches and the contingency endemic to the digital is informative. She departs from the work of media theorist Luciana Parisi, who identifies the algorithmic as an autonomous mode of thinking and randomness as 'the condition of *programming culture*' (Parisi, 2013, p. ix, emphasis in original). For Marenko, the glitch signifies the 'tangible, yet undesigned […] evidence of the autonomous capacities of digital matter' (Marenko, 2015, p. 112). Properties may unintentionally emerge from the actual activity of algorithms (or from the communication between different algorithmic systems), rather than having been planned and programmed; 'the contingencies revealed in the opening of spaces of possibilities, in the manifestation of an otherwise potential, in the interstices of the present, are what allows the irruption of the virtual' (Marenko, 2015, p. 113). Glitches become events that reveal the autonomous agency of the algorithmic. Crucially, in light of our discussion of transparency, a model's potentiality always exists in excess of its formulation: 'There is something within algorithmic procedures that cannot be exhausted by their formulation, no matter how complex or elegant' (Marenko, 2015, p. 115).

To conclude: while Mazzotti compellingly argues for a standard of transparency to elucidate 'the technical choice of certain expert groups' (Mazzotti, 2017), the reality of algorithms today is that even for these experts the precise operations and potential ramifications of an assemblage of algorithms remain obscure. Transparency supposes a holistic model of planned codes and instructions that captures all possibilities, but the nature of algorithmic mediation complicates such a vision. Even if an algorithmic model is made entirely transparent, not all of its potential effects and faculties can be inferred from this gesture. The complexity of the algorithmic along with its autonomous capacities necessarily entails that part of its potentiality remains closed off.

Concluding remarks

Algorithmic models have become entrenched in virtually all spheres of human activity. Despite the many advantages that these quantification objects may bring about, concerns have risen about the susceptibility of algorithmic models to human-like bias (Barocas & Selbst, 2016; Caliskan-Islam, Bryson, & Narayanan, 2016; van Doorn, 2017). This stands in contrast with the promise of algorithmic models of bringing 'reliability and objectivity to otherwise uncertain procedures' (Mazzotti, 2017). It is not surprising that the academic community has moved to discuss this issue and has moved to develop guidelines for the responsible management, analysis, and use of algorithmic models. We provided a snapshot of this discussion here to consolidate the pervasive discourse and intervene in the debate. We identify transparency as a key factor in these guidelines and criticize the concept by drawing on glitch studies and introducing the dimensions of speed, critical audiences, and complexity and irreducibility. The argument that we have developed is twofold.

First, and foremost, we have emphasized and interrogated the role of a critical audience when it comes to matters of transparency. Measures of transparency are at risk of remaining empty signifiers if no critical and unbiased engagement emerges from their installment. The key argument that this paper has thus developed is that uncoupling the value of transparency from the practical matter of how that transparency takes shape and how it is likely to be engaged with ultimately paints a limited picture. Measures toward algorithmic accountability are most effective if we consider them a property of socio-technical assemblages of people and machines. Within such assemblages, the value of transparency fundamentally depends on enlisting and maintaining critical and informed audiences.

A second point developed is that the fostering of such audiences meets its own share of issues; the discrepancy between cyberspace and cybertime (Berardi, 2015), the rapid turnout of new versions, the exponential complexity of algorithmic architectures, and the irreducible autonomy of the algorithmic are all phenomena that hamper an effective assessment of algorithms and that are not sufficiently remedied by transparency alone. We thus conclude that the call for transparency as articulated in the discussed guidelines is in itself unlikely to have the desired effect. We urgently need more empirical studies of algorithmic models used in practice; in particular, we need to assess the conditions in which measures of transparency actually yield positive effects by fostering a productive relationship with an audience, while also acknowledging the necessary limits of such a relationship. Such research can help us to test guidelines for algorithmic accountability. Only then can we begin to develop guidelines that are sound and fit within existing practices.

Notes

1. While transparency marks the focus of this paper, the other FACT principles are of course implicated in the valuation of transparency ‑ here follows a brief outline: Fairness ‑ How to avoid unfair conclusions even if they are true? Accuracy – How to answer questions with a guaranteed level of accuracy? Confidentiality – How to answer questions without revealing secrets?
2. In discussing the characteristics of the socio-technical assemblages surrounding algorithms which can impede critical inspection altogether, or prevent the development of a critical audience, we build on previous work on applications of algorithms in practice. We also draw on fieldwork that has been presented extensively elsewhere (Author, forthcoming). We studied 8 cases of algorithmic model use in government and analytic industry over a period of 2.5 years. Data were collected in the form of interviews, participant observations, and (digital) documents analysis.
3. Pensim2 has been reviewed on several occasions with varying degrees of formality. Examples include the review by the US congressional office mentioned in the text and a review by the UK Institute for Fiscal studies (Emmerson, Reed, & Shephard, 2004).

