HUMAN RIGHTS IN THE AGE OF PLATFORMS

EDITED BY RIKKE FRANK JØRGENSEN
FOREWORD BY DAVID KAYE

3 Data as Humans: Representation, Accountability, and Equality in Big Data

Anja Bechmann

Introduction

This chapter examines the democratic implications of how we treat data as humans in the datafied society, and how we process such data through machine learning algorithms. Democracy as a concept has a long history in political and social sciences. The focus of the chapter will be limited to the connection between the processing of *data as humans* in learning algorithms and the democratic values of representation (including participation), accountability, and equality.[1] In line with existing studies, I illustrate how systematic discrimination and inequality may occur through machine learning if we do not take the preliminary measure of inscribing these democratic values in the machine learning algorithms executed by, for instance, social media platforms. Moreover, I argue that free and open communication is an ideal that we must strive for if we wish to avoid democratic deficits. The chapter theorizes on whom data represents, what we (as society) do with data, and how we govern these practices. I argue that while some problems of representation in the datafied society are not new, problems of discrimination may now happen in a more systematic manner without yet receiving the same regulatory impact.

In the pursuit of as many different data points as possible, technology companies develop products that intersect and datafy every aspect of human existence from self-reports (social media) to location data (self-driving cars and maps) and biometrics (health apps, exercise wearables, and biojewelry). These data traces are increasingly used to inform product and processual decisions by companies that want to "listen" to the user and optimize recommendations, products, and revenue accordingly or by politicians and governments that want to "adjust" behavior using large data streams and big data methods. One example would be the recent Cambridge Analytica controversy and the use of social media data and microtargeting campaigns during the 2016 election of Donald Trump as US president that came as a shock to many citizens, including the press. Many blamed social media, because these platforms insisted on preserving the algorithm that provides users primarily with content that confirmed their social and political adherences (Bakshy, Messing, and Adamic 2015) and at the same time allowed for third parties to implement microtargeted campaigns.

It is important for social media companies to keep users on the platform for as long as possible in order to increase advertising profits through monetizing data (see Bermejo, this volume), and one way to do this is to display content that users agree with (Bakshy, Messing, and Adamic 2015). A platform like Facebook represents a large public forum for reading, viewing, listening to, and participating in discussions; however, the company is registered as a technology company, not a media company with press responsibilities. On Facebook, we do not know the principles and values (as opposed to journalistic values as outlined in International Federation of Journalists 1986) behind the editing done by the algorithm and are unable to see each other's personalized news feeds. This is radically different from traditional editorials, where we can discuss the principles, and printed papers where we were/are able to purchase or subscribe to full papers with different viewpoints.

The chapter seeks to discuss such problems not only on the algorithmic level but also on the level of data, which plays an increasingly important role in everyday life (Schäfer and van Es 2017). The data that we leave behind when we use online platforms are central elements of the global online economy and a defining and pertinent characteristic of a citizen in the digital age, as is the processing of such data. Critical voices question whether informed consent is an option anymore, as it is impossible for companies to provide a comprehensible account of the vast places data is being used (Bechmann 2014; Nissenbaum 2011). Despite the right to access and transport one's own data (Regulation [EU] 2016/679 of the European Parliament and of the Council L 119/1), it is debatable how this right should be executed and controlled. Opting out of the datafied society is no longer an option. Even so, exclusion from data-enriched decisions may also have profound consequences for the equal representation of individuals in society (Ananny 2016). In response to these challenges, my questions are these:

How does democratic society strive to ensure that all humans are properly and equally represented—that is, that data traces actually represent the user and that all users are part of the data processing on equal terms? And how does it ensure that algorithmic decisions are transparent and reliable?

These questions cannot be answered easily, but they must be addressed; therefore, they drive the discussion in this chapter. Theoretically, the chapter draws on pragmatism (Dewey 1927) and cyberfeminist theory (Bowker and Star 1999; Haraway 1991; Star 1990) to account for the meaning of underrepresentation, unaccountability, discrimination, and inequality as constituting democratic deficits in the use of big social data and machine learning. The chapter will draw on previous empirical work carried out primarily with social media data, focusing on Facebook data as one of the most well-known sources of data enrichment and the use of big data methods such as cluster analysis, neural networks, and deep learning to account for usage patterns as a source of insight into human behavior and preferences. The purpose of the chapter is to provide critical insights into the consequences of (a lack of) data quality in machine learning processing.

Representation and Participation as Democratic Values

Let us begin by examining social media platforms and how such platforms themselves, and the datafied society at large (third parties), use social media as a data source for prediction through machine learning. In considering democracy in the datafied society, I will turn to the encounter between Lippmann and Dewey in the 1920s in which Lippmann criticizes the public as "the phantom public" (Lippmann 1927) and the ability for citizens to represent public opinion based on democratic values. In brief, Lippmann (ibid.) sees experts, facts, and science as the solution to the problem of the public and the sustainability of democratic values. Dewey (Dewey 1927; Bybee 1999) recognizes the problem but disagrees on the solution. Instead of relying on experts, facts, and science, he argues that democracy is created, situated, and negotiated through the agency of citizens and their participation in the construction of democracy, thereby empowering both individuals *and* the social group:

