Received 30 November 2020
Revised 24 February 2021
22 March 2021
6 April 2021
Accepted 6 April 2021

Towards a political theory of data justice: a public good perspective

Chi Kwok
Department of Philosophy and Religious Studies, Utrecht University, Utrecht, The Netherlands and Department of Political Science, Lingnan University, Hong Kong, China, and

Ngai Keung Chan
School of Journalism and Communication, The Chinese University of Hong Kong, Hong Kong, China

Abstract

Purpose – This study aims to develop an interdisciplinary political theory of data justice by connecting three major political theories of the public good with empirical studies about the functions of big data and offering normative principles for restricting and guiding the state's data practices from a public good perspective.

Design/methodology/approach – Drawing on three major political theories of the public good – the market failure approach, the basic rights approach and the democratic approach – and critical data studies, this study synthesizes existing studies on the promises and perils of big data for public good purposes. The outcome is a conceptual paper that maps philosophical discussions about the conditions under which the state has a legitimate right to collect and use big data for public goods purposes.

Findings – This study argues that market failure, basic rights protection and deepening democracy can be normative grounds for justifying the state's right to data collection and utilization, from the perspective of political theories of the public good. The state's data practices, however, should be guided by three political principles, namely, the principle of transparency and accountability; the principle of fairness; and the principle of democratic legitimacy. The paper draws on empirical studies and practical examples to explicate these principles.

Originality/value – Bringing together normative political theory and critical data studies, this study contributes to a more philosophically rigorous understanding of how and why big data should be used for public good purposes while discussing the normative boundaries of such data practices.

Keywords Big data, Data justice, Political theory, Public good, The state

Paper type Research paper

1. Introduction

Big Data, explain boyd and Crawford (2012), is not simply about the *quantities* of information, but "a capacity to search, aggregate and cross-reference large data sets" and a mythology that data can generate "objective" and "accurate" knowledge to predict and inform public decisions (p. 663). Contemporary studies of big data and surveillance capitalism are critical of the repressive potential of data (Couldry and Mejias, 2019; Zuboff, 2019); for example, the deployment of algorithmic and data-driven systems may amplify social biases (Barocas and Selbst, 2016; Lee, 2018), exacerbate social inequalities (Eubanks, 2018), naturalize the exploitation of data subjects (Couldry and Mejias, 2019) and result in infringement of individual

Authors would like to thank the special issue editor Charles Ess and the anonymous reviewer for their invaluable help with earlier drafts of this article. The earlier version of the article was presented at the AoIR 2020 conference. Authors would like to thank the audience for their helpful suggestions. All errors and omissions are, of course, their own.

© Emerald Publishing Limited
1477-996X
DOI 10.1108/JICES-11-2020-0117

privacy (Nissenbaum, 2017). Critical scholars, therefore, have advocated that the access, collection, processing and uses of personal data must be constrained by values of justice, such as transparency, fairness and accountability (Dencik *et al.*, 2016; Kemper and Kolkman, 2019; Lepri *et al.*, 2018; Taylor, 2017; Wieringa, 2020), while questioning the limitations of transparency (Ananny and Crawford, 2018) [1]. However, the creation and dissemination of high-quality and comprehensive data also have the potential to increase transparency in market competition (Acquisti, 2014), improve public health conditions (Ginsberg *et al.*, 2009) and benefit the development of cities where services and resources can be more efficiently delivered to citizens (Goerge, 2014).

Pinpointing the emergence of "data politics," Bigo *et al.* (2019) argue that "[d]ata is not only shaping our social relations, preferences and life chances but also our very democracies" (p. 5). The epistemic opacity of big data – the difficulties for data subjects to understand what is being collected and how data is potentially used (Andejevic, 2014) – has made datafication a form of power that the state could exercise without being accountable to the citizens. Hence, democratic citizens simply cannot monitor and enforce political accountability to what they do not comprehend. Importantly, the collection and analytics of data could draw worrying inferences about data subjects (Wachter and Mittelstadt, 2019) and potentially pose a serious threat to democracy. A recent example is how Cambridge Analytica "allegedly influenced both the US election and the UK referendum by mining data from Facebook and using it to create profiles predicting people's personalities and then tailoring advertising to their psychological profiles" (Bigo *et al.*, 2019, p. 5). The state's alignment with the force of data raises legitimate concerns over issues pertaining to the right of the state's collection and utilization of data and its normative and practical implications.