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Jakko Kemper is a PhD candidate at the Amsterdam School for Cultural Analysis (University of Amsterdam). His research focuses on digital culture, the aesthetics of imperfection, and media theory [email: j.kemper@uva.nl].

Daan Kolkman is a research fellow in decision-making at the Jheronimus Academy of Data Science. His research revolves around the sociology of quantification, intelligent decision support systems (e.g. machine learning, artificial intelligence), and organizational decision-making. Daan received his PhD in sociology from the University of Surrey (England) for his work on computational models in government. He was supervised by Nigel Gilbert, Tina Balke, and Paolo Campo at the Centre for Research in Social Simulation. Daan develops data-intensive products and services for – and in collaboration with – retailers and SMEs [email: d.kolkman@tue.nl].

References

Aalst, W. M. P., Bichler, M., & Heinzl, A. (2017). Responsible data science. *Business & Infor-Mation Systems Engineering*, 2–4. doi:10.1007/s12599-017-0487-z

Angwin, J. (2016). Make algorithms accountable. *The New York Times*, 1.

Annany, M., & Crawford, K. (2016). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, *20*(3), 973–989.
Barocas, S., & Selbst, A. (2016). Big data's disparate impact. *California Law Review*, *104*(1), 671–729. doi:10.15779/Z38BG31
Berardi, F. (2015). *And: Phenomenology of the end*. South Pasadena, CA: Semiotext(e).
Betancourt, M. (2017). *Glitch Art in theory and practice: Critical failures and post-digital Aes-thetics*. New York, NY: Routledge.
Beunza, D., & Garud, R. (2007). Calculators, lemmings or frame-makers? The intermediary role of securities analysts. In M. Callon, Y. Millo, & F. Muniesa (Eds.), *Market devices* (pp. 13–39). Oxford: Blackwell. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1467-954X.2007.00728.x/full
Bodó, B., Helberger, N., Irion, K., Borgesius Zuiderveen, F. J., Moller, J., van der Velde, B., & Vreese, C. H. d. (2017). Tackling the algorithmic control crisis – the technical, legal, and ethical challenges of research into algorithmic agents. *Yale Journal of Law & Technology*, *19*, 133.
Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, *356*(6334), 183–186. Retrieved from http://science.sciencemag.org/content/356/6334/183.abstract
Caliskan-Islam, A., Bryson, J. J., & Narayanan, A. (2016). *Semantics derived automatically from language corpora necessarily contain human biases*. arXiv:1608.07187v2 [cs.AI] 30 Aug 2016, 6334, 1–14. doi:10.1126/science.aal4230
Chen, H., Chiang, R., & Storey, V. (2012). Business intelligence and analytics: From Big data to Big impact. *MIS Quarterly*, *36*(4), 1165–1188. Retrieved from https://noppa.aalto.fi/noppa/kurssi/37e44000/materiaali/37E44000_3_generations_of_business_analytics_overview.pdf
Chun, W. H. K. (2016). *Updating to remain the same: Habitual New media*. Cambridge, MA: MIT press.
Cloninger, C. (2013). Manifesto for a theory of the 'New aesthetic'. In J. B. Slater & A. Iles (Eds.), *Slave to the algorithm.* (pp. 16–27). London: Mute.
Contreras-Koterbay, S., & Mirocha, Ł. (2016). *The new aesthetic and art: Constellations of the post-digital*. Amsterdam: Institute of Network Cultures.
Council of Europe. (2017). *Guidelines on the protection of individuals with regard to the processing of personal data in a world of Big Data*.
Cubitt, S. (2017). Glitch. *Cultural Politics*, *13*(1), 19–33. doi:10.1215/17432197-3755156
Dekker, C. M., Groenendijk, A., Sliggers, C. J., & Verboom, G. K. (1990). *Quality criteria for models to calculate air pollution*. Lucht (Air) 90. Leidschendam: Ministry of Housing, Physical Planning and Environment.
Department for Transport. (2012). *Report of the Laidlaw inquiry*.
van Doorn, N. (2017). Platform labor: On the gendered and racialized exploitation of low-income service work in the 'on-demand' economy. *Information, Communication & Society*, *20*(6), 898–914. doi:10.1080/1369118X.2017.1294194
Emmerson, C., Reed, H., & Shephard, A. (2004). *An assessment of PenSim2* (No. 04/21). IFS Working papers. Institute for Fiscal Studies (IFS).
Espeland, W. N., & Stevens, M. L. (2008). A sociology of quantification. *European Journal of Sociology*, *49*(3), 401. doi:10.1017/S0003975609000150
Faulkner, P., & Jochen, R. (2013). Technological objects, social positions, and the transformational model of social activity. *Mis Quarterly*, *37*, 803–818.
Floridi, L., & Taddeo, M. (2016). What is data ethics? *Philosophical Transactions of the Royal So-Ciety A: Mathematical, Physical and Engineering Sciences*, *374*, 2083. Retrieved from http://rsta.royalsocietypublishing.org/content/374/2083/20160360.abstract
Frigg, R., & Hartmann, S. (2012). *Models in Science*. The Stanford encyclopedia of philosophy. Retrieved from http://plato.stanford.edu/archives/fall2012/entries/models-science/
Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137–144. doi:10.1016/j.ijinfomgt.2014.10.007