> [Democracy] consists in having a responsible share according to capacity in forming and directing the activities of the groups to which one belongs and in participating according to need in the values which the groups sustain. From the

> standpoint of the groups, it demands liberation of the potentialities of members of a group in harmony with the interests and goods which are common. Since every individual is a member of many groups, this specification cannot be fulfilled except when different groups interact flexibly and fully in connection with other groups . . . there is a free give-and-take: fullness of integrated personality is therefore possible of achievement, since the pulls and responses of different groups reinforce one another and their values accord. (Dewey 1927, 148)

In this sense, Dewey wants to restore agency to the users or citizens as a way of negotiating values and meanings in smaller or larger groups.[2] He does not consider democracy to be something that relates solely to politics and the public sphere but rather as a basic social construct in groups. Democracy happens in both the public and private spheres and does not connect only to public opinion. He also argues for an epistemological politics of "By what right do we act?" instead of "What are the facts?" (Dewey 1927, 69). Open and free communication plays a central role in this value creation, enabling "a public to act as a public" (Dewey 1927, 55) and to judge how actions influence shared interests. Communication thereby creates the "very meaning that will be called knowledge" (ibid.).

Lippmann and Dewey's debate on democratic values and the public is relevant when it comes to user participation in social media today (posts, comments, likes, and shares) and the way algorithms control how communication is handled—whether on social media sites themselves or for data collected from them for predictions in other domains such as targeting and manipulation in political campaigns, risk assessments in financial sectors, or diagnoses and treatments in the health sector. On social media, individuals are represented through their data and connected and processed through algorithms. Users develop networks of, and memberships in, several groups and communicate on broad or narrow topics of interest, with a broader or narrower group of people, and with strong or weak ties (Bechmann, Kim, and Søgaard 2016). In many ways, social media is the ideal participation platform in Dewey's terminology, as it allows people to participate in debates across spheres. However, the transparency of who we are talking to and sharing behavioral data with and the overlap between groups have been the subject of extensive critical analysis (Bakshy, Messing, and Adamic 2015; Marwick and boyd 2014; Stutzman, Gross, and Acquisti 2012). This topic has received renewed interest in the light of recent cases and events such as Cambridge Analytica, Brexit, and the election of Trump as US president, raising the issues of informed consent (Bechmann 2014), free and open communication, and the need for different groups to meet as the basis for participatory democracy in Dewey's sense. This debate reinforces the necessity of considering how data is constructed as a representation of the individual and the way algorithms encourage exchanges across groups *with different interests*.

Equality and Accountability as Democratic Values

Dewey's focus is not primarily on whether all individuals have the same premises for participating in the creation of knowledge and democratic values and how free and open communication can be accounted for. This is a key concern of cyberfeminist theory as set out in Haraway (1991) and Star (1990). Haraway's theory of the integration of technology and humans generates an interesting perspective: data not as something "apart" or alienated from the individual but as an equal part of humans just like the body. Data cannot then be rejected as something alienating or "out there" (see also Mai's discussion on personal information in this volume). Users may experience data as something bad or something that has been violated; the sense of "embodiment" contained in this feeling is striking in earlier studies on, for instance, cyber rape (Turkle 1995). Thus, from an algorithmic or developer's point of view, data cannot be treated as something that is *nonhuman*—if we view it from the perspective of Haraway's cyborg discussion, it is indeed an integrated part of the human being. There is no mother-and-child or host-and-guest relationship, nor any extension of the body as described in McLuhan (1964). Haraway's point is that developing a perspective on data similar to that set out in McLuhan's medium theory allows for alienation and a critical discussion of technology and data as something potentially harmful that can turn against humans. On the other hand, if we do not hierarchize the relationship, then we are already technology, data is already us.

Still, just as earlier cases of census and statistical data in aggregate show (Anderson 2015; Desrosières 2002), data does not equally represent all humans. The difference between traditional survey data and social media data is that in social media settings some humans create more data points than others. Star (1990) has a strong focus on the underrepresented in specific socio-technical networks. Building on Law (1990), she also argues for technology as an arena for modulation, tacit power relations, interests, and conflicts. I will argue that questions of inclusion and exclusion (Kroll et al. 2017; Law 1990) also become relevant in the discussion on democracy, social big data, and machine learning processing. Star (Bowker and Star 1999; Star 1990) has a strong focus on the outliers—the underrepresented or “monsters” (Law 1990) that give meaning to the normal (Crawford and Calo 2016; Metcalf and Crawford 2016). To her, underrepresentation and abnormality can take many forms, from sexuality and gender to being allergic. What they all have in common is that such individuals do not decide on the shared knowledge or meaning that binds the socio-technical network together in the manner suggested by Dewey. Unable to act, they are nonetheless important as a nonagent and as a confirmation of the rules for inclusion and exclusion. In the next section I will exemplify how I see such underrepresentation encoded into the algorithmic processing of social big data and how this may subsequently lead to discrimination against protected classes (Charter of Fundamental Rights of the European Union, Article 21).