The state's collection and utilization of big data, therefore, appears to be a double-edged sword: big data simultaneously enables the state's ability to improve people's living conditions *and* its ability to abuse its power, which threatens privacy and freedom of democratic citizens. boyd and Crawford (2012) posed two central questions about the relations between big data and politics:

(1) Whether big data can become a public good that is beneficial to people's well-being and good life.

(2) Whether the state should be granted the right to collect big data.

Following concerns over data governance (Rogerson *et al.*, 2017), this article argues that big data can serve as a public good, but the legitimacy of the state to collect and use them depends on whether it fulfills all three conditions:

(1) The principle of transparency and accountability.

(2) The principle of fairness.

(3) The principle of democratic legitimacy.

The contributions of this article are twofold, namely, first, it develops an interdisciplinary political theory of data justice by connecting three major political theories of the public good (Kohn, 2021) with empirical studies about the functions of big data. It systematically defends the state's right to collect and use big data from a public good perspective, while acknowledging the limitations of such data practices. Second, it offers a preliminary normative framework to qualify the conditions under which the state's right to collect big data for beneficial public purposes can be legitimate in a democratic context. Following Lane *et al.* (2014), our primary goal is to consider the ethical requirements of justice for "government officials seeking to use big data to serve the public good without harming individual citizens" (p. xi).

This article is structured as follows. Section 2 discusses three major political theories of the public good (market failure, basic rights and democratic theory) and their normative implications for the right of the state to collect and use big data for public good purposes. Section 3 outlines three major political principles: transparency and accountability, fairness and democratic legitimacy. The three principles serve to constraint and guide the state's data practices for minimizing the potential political abuse of data power. Section 4 summarizes our major arguments and points to the need to pay more attention to issues concerning what institutional arrangements can help to apply the three principles in practice.

2. Theories of the public good and the state's collection and use of big data

Public goods generally refer to goods that are publicly beneficial and yet cannot be sufficiently supplied by the market or should not be supplied by it due to moral and ethical concerns. Following Kohn's (2021) typology, we examine how three major approaches of the public good (market failure, basic rights and democratic) in the political theory literature could provide useful intellectual resources for theorizing different types of big data as public goods. Below we briefly define each approach and its relation to big data [2].

2.1 The market failure approach

This approach suggests that when goods are widely beneficial to the public and yet are not profitable, the inability of the market to provide these goods to a sufficient degree renders the state a legitimate claim to provide them (Kohn, 2021; Batina and Ihori, 2005). Consider, for example, real-time traffic data. They could inform drivers to avoid traffic congestion and thereby improve road safety, but they would only be widely used by drivers when the data are freely available to them (Shi and Abdel-Aty, 2015). From the perspective of public safety, it would be self-contradictory to restrict access to these traffic data to only drivers who have paid a fee. Nonetheless, "gate-keeping" is usually a precondition of profit-making. If corporations consider the non-exclusive use and sharing of data in the market as non-profitable, they lack the incentive to produce them. Highlighting the distinction between digital infrastructures and digital *public* infrastructures, Zuckerman (2020) aptly argues that *[m]arkets do not always provide the infrastructures we need* (p. 6; emphasis original). For gate-keeping to be effective, the government must also enforce rules and laws that protect the exclusion of free-riders. Thus, if publicly beneficial data is only distributed through the market, the market will distribute data in ways that are unable to maximize the public good functions.

Another structural problem associated with the private provision of big data is that individual corporations are unable to ensure the consistency of the design of data collection for public purposes. Poom *et al.* (2020) have rightly pointed out that the methodologies behind large platform companies, such as Google and Facebook, are often "black-boxed" and "it is difficult to evaluate their [ad hoc data collected through these platforms] usefulness or potential for future use" (p. 3). The algorithms that platform firms rely on to collect data and produce results are commercial secrets that firms intentionally attempt to hide from competitors and the public (Burrell, 2016; Pasquale, 2015). Because the state is unable to take control over the ways how the data collecting algorithms are designed, the data from these platforms, therefore, do not necessarily fit for social good purposes (Poom *et al.*, 2020).