Geiger, R. S. (2014). Bots, bespoke, code and the materiality of software platforms. *Information, Communication & Society*, *17*(3), 342–356. doi:10.1080/1369118X.2013.873069

Goodman, B., & Flaxman, S. (2016). *EU regulations on algorithmic decision-making and a "right to explanation."* 2016 ICML Workshop on Human Interpretability in Machine Learning (WHI 2016), (Whi), 26–30. Retrieved from http://arxiv.org/abs/1606.08813

Gross, D., & Strand, R. (2000). Can agent-based models assist decisions on large-scale practical problems? *A Philosophical Analysis: Complexity*, *5*(6), 26–33. doi:10.1002/1099-0526(200007/08)5:6≤26::AID-CPLX6≥3.0.CO;2-G

Haag, D., & Kaupenjohann, M. (2001). Parameters, prediction, post-normal science and the precautionary principle – a roadmap for modelling for decision-making. *Ecological Modelling*, *144*(1), 45–60. doi:10.1016/S0304-3800(01)00361-1

Halter, E. (2010, November). *The matter of electronics*. Retrieved from http://heavysideindustries.com/wp-content/uploads/2010/11/THE-MATTER-OF-ELECTRONICS-by-Ed-Halter.pdf

Hardie, I., & Mackenzie, D. (2007). Constructing the market frame: Distributed cognition and distributed framing in financial markets. *New Political Economy*, *12*(3), 389–403. doi:10.1080/13563460701485649

Heald, D. (2006). Varieties of transparency. *Proceedings of the British academy*, 135, 25–43. doi:10.5871/bacad/9780197263839.003.0002

Hildebrandt, M. (2006). Profiling: From data to knowledge. *Datenschutz Und Datensicherheit-DuD*, *30*(9), 548–552.

Hildebrandt, M. (2012). The Dawn of a critical transparency right for the profiling era the dawn of a critical transparency right for the profiling era. *Digital Enlightenment Yearbook 2012*, *12*(2008), 40–56.

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information Systems Management*, *29*(4), 258–268. doi:10.1080/10580530.2012.716740

Janssen, M., & Kuk, G. (2016). The challenges and limits of big data algorithms in technocratic governance. *Government Information Quarterly*, *33*(3), 371–377. doi:10.1016/j.giq.2016.08.011

Kallinikos, J., Aaltonen, A., & Marton, A. (2013). The ambivalent ontology of digital artifacts. *Mis Quarterly*, *37*(2), 357–370.

Kitchin, R. (2017). Thinking critically about and researching algorithms. *Information, Communication & Society*, *20*(1), 14–29. doi:10.1080/1369118X.2016.1154087

Knoespe, K. J., & Zhu, J. (2008). Continuous materiality: Through a hierarchy of computational codes. *Théorie, Littérature, Epistémologie*, *25*, 235–247.

Kolkman, D. (in press). The (in)credibility of data science methods to non-experts. *Science Technology Studies*.

Lathrop, D., & Ruma, L. (2010). *Open government: Collaboration, transparency, and participation in practice*. Sebastopol, CA: O'Reilly Media.

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: Traps in Big Data Analysis David. *Policy Forum*, *343*(March), 1203–1205.

Leonardi, P. (2010). Digital materiality? How artifacts without matter, matter. *First Monday*, *15*(6). doi:10.5210/fm.v15i6.3036

Lepri, B., Oliver, N., Letouzé, E., Pentland, A., & Vinck, P. (2017). Fair, transparent, and accountable algorithmic decision-making processes the premise, the proposed solutions, and the open challenges. *Philosophy and Technology*. doi:10.1007/s13347-017-0279-x

Liu, A. (2004). *The laws of cool: Knowledge work and the culture of information*. Chicago, IL: University of Chicago Press.

Macpherson, N. (2013). *Review of quality assurance of government models*. London: HM Treasury.