Bowker and Star (1999) suggest that by accounting for the different levels of exclusion, we are able to understand how the socio-technical is political by nature. Accountability then also becomes an interesting aspect in terms of democracy. Although Star does not explicitly discuss democracy, I will combine her proposals with those of Dewey to argue that accountability of underrepresentation and inequalities (Calo 2017; Crawford and Calo 2016; Kroll et al. 2017) in the socio-technical is part of the transparency of participation processes that Dewey considers to be the core of democracy. I will argue that the accountability of data input and machine learning processing workflows rather than a focus on the transparency of the algorithm itself (Ananny and Crawford 2018) is essential if we are to maintain an understanding of inclusion and exclusion rules *as well as* transparency in data processing. Still, this accountability does not solve the participation dilemma generated by social media platforms—thus the difficulty for citizens of both participating *and* resisting datafication. This dilemma shows that Dewey-inspired participation in a datafied society may conflict with the right to privacy (see also van Hoboken’s analysis of datafication and privacy in chapter 10 of this volume).

The Politics of Algorithms

Machine learning is built into social media algorithms as the backbone of the service, and by third-party companies as a way to interpret user behavior and preferences. Algorithms are programs that control the logic and presentation of digital platforms and services, the specific recipe behind any computational decision:

> Algorithms are now a communication technology; like broadcasting and publishing technologies, they are now "the scientific instruments of a society at large," (Gitelman 2006, 5) and are caught up in and are influencing the ways in which we ratify knowledge for civic life, but in ways that are more "protocological" (Galloway 2004), i.e. organized computationally, than any medium before. (Gillespie 2014, 169)

This chapter builds on the basic argument that the algorithms we use in data-driven decision-making are not objective tools that simply compute data. They are highly error prone, interpretive, and in need of adjustments to perform optimally, and in that sense, they are political and normative in nature. This argument is present in many existing critical contributions on algorithms within communication and media studies (Ananny and Crawford 2018; Bodle 2015; Bucher 2012; Cheney-Lippold 2011; Kitchin 2014; Leese 2014; Pasquale 2015; Turow 2011).[3] These studies focus on algorithms as a cultural phenomenon with unintended consequences for society. One such consequence relates to the exploitation of user data, the commodification of personal data, and resulting challenges concerning the right to privacy (Cheney-Lippold 2011; Leese 2014; Solove 2004; Turow 2011).

Other critical accounts focus on surveillance mechanisms, where the discourse on algorithms does not relate primarily to how such computational processes violate individual privacy but rather, on the basis of Foucault (1977), considers how they function as power tools for centralized entities such as a state or a government to control, adjust, and impose certain values upon the potentially surveilled citizen (Introna and Wood 2004; Lyon 2007).

Another array of studies show how algorithms have consequences for what is presented to us as relevant information and communication. Algorithms as filters are discussed in terms of filter bubbles, echo chambers, and digital divides (Bodle 2015; Bucher 2012; Introna and Nissenbaum 2000; Mager 2012; Rogers 2009). These bodies of literature focus on the democratic values of freedom of expression and social cohesion as threatened in a personalized online space controlled by a small number of powerful gatekeepers such as Google, Facebook, Microsoft, Amazon, and Apple.

Taken together, these critical algorithmic studies offer important insights into the societal consequences of algorithms and, indirectly, into how basic human rights and democratic values can potentially be violated through algorithms in a manner that is subtle, sometimes undeliberate, yet highly effective. A fairly new interdisciplinary approach to critical algorithmic studies consists of "audit" (Sandvig et al. 2014), "decipher" (Rieder 2005), or reverse engineering studies that seek to discuss the communicative consequences of algorithmic processing through a close analysis of the actual structures and logics of specific statistical models or algorithmic constructions (Diakopoulos and Koliska 2017; Mackenzie 2015; Rieder 2017). These studies show that algorithms encode certain types of values (Mackenzie 2015) in the way they classify (Rieder 2017), cluster, or sort the data for a certain purpose without the developers knowing exactly how the statistical model or algorithm leads to a particular optimal pattern or outcome predictor. I will supplement this interesting body of literature with my own empirical work to provide examples of how underrepresentation and inequality can occur in practice.

Big Social Data as Population and Census Data

All machine learning processing is conducted with the aim of recognizing patterns in order to predict outputs, which for instance can then be used for persecution, credit scores and subsequent insurance and loan offerings, health care, and propaganda. As algorithms become widely used to structure our culture and democracy, it is crucial for society, in an interdisciplinary manner, to illustrate errors and interpretative spaces and to inscribe the "human" in standardized processes carried out to execute decisions fast and seamlessly. This is of particular importance when algorithmic decision-making moves from product and service optimization and marketing into the realm of governmental data and when such data is paired with social media data. Algorithmic decisions made here not only affect the media "bubbles" we live in, the people we engage with or exclude in consumer society (Bauman 2000)—they also affect our health care, educational opportunities, and probability of being a political target (Noyes 2015). In this section, I will focus on Facebook as an arena for the creation of big social data that is often used for data enrichment to understand and predict user behavior both on and outside Facebook.

I will then examine more closely the critical algorithmic theories on the processes at work in machine learning models and provide examples from my own experience applying such models to Facebook data in order to illustrate the interpretive spaces and politics of such algorithms. In this way, I will focus on both the data layer and the model layer in algorithmic processing (Diakopoulos and Koliska 2017).