Even if business corporations are interested in taking part in the production of publicly beneficial big data, the provision of such type of big data through business actors might lead to concerns about the potential commercialization of personal data and its negative effects, which could disincentivize people from using them. Fourcade and Healy (2017) observe that modern big data are increasingly a form of capital that business corporations are under a data imperative to collect as much data as possible: “organizations believe they should be in the data collection business, even when they do not yet know what to do with what they collect” (p. 17). In addition, even if “a firm is not sure how to extract its value, there are other organizations that know or claim to know” (Fourcade and Healy, 2017, p. 17). The major worry here is that publicly beneficial data, when being collected, used and sold by business corporations, are more likely to result in the abuse of data and infringement of privacy. This is a legitimate concern provided that recent examples have shown that big tech firms, such as Google and Facebook, often disparage norms and legal restrictions that regulate the appropriate use of big data until public outrage explodes (Zuboff, 2019). As Zuckerman (2020) argues, “Facebook was designed not to enable citizenship but to display ads to users” (p. 8).

What is more problematic is the fact that large business corporations often actively seek to change and reshape rules that regulate them directly and indirectly. For example, the federal trade commission (FTC) antitrust case in the USA was unanimously dropped by its members despite a leaked FTC staff report showed that members had concluded that “Google had unlawfully maintained its monopoly over general search and search advertising” through “scraping content from rival vertical websites,” “by entering into exclusive and highly restrictive agreements with web publishers that prevent publishers from displaying competing search results or search advertisement” and “by maintaining contractual restrictions that inhibit the cross-platform management of advertising campaigns” (Zingales, 2017, p. 123). As Zingales (2017) suggests, “one wonders if the frequent visits paid by Google employees to the White House played a role,” since “between Obama’s first in the inauguration and the end of October 2015, employees of Google and associated entities visited the White House 427 times, including 21 small, intimate meetings with President Obama” (p. 123). The possession of data extracted from a large user base would only reinforce corporations’ ability to reshape the rules of regulation through not only non-democratic means but also “democratic” means. Platform firms, for example, could identify and mobilize sympathetic democratic citizens to contest democratic institutions through their large databases (Pollman and Barry, 2017). Business corporations also tend to shield themselves from democratic scrutiny by appealing to the discourse that they are private associations whose primary purposes are innovation and efficiency (Zuboff, 2019). The paradox, therefore, is that if the responsibility for the provision of publicly beneficial big data falls entirely on the market, it is likely that it will result in a situation in which these providers become a greater threat to both the privacy of individual citizens and the core values of liberal democracy. The cost of reinforcing the power of these large corporations might even outweigh the benefits generated by the publicly beneficial data that they provide.

Thus, when it comes to the provision of big data for common good purposes, the market is not a reliable mechanism for three reasons. First, it lacks the incentive to produce publicly beneficial data that are unprofitable. Second, when the duty to collect and distribute publicly beneficial data is allocated to the market, it may not be able to maximize the common good functions of big data due to the gate-keeping precondition for profit. Third, even if business corporations are willing to bear the duty to produce such data and provide open access, their involvement might raise concerns over their potential commercialization behaviors, which could disincentivize citizens who take privacy seriously from using such data. The failure of the market to provide safe and free publicly beneficial data renders a reason for the intervention of the state.

The claim here is not that the state will not infringe upon data privacy. Instead, we argue that it is easier for democratic citizens to gain control over how data is collected, managed and distributed by the state than by business corporations in a democratic context. On the one hand, business corporations increasingly frame themselves as private actors in the market and rely on this framing to shield themselves from democratic accountability. On the other hand, business corporations have also successfully produced discourses that project state intervention as a hindrance to progress and innovation (Zuboff, 2019). These two factors make the state, rather than business corporations, a better provider of publicly beneficial data.

2.2 The basic rights approach

Shue (1996), in his work *Basic Rights*, argues that goods related to physical security and basic subsistence – because of their paramount significance to the good life of individual citizens – *ought* to be provided and guaranteed by the state. As Kohn (2021) suggests, the basic rights approach to public goods "is a normative theory that holds that state provision is justified when it is necessary to supply primary goods" (p. 1106). This approach emphasizes that the state should provide these essential goods not because the market cannot, but because the state has the duty to ensure equal and fair access to these goods, which the market is normally unable to. When certain goods are fundamental to the interests of citizens and whose distribution therefore, needs to adhere to normative principles of fairness and equality, which cannot be guaranteed by the market, then there is a normative case for the state to produce and distribute such goods. Many political theorists perceive that the political legitimacy of the state, at least to a significant extent, hinges on its capacity to deliver essential goods to citizens (Rawls, 2001; Dworkin, 2002). The state's provision of essential goods that the market could also provide is not a new idea. Water, electricity and health care are examples of essential goods that many countries have decided to take on the responsibility to deliver such goods. In principle, local police force and national defense could also be provided by the market through security services and mercenaries and yet, as the provision of these goods must adhere to principles of equality and fairness, most states have chosen not to delegate such responsibility to the market.