Manon, H., & Temkin, D. (2011). Notes on Glitch. *World Picture*, *6*, 1–15.

Marenko, B. (2015). When making becomes divination: Uncertainty and contingency in computational glitch-events. *Design Studies*, *41*, 110–125. doi:10.1016/j.destud.2015.08.004

Mazzotti, M. (2017). *Algorithmic life*. Retrieved from https://lareviewofbooks.org/article/algorithmic-life/

Menkman, R. (2010, June 1). *The collapse of PAL*. Retrieved from https://vimeo.com/12199201

Miller, S. (2014). Collaborative approaches needed to close the big data skills gap. *Journal of Organization Design*, *3*. doi:10.7146/jod.9823

Miller, H. G., & Mork, P. (2013). From data to decisions: A value chain for big data. *IT Professional*, *15*(1), 57–59.

Millo, Y., & MacKenzie, D. (2009). The usefulness of inaccurate models: Financial risk management" in the wild. *Journal of Risk Model Validation*, *3*(1), 23–49. Retrieved from http://eprints.lse.ac.uk/id/eprint/28966

Minsky, M. L. (1965). Matter, mind, and models. *Proceedings of International Federation of Information Processing Congress* 1, 45–49.

Miyazaki, S. (2012). Algorhythmics: Understanding micro-temporality in computational cultures. *Computational Culture*, *2*. Retrieved from http://computationalculture.net/algorhythmics-understanding-micro-temporality-in-computational-cultures/

National Institute of Food and Agriculture. (2017). *National Institute of Food and Agriculture* (pp. 1–2). Data science in agriculture summit. Retrieved from http://nifa.usda.gov/extension

Ojeda, T., Murphy, S. P., Bengfort, B., & Dasgupta, A. (2014). *Practical data science cookbook*. doi:10.5326/50.5.toc

Parisi, L. (2013). *Contagious architecture: Computation, aesthetics, and space*. Cambridge, MA: MIT Press.

Pasquale, F. (2015). *The black Box society*. Cambridge, MA: Harvard University Press.

Peng, G., Ritchey, N. A., Casey, K. S., Kearns, E. J., Privette, J. L., Saunders, D., & Ansari, S. (2016). Scientific stewardship in the open data and big data era – roles and responsibilities of stewards and other major product stakeholders. *D-Lib Magazine*, *22*(5-6), 1–1. doi:10.1045/may2016-peng

Pielke, R. A. (1999). *Who decides? Forecasts and Responsibilities in the 1997 Red River Flood. Applied Behavioral Science Review*, *7*(2), 83–101.

Rae, A., & Singleton, A. (2015). Putting big data in its place: A regional studies and regional science perspective. *Regional Studies, Regional Science*, *2*. doi:10.1080/21681376.2014.990678

Responsible Data Science Initiative. (2016). Responsible Data Science Initiative.

Rhizome. (2012, May 31). Rhizome The Collapse of PAL. Retrieved from http://classic.rhizome.org/portfolios/artwork/54452/

Rodríguez-Iglesias, A., Rodríguez-González, A., Irvine, A. G., Sesma, A., Urban, M., Hammond-Kosack, K. E., & Wilkinson, M. D. (2016). Publishing FAIR data: An exemplar methodology utilizing PHI-base. *Frontiers in Plant Science*, *7*, 641. doi:10.3389/fpls.2016.00641

Sagiroglu, S., & Sinanc, D. (2013, May). *Big data: A review*. International conference on collaboration technologies and systems (CTS), San Diego, CA (pp. 42–47). IEEE.

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). *Auditing algorithms: Research methods for detecting discrimination on internet platforms*. Paper presented to Data and Discrimination: Converting Critical Concerns into Productive Inquiry, Seattle, WA.

Seaver, N. (2013). Knowing algorithms. *Media in Transition*, *8*, 1–12.

Stone, H. S. (1971). *Introduction to computer organization and data structures*. New York: McGraw-Hill.

Stoyanovich, J., Howe, B., Abiteboul, S., Miklau, G., Sahuguet, A., & Weikum, G. (2017). *Fides: Towards a platform for responsible data science*. International Conference on Scientific and Statistical Database Management, Chicago.

Suchman, L. (2007). *Human-machine reconfigurations: Plans and situated actions*. Cambridge: Cambridge University Press.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.

Wilkinson, M. D. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, *3*. doi:10.1038/sdata.2016.18

Wilkinson, M. D., Verborgh, R., Bonino da Silva Santos, L. O., Clark, T., Swertz, M. A., Kelpin, F. D. L., … Dumontier, M. (2017). Interoperability and FAIRness through a novel combination of Web technologies. *PeerJ Computer Science*, *3*. doi:10.7287/peerj.preprints.2522v1