Facebook data is an example of an overwhelming pool of data that developers could access and use for various purposes, ranging from systematic surveillance to recruiting, political campaigning, and service optimization.[4] Currently, Facebook is globally only surpassed by Reddit as the online platform people spend the most time on (alexa.com), and Facebook data contains a wealth of different data points from self-reported demographics, interests (likes and shares), and personal accounts and opinions (status updates, photos, links, and comments) to network and behavioral data (visit to external sites with Facebook plug-ins). The overwhelming amount of data both vertically and horizontally (over time) often lead to *data rush*—overly enthusiastic and bold uses of the data as an example of human behavior and opinions worldwide. Often, people are portrayed solely in terms of data for the purposes of predictions and subsequent decisions. This has fundamental consequences for the representation of individuals in decisions based on those data. For example, a study of the total amount of private status updates, shared links, and photos among 1,000 Danes over a period of eight years (Bechmann, 2019 in press) shows that the number of data traces created varies greatly when broken down into demographics, especially age. This means that when one is using such data—for example, in connection with the provision of public services —those who only read or listen are underrepresented and excluded from the data set that informs decision-making. Furthermore, although Facebook has a large penetration rate in many countries—often higher than Twitter, Snapchat, Instagram, and noninternational platforms—the data fails to represent those who are not active users of these platforms. Active users are here understood as those who leave data traces behind to be processed by algorithms (Hargittai and Walejko 2008). Thus, when using social big data as an input in machine learning processing to represent populations, the data pool may be extremely large but the sample bias is also significant. Often people "game" the Facebook advertising algorithm (Bechmann 2015; Marwick and boyd 2014) by deliberately reporting fake demographic parameters, rendering the data quality weaker and, potentially, the predictions made from such data false if the calculations are not enriched with other data.

Sample bias is nothing new and as a concept has existed throughout the history of the political and social sciences (Desrosières 2002). Critics would argue that representation is always a problem when we have to reduce populations via aggregation and work with census data (Anderson 2015), so what is new? There are at least two differences here. Social media data is produced in private domains with limited transparency obligations. When enriching public data with, for instance, social media data to create a more detailed understanding of personal behaviors and preferences, it is difficult to account for sample biases in detail and, consequently, for how the data sets used for predictions are effectively balanced. Furthermore, the data functions not only as a one-step analytical phase but also as training data for machine learning algorithms. This training data is often not provided from the same data pool. These data steps therefore obscure even further the results of the analysis and research phase before decision-making takes place. Cambridge Analytica is a good example of this, where people's data was used to train a model to find the most predictive Facebook behaviors and attributes (e.g., like profiles) for a certain psychological profiles that again allowed the company to target specific voters with carefully tailored content.

Machine Learning and Training Data

Data in machine learning processing therefore becomes an issue not only in terms of the quality of data input itself but also in terms of its suitability for training the algorithms to recognize patterns and clusters and to create classifications. The more data and the more diversified training data you have, the better your algorithm potentially is at recognizing new data. The algorithm can only interpret data and predict patterns from the data that it has already seen (training data). However, studies and incidents have shown that training data is biased historically, culturally, and contextually. Google, for instance, labeled black people as gorillas (Cohen 2019), and a study has shown how women were described with discriminatory words due to the historical role of women as, for instance, housewives (Bolukbasi et al. 2016). To generate enough training data for the algorithm to recognize patterns and connect those patterns to certain labels, researchers and developers often use data that spans a wide historical period. In such cases, training data creates a preservative construct of associated meanings and words with key concepts that may, for instance, enforce a conservative cultural understanding of the role of women. While societal values and interpretations are in general moving toward a more inclusive and diverse society and the nondiscrimination principle operates with protected classes, decisions and predictions automated through machine learning may reinforce historical biases. This, in turn, pulls societal values in the opposite direction from inclusiveness and diversity. Still, one could claim that the data is sound proof of Dewey's participatory democracy in the sense that the data is a result of what people do with data in a particular domain or context, not what they ought to be doing according to democratic values of representation, accountability, and equality.

Similar problems arise when training data is used in, for instance, picture recognition and classification algorithms through deep convolutional neural networks. Here, training data is also the most important factor in high performance. Such algorithms are usually trained on what is available, which often means large picture databases such as ImageNet with its 1,000 classes of pictures (e.g., dogs, trees, flowers) as the potential outcome of the algorithmic processing (the last layer in the network). To Internet industries such as IBM (Watson), Alphabet (Google), Amazon, Facebook, Tesla, and Microsoft, having a large and diverse pool of annotated training data becomes a lucrative business that potentially puts their algorithms and products at the forefront of the machine learning field. However, using algorithms trained on certain types of data may lead to decisions based on a false interpretation of the data source. If a picture classification algorithm is trained on a data set containing various types of annotated data (human as well as animal), the interpretation of human faces may resemble that of animal faces and thereby lead to false, and in terms of inclusion deeply problematic, classifications and decisions. In our own research lab, we tried to use vision convolutional neural networks pretrained on ImageNet to classify social media pictures (Bechmann 2017). The performance was very poor when compared to manual annotation and may indicate that contextual sensitivity in the training data is essential for the performance of the algorithms and consequently as the basis for sustainable predictions and actions. In our case sensitivity toward social media features such as the importance of text in pictures and image focus led to false classifications of the picture from a sociological perspective. Also, the interpretation of clothes generated false output categories because the training data presumably contained health care–related pictures in which staff wore uniforms that resemble white leisure wear; thus the model tried to interpret leisure clothes as business attire and provided a completely wrong prediction of the picture.