Does big data belong to the category of essential goods? Some researchers have suggested that the provision of most basic services and goods is increasingly mediated by digital platforms and big data (Pistor, 2020). Consider the example of pandemic data. The Centers for Disease Control and Prevention in the USA has long collected data regarding the spread of various diseases and used these data to advise people to take appropriate preventive measures. Big data for public health purposes can significantly improve prediction speed and precision (Ginsberg *et al.*, 2009), and hence, better protect the lives of many. The COVID-19 pandemic has shown that the capacity of the state to collect and distribute pandemic-related demographic data is the key to facilitate effective policy-making (Poom *et al.*, 2020), and the availability of such free and massive scale data also enables citizens to take into account risk factors when they need to travel to regions and areas with high infection rates.

Indeed, the collection and uses of pandemic data, and more broadly, public health data raise questions about the *normalization* of state surveillance (Deibert, 2020). As discussed previously, data collection and analytics are closely related to power relations between infrastructures, political and economic agents and data subjects. The state's data collection comes with concerns over vulnerability (D'Ignazio and Klein, 2020). It is important to problematize the "objective" nature of data and recognize what such data might neglect.

One might, therefore, reject the idea that the state ought to have the right to nudge individual citizens' behaviors through the collection and use of big data. This might be a pressing issue in non-emergency contexts, where the role of big data is not closely connected to the basic security of individual citizens, and thus, the employment of big data to modify behaviors of individual citizens in non-emergency contexts ought to face more serious constraints than in a crisis. However, in times of crises, it is important to recognize that the case against the use of big data to inform and moderate citizens' behaviors becomes weaker. One of the reasons for the provision of real-time national and regional COVID-19 infection data aims precisely to change individual behaviors – minimizing their intention to travel to areas of high risk and take precautions. Indeed, medical researchers have found that a major factor that explains Taiwan's low infection rate is its ability to leverage its "health insurance database and integrated it with its immigration and customs database to begin the creation of big data for analytics," which has helped to generate "real-time alerts during a clinical visit based on travel history and clinical symptoms to aid case identification" (Wang *et al.*, 2020, p. E1).

The need to deliver essential goods or protect citizens' basic interests in times of crisis would also warrant the right of the state to collect relevant big data in normal times. Data that can be preserved and used for specific purposes for the future require a certain degree of methodological consistency and that further requires institutional design and configuration (Kitchin, 2014). The Finnish example about COVID-19 data is worth quoting in some length here:

> [I]n Finland, access to mobile phone data has been rather limited all the time due to strict interpretation of privacy related legislation [. . .] Recently, however, a main mobile network operator, Telia, developed an aggregated and anonymized data product allowing mobility analysis at the scale of the entire population. When the COVID-19 pandemic started, the existence of this ready-made data product allowed governmental officials and researchers quick data to uncover changing mobility flows brought about by closing the borders of the capital region and instructing citizens to avoid visiting secondary homes. However, the relatively simple data product did not leverage or allow access to individual-level raw data necessary to create custom spatial and categorical aggregations. Moreover, because Telia's preconstructed data products were designed to answer specific questions, they could not always address the new questions resulting from COVID-19. Further complicating the application of these data products was that the methodology behind them was not transparent enough to understand fully how the resulting values are derived. *Thus, even when access to mobile Big Data is available, it may not be structured in ways that fit the specific needs that arise during a crisis* (Poom *et al.*, 2020, p. 3, emphasis added).

The Finnish example demonstrates that to have ready-at-hand data for crisis management purposes, central governmental agencies must have well-crafted comprehensive data that are designed specifically for such purposes far in advance. This implies that they need to have standard practices of data collection and stable collaboration with research institutions for meaningful data analysis.