So how can we create context-sensitive annotated training data? Natural language processing researchers Derczynski, Bontcheva, and Roberts (2016) suggest that if the right type of data is available as training data, social media data is best annotated by a combination of experts/researchers and a diverse crowd of social media users. They also show how crowd training and continual performance measurements are a clear feature of social media annotations. Furthermore, they discuss how the reduction of output categories to between seven and ten is important for automatic clustering. However, despite higher performance with reduced multidimensional complexity in output categories, this reduction of complexity can be problematic for representation and democratic actions. Although a small number of computationally isolated categories do enable higher performance, the reduction in itself may give a false picture of data as humans. For instance, Facebook operates with six different categories of "like" in the data structure in order to understand the emotional reactions toward a post, despite the psychological field suggesting eight categories and twenty-four associated emotional dimensions (Plutchik 2001). The reduction of multidimensionality in data processing may simply lead to measurement errors and false conclusions, predictions, and decisions in the use of data as humans.

Mathematical Models, Abnormality and Outliers

In the model layer (Diakopoulos and Koliska 2017), working with statistical learning models in machine learning and related interpretations and decisions may also lead to underrepresentation and inequality, as with the selections previously considered in the data layer. Rieder (2017) shows how Bayes classifiers, the most widespread prediction model within big data, produce "a basis for decision-making that is not a clear-cut formula, but an adaptive statistical model containing potentially hundreds of thousands of variables" (Rieder 2017, 110). Accountability for the calculation of such variables may be very complicated. This was also true for the use of learning models (e.g., topic models and convolutional neural networks) tested at my research lab (datalab.au.dk). In our experience with topic models, it can be very time-consuming, if at all possible, to understand the statistical logic of clusters and the implications of such logic for clusters. In other words, why do these particular clusters result from the machine learning processing? The logic appears on a linguistic level (Jurafsky and Martin 2008; Manning and Schütze 1999) that does not necessarily relate to human field-specific interpretations. This discrepancy potentially distorts the actions carried out on the basis of the clusters found. The choices made in the preprocessing of data can create very different cluster predictions depending on what kind of words are included or omitted from the data set. Omitting words that have a tendency to occur in connection with certain groups of people or minorities in order to normalize data and create "meaningful" clusters reduces their representation and "voices" in the final actions, just as omitting data from abnormal users would in classical social and behavioral sciences. Another example is that the developer's choice of the optimal number of topics or distance between the different clusters significantly influences the visibility of less normal behavior or "monsters" (Law 1990) in the data. A larger distance creates less sensitivity toward diversity, whereas a smaller distance will potentially provide sensitivity to differences and diversity in the data. Setting the optimal numbers of clusters or distance requires closer examination and judgment based on data explorations, as we sought to ensure when balancing redundancy against diversity (Bechmann, Kim, and Søgaard 2016). On the other hand, microsegmentation (Bechmann, Bilgrav-Nielsen, and Jensen 2016) creates potential challenges to privacy, as abnormal or deviant usages and users *light up* in models with a large number of clusters. Such visibility can be used against users' interests in risk assessments and behavioral adjustments by authoritarian regimes and other oppressive entities.

In our work with convolutional neural networks on pictures it is equally difficult, if not more difficult, to account for meaningful subclusters created in the various layers of the network, even though researchers are able to account for the mathematical logic in different machine learning models (Davenport and Harris 2007; Freedman 2005; Zumel and Mount 2014).

Consequently, accounting for the exact reasons for certain predictions is very difficult (Bechmann 2017). For businesses with an interest in performance this is not relevant, but for a democratic society that values accountability in decision-making processes it is a major issue (Calo 2017). False positives and false negatives may be hard to account for and instead have to be adjusted for through manual/human alerts and new training and test data iterations. With the development of still more complex models with an increasing number of layers, the challenge of accountability grows. And if such machine learning models have a still more widespread and seamless use in various products, services, and decisions, the processing of manual reports of false positives and negatives may not be prioritized enough, as this requires significant resources from industry and government. Abnormal patterns or "monsters" thereby have a potential to create false positives and negatives because the training data is not sufficient to take these patterns into account in the interpretation and prediction of clusters, or the training data is not labeled sufficiently to take into account such rare or abnormal occurrences. This was also the case before machine learning, but the layers of adjusting for abnormality now become more complex. The lack of training data or labels for innovative structures, minorities, and deviant picture patterns have consequences for the ability of the datafied society to process data as humans on equal terms, especially if this is not a focal point in the design, training, and documentation of the models and their use in specific contexts.

These applied examples illustrate cases in which accountability is difficult and underrepresentation, inequality, and discrimination may easily occur. However, these are just examples of choices made when working with big social data and machine learning that have profound consequences for the democratic values of representation, accountability, and equality. Other examples include the choice of accuracy measures, confidence values, and interpretations of uncertainty information (Diakopoulos and Koliska 2017). The scientific field of machine learning thus filters out outliers but at the same time strives to achieve near-human processing (Harnad 2000; Turing 1950). Machine learning tries to avoid a simple human model but at the same time uses models that normalize data and find similar results (Ananny 2016) instead of concentrating on diversity, for instance in terms of outliers. To create a deeper understanding of the data processing that occurs and ideally enable equal treatment of all humans, documentation of the algorithmic processing is *not* a stand-alone solution (Ananny and Crawford 2018; Cohen 2019). Even if we demanded a circumvention of the intellectual property rights of companies and/or accountability for public data processing, this would be too complex to account for and also insufficient, because discrimination happens when the algorithm is applied to the data as described in this section.

Conclusion

The chapter has provided a conceptual background for data as humans on the basis of Dewey's theory of participatory democracy, and a further development of this theory using cyberfeminist theory to highlight problems with the interpretation of data as humans. It has highlighted challenges to representation, accountability and equality as democratic values and exemplified how such underrepresentation, discrimination, and unaccountability can take place in specific uses of machine learning processing on social big data. These examples have shown how the selection and processing of social big data is profoundly political in nature, thus the examples support existing critical voices in algorithmic studies.

Taking social media as its starting point, the chapter has analyzed democracy on two different levels; as media for democratic debate in Dewey's sense and as a data source for decision-making on a broader scale that defines the citizen in a datafied society (here, social media data plays a role for interpreting the citizen).

Firstly, the chapter has suggested that we should focus not on accounting for the algorithm itself as a standalone solution, but on the *social values* that have been encoded into the algorithm directly or indirectly as (political) choices made by developers. Examples might be: What are the social values of the choices made by Facebook to only show posts people agree with in order to maximize time spent on the platform, or to censor nude pictures? Western media (which social media platforms deliberately avoid registering as) are accustomed to transparency on such issues, but they are challenged by increasingly global social media that also target non-Western societies. Furthermore, such transparency is challenged by Facebook's content moderators situated in different cultures from the ones they serve and with very little editorial education compared to journalists. Can we internationally agree on shared values such as human rights and if not, would it serve democracy and the online public debate to regulate for market cultural sensitivity, that is, sensitivity toward different cultural interpretations of gender equality?

Secondly, the chapter has suggested a political and regulatory focus on documentation and accountability—not as open access to the actual (and often, in private domains, legally protected) algorithm as the only solution, but in terms of an increased focus on normalization logics and the potential negative consequences for protected classes, abnormal behavior, minorities, and underrepresented groups as a part of documenting compliance with international human rights law and the democratic values of representation, accountability and equality. We need new standards for how we create balanced big data data sets, and how we document such balance. Ultimately, if we do not find an effective modus operandi for demanding this documentation of balance and compliance, we will widen the divides already experienced in society more systematically. Again, critical voices would claim that these divides already exist, but I would argue that the systematic and integrated nature of these divides is what sets them apart as new. We no longer have a process of analysis and research followed by actions: instead, we now have a single computational process that operates in loops. If we do not account for balance and compliance with social values, citizen data mining (including in the welfare state) may develop into systematic self-reinforcing loops of discrimination due to the closed learning cycles of machine learning algorithms, informed by big social data.

Dewey's ideal of participatory democracy is not without negative consequences, especially in terms of user privacy when data points increase in tandem with participation. Apart from this big dilemma, future research needs to focus not only on how accountability is created in the different stages of machine learning processing for big (social) data, but also on how such accountability is made accessible to society at large. Finally, "Underrepresented" and outlier focus accounts (e.g., due to false negatives and positives) are a radically different way of approaching big data compared to engineering approaches tackling data-processing techniques and computational optimization. This emphasizes the need for media and communication sociologists to engage and contribute to the field of machine learning so that the analytical models we use in future research as well as in wider society can be tailored to a humanistic approach. This means accounting in detail for the human, both in terms of data and in terms of human decisions in the data processing through machine learning algorithms. By doing so, we may provide a deeper and more detailed account of how power relations are enacted through algorithms and how platforms and services are shaped by designers and users, and in turn shape society.

Notes

1. The right to equality and nondiscrimination is recognized in Article 2 of the Universal Declaration of Human Rights and is a crosscutting standard in different UN human rights instruments. Accountability is a process value related to transparency and the rule of law (see also McGonagle, chapter 9, this volume).

2. This perspective is also found in the growing body of literature on data activism (see, e.g., Gray 2018; Milan 2013).

3. For more references, see https://socialmediacollective.org/reading-lists/critical-algorithm-studies.

4. Facebook restricted the API for third party use in 2019 and has now chosen to commercialize the data and knowledge within the company's own platforms, preventing anyone from monitoring their use of this data.

References

Ananny, Mike. 2016. "Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness." *Science, Technology, & Human Values* 41 (1): 93–117. https://doi.org/10.1177/0162243915606523.

Ananny, Mike, and Kate Crawford. 2018. "Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability." *New Media & Society* 20 (3): 973–989.

Anderson, Margo J. 2015. *American Census: A Social History*. 2nd ed. New Haven, CT: Yale University Press.

Bakshy, Eytan, Solomon Messing, and Lada A. Adamic. 2015. "Exposure to Ideologically Diverse News and Opinion on Facebook." *Science* 348 (6239): 1130–1132. https://doi.org/10.1126/science.aaa1160.

Bauman, Zygmunt. 2000. *Liquid Modernity*. Cambridge: Polity Press.

Bechmann, Anja. 2014. "Non-Informed Consent Cultures: Privacy Policies and App Contracts on Facebook." *Journal of Media Business Studies* 11 (1): 21–38.

———. 2015. "Managing the Interoperable Self." In *The Ubiquitous Internet: User and Industry Perspectives*, edited by Anja Bechmann and Stine Lomborg, 54-73. New York: Routledge.

———. 2017. “Keeping It Real: From Faces and Features to Social Values in Deep Learning Algorithms on Social Media Images.” *Proceedings of the 50th Hawaii International Conference on System Sciences*, 1793–1801. Hilton Waikoloa Village, Hawaii. https://doi.org/10.24251/HICSS.2017.218.

———. 2019 (in press). “Inequality in Posting Behavior over Time: A Study of Danish Facebook Users.” *Nordicom Review* 40 (1).

Bechmann, Anja, Kirstine Bilgrav-Nielsen, and Anne-Louise K. Jensen. 2016. “Data as a Revenue Model.” *Nordicom Information* 38 (1): 76–82.

Bechmann, Anja, Jiyoung Ydun Kim, and Anders Søgaard. 2016. “What We Use Facebook Groups For: A Cross-National Comparison of Private Facebook Groups.” Paper presented at the Annual Seminar for Association of Media Researchers in Denmark. Middelfart, Denmark: SMID.

Bodle, Robert. 2015. “Predictive Algorithms and Personalization Services on Social Network Sites: Implications for Users and Society.” In *The Ubiquitous Internet*, edited by Anja Bechmann and Stine Lomborg, 130–145. New York: Routledge.

Bolukbasi, Tolga, Kai-Wei Chang, James Y. Zou, Venkatesh Saligrama, and Adam T. Kalai. 2016. “Man Is to Computer Programmer as Woman Is to Homemaker? Debiasing Word Embeddings.” *Advances in Neural Information Processing Systems*, 4349–4357.

Bowker, Geoffrey C., and Susan Leigh Star. 1999. *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press.

Bucher, Taina. 2012. “Want to Be on the Top? Algorithmic Power and the Threat of Invisibility on Facebook.” *New Media & Society* 14 (7): 1164–1180.

Bybee, Carl. 1999. “Can Democracy Survive in the Post-Factual Age? A Return to the Lippmann-Dewey Debate about the Politics of News.” *Journalism & Communication Monographs* 1 (1): 28–66.

Calo, Ryan. 2017. “Artificial Intelligence Policy: A Primer and Roadmap.” *U.C. Davis Law Review* 51: 399–436.

Cheney-Lippold, John. 2011. “A New Algorithmic Identity: Soft Biopolitics and the Modulation of Control.” *Theory, Culture & Society* 28 (6): 164–181.

Cohen, Julie E. 2019. “Turning Privacy Inside Out.” *Theoretical Inquiries in Law* 20 (1). https://papers.ssrn.com/abstract=3162178.

Crawford, Kate, and Ryan Calo. 2016. “There Is a Blind Spot in AI Research.” *Nature News* 538 (7625): 311. https://doi.org/10.1038/538311a.

Davenport, Thomas H., and Jeanne G. Harris. 2007. *Competing on Analytics: The New Science of Winning*. Boston: Harvard Business Press.

Derczynski, Leon, Kalina Bontcheva, and Ian Roberts. 2016. "Broad Twitter Corpus: A Diverse Named Entity Recognition Resource." In *Proceedings of COLING 2016,* 1169–1179. https://aclanthology.info/papers/C16-1111/c16-1111.

Desrosières, Alain. 2002. *The Politics of Large Numbers*. Cambridge, MA: Harvard University Press.

Dewey, John. 1927. *The Public and Its Problems*. New York: Holt, reprint 1946.

Diakopoulos, Nicholas, and Michael Koliska. 2017. "Algorithmic Transparency in the News Media." *Digital Journalism* 5 (7): 809–828.

Foucault, Michel. 1977. *Discipline and Punish: The Birth of the Prison*. New York: Vintage Books.

Freedman, David. 2005. *Statistical Models: Theory and Practice*. New York: Cambridge University Press.

Galloway, Alexander R. 2004. *Protocol: How Control Exists after Decentralization*. Cambridge, MA: MIT Press.

Gillespie, Tarleton. 2014. "The Relevance of Algorithms." In *Media Technologies: Essays on Communication, Materiality, and Society*, edited by Tarleton Gillespie, Pablo J. Boczkowski, and Kirsten A. Foot, 167–193. Cambridge, MA: MIT Press.

Gitelman, Lisa. 2006. *Always Already New*. Cambridge, MA: MIT Press.

Gray, Jonathan. 2018. "Three Aspects of Data Worlds." *Krisis: Journal for Contemporary Philosophy* (1): 4–17.

Haraway, Donna. 1991. *Simians, Cyborgs and Women: The Reinvention of Nature*. New York: Routledge.

Hargittai, Eszter, and Gina Walejko. 2008. "The Participation Divide: Content Creation and Sharing in the Digital Age." *Information Communication & Society* 11 (2): 239–256.

Harnad, Stevan. 2000. "Minds, Machines and Turing." *Journal of Logic, Language and Information* 9 (4): 425–445.

International Federation of Journalists. 1986. *Principles of Conduct of Journalism*. https://www.ifj.org/who/rules-and-policy/principles-on-conduct-of-journalism.html.

Introna, Lucas, and Helen Nissenbaum. 2000. "Shaping the Web: Why the Politics of Search Engines Matters." *The Information Society* 16 (3): 169–185.

Introna, Lucas, and David Wood. 2004. "Picturing Algorithmic Surveillance: The Politics of Facial Recognition Systems." *Surveillance & Society* 2 (2–3). https://ojs.library.queensu.ca/index.php/surveillance-and-society/article/view/3373.

Jurafsky, Daniel, and James H. Martin. 2008. *Speech and Language Processing*. 2nd ed. Upper Saddle River, NJ: Prentice Hall.

Kitchin, Rob. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: Sage.

Kroll, Joshua A., Joanna Huey, Solon Barocas, Edward W. Felten, Joel R. Reidenberg, David G. Robinson, and Harlan Yu. 2017. "Accountable Algorithms." *University of Pennsylvania Law Review* 165: 633–699.

Law, John. 1990. "Introduction: Monsters, Machines and Sociotechnical Relations." *The Sociological Review* 38 (S1): 1–23.

Leese, Matthias. 2014. "The New Profiling: Algorithms, Black Boxes, and the Failure of Anti-Discriminatory Safeguards in the European Union." *Security Dialogue* 45 (5): 495–511.

Lippmann, Walter. 1927. *The Phantom Public*. New Brunswick, NJ: Transaction.

Lyon, David. 2007. *Surveillance Studies: An Overview*. Cambridge: Polity Press.

Mackenzie, Adrian. 2015. "The Production of Prediction: What Does Machine Learning Want?" *European Journal of Cultural Studies* 18 (4–5): 429–445.

Mager, Astrid. 2012. "Algorithmic Ideology." *Information, Communication & Society* 15 (5): 769–787.

Manning, Christopher D., and Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.

Marwick, Alice E., and danah boyd. 2014. "Networked Privacy: How Teenagers Negotiate Context in Social Media." *New Media & Society* 16 (7): 1051–1067.

McLuhan, Marshall. 1964. *Understanding Media: The Extensions of Man*. New York: Signet.

Metcalf, Jacob, and Kate Crawford. 2016. "Where Are Human Subjects in Big Data Research? The Emerging Ethics Divide." *Big Data & Society* 3 (1). https://doi.org/10.1177/2053951716650211.

Milan, Stefania. 2013. *Social Movements and Their Technologies: Wiring Social Change*. New York: Palgrave MacMillan.

Nissenbaum, Helen. 2011. "A Contextual Approach to Privacy Online." *Daedalus* 140 (4): 32–48.

Noyes, Katherine. 2015. "Will Big Data Help End Discrimination—Or Make It Worse?" *Fortune*. http://fortune.com/2015/01/15/will-big-data-help-end-discrimination-or-make-it-worse.

Pasquale, Frank. 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.

Plutchik, Robert. 2001. "The Nature of Emotions: Human Emotions Have Deep Evolutionary Roots, a Fact That May Explain Their Complexity and Provide Tools for Clinical Practice." *American Scientist* 89 (4): 344–350.

Rieder, Bernhard. 2005. "Networked Control: Search Engines and the Symmetry of Confidence." *International Review of Information Ethics* 3 (1): 26–32.

———. 2017. "Scrutinizing an Algorithmic Technique: The Bayes Classifier as Interested Reading of Reality." *Information, Communication & Society* 20 (1): 100–117.

Rogers, Richard. 2009. "The Googlization Question, and the Inculpable Engine." In *Deep Search: The Politics of Search beyond Google*, edited by Konrad Becker and Felix Stalder, 173–184. New Brunswick, NJ: Transaction.

Sandvig, Christian, Kevin Hamilton, Karrie Karahalios, and Cedric Langbort. 2014. "Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms." Paper presented at Data and Discrimination: Converting Critical Concerns into Productive Inquiry, May 22, Seattle, WA.

Schäfer, Mirko Tobias, and Karin van Es. 2017. *The Datafied Society*. Amsterdam: Amsterdam University Press.

Solove, Daniel J. 2004. *The Digital Person: Technology and Privacy in the Information Age*. New York: NYU Press.

Star, Susan Leigh. 1990. "Power, Technology and the Phenomenology of Conventions: On Being Allergic to Onions." *The Sociological Review* 38 (S1): 26–56.

Stutzman, Fred, Ralph Gross, and Alessandro Acquisti. 2012. "Silent Listeners: The Evolution of Privacy and Disclosure on Facebook." *Journal of Privacy and Confidentiality* 4 (2): 7–41.

Turing, Alan M. 1950. "Computing Machinery and Intelligence." *Mind* 59 (236): 433–460.

Turkle, Sherry. 1995. *Life on the Screen: Identity in the Age of the Internet*. New York: Simon & Schuster.

Turow, Joseph. 2011. *The Daily You: How the New Advertising Industry Is Defining Your Identity and Your Worth*. New Haven, CT: Yale University Press.

Zumel, Nina, and John Mount. 2014. *Practical Data Science with R*. Greenwich, CT: Manning.