Once we accept the claim that big data *can* and *should* be used to protect human lives and other essential goods, then controversies around the scope of the collection and the appropriate deployment of big data do not preclude the right of the state to collect such data. Nonetheless, it must be stressed that the basic rights approach does not entail that big data could be collected or used for any governmental purpose. Instead, the normative justification lies at the connection between big data and essential goods. It is only when big data becomes an essential good or is tightly connected to the state's capacity to deliver essential goods, such as the physical security of individual citizens, then the state's collection and employment of big data is justifiable. What makes big data an essential good is certainly contextual. These political questions must be determined by citizens themselves through a fair and open democratic process.

2.3 The democratic approach

The democratic approach points out that "public goods are goods that provided by the "public" (e.g. the state) to the 'public' (e.g. citizens or residents)" and in "democratic states, the term public goods is often used to describe things that the majority of the citizens, through their elected representatives, choose to provide" (Kohn, 2021, p. 1107). This approach implies at least two categories of democratic public goods. The first category is goods that democratic citizens collectively want (call them *democratic public goods*) and the second category is goods that can foster and strengthen the democratic process to better account for what goods democratic citizens collectively want (call them *foundational democratic public goods*). The second category is foundational in the sense that an effective provision of democratic public goods partially depends on the existence of foundational democratic public goods. Put differently, foundational democratic public goods are democratic public goods for deepening democracy. Foundational democratic public goods are the focal point of this section. Examples of such goods include free and high-quality news information and free basic education (Claassen, 2018). Typically, these goods are regarded as duties of the state, despite the fact that, in principle, they could be provided by other agents through different distributive mechanisms. The case in favor of the collective provision of foundational democratic public goods is usually made on the basis that a democratic state has the responsibility to ensure the well-functioning of the democratic procedure, and that further implies that a democratic state must be able to secure channels of democratic voice-taking and participation to ensure the legitimacy of the democratic procedure. One of the major ways to secure democratic input is to diversify channels of input for information dissemination and policy-making purposes.

Is big data a foundational democratic public good? As Smith (2010) suggests, an area of information technology that public authorities have been interested and highly active is "the provision and dissemination of information: most public authorities have websites that provide access to reams of official documentation" (p. 143). For example, an increasingly important open data movement has perceived a major condition of an open government to be "opening public sector information data and enabling citizens and entrepreneurs to access government data in a uniform way" (Veljković *et al.*, 2014, p. 279). Thus, "information is a necessary resource for informed political participation and can increase transparency" (Smith, 2010, p. 143). This includes providing open access to mass data collected by the government to create opportunities not just for citizen engagement but also for "enabling cooperation across different levels of government, between the government and private institutions and between the government and the citizens" (Veljković *et al.*, 2014, p. 279). Consider the example that some local governments in the US collaborate with civic technology firms, such as SeeClickFix and PublicStuff, to provide convenient ways for citizens to report local infrastructural problems directly to local governments. Such initiatives not only strengthen local governments' understanding of communal needs but also incentivize citizens to actively participate in local governance (Graeff, 2018).

A common critique of representative democracy is that it essentially reduces citizens' role to be merely passive and their major duty is simply to elect the right groups of elites to govern them (Schumpeter, 2008). Thus, deepening democracy requires the state to strengthen democratic participation beyond those provided by formal representative institutions. "E-democracy" could be one of the methods (Akman *et al.*, 2005; Smith, 2010; Veljković *et al.*, 2014) because, through establishing relatively accessible forms of e-interactive channels, data collected through these interactions are foundational democratic goods in the sense that they enable the government to understand better the needs of citizens, especially those that are less salient in formal representative channels.

The state as a political agent, which holds any amount of resources that is unavailable to agents in private realms is, therefore, able to collect data on a grand scale (Kitchin, 2014). Enabling the state to do this for common good purposes can deepen democracy for at least two reasons. First, policy innovations and recommendations require data. The more comprehensive open data are, the more likely that the government can provide a better data infrastructure for high-quality civil society engagement. In addition, comprehensive, transparent and open big data could empower civil society actors to critically examine and contest government policies. Following Tocqueville, Smith (2010, p. 144) describes this as an opportunity to rebuild "twenty-first century town meetings," as evidence has shown that in participatory budgeting in Porto Alegre, the "transparency of information allows citizens involved in different elements of the budgetary process to keep track of investments, undertake research on the administration and its agencies' activities and hold the administration and budget delegates and councilors to account." Organizers of these new types of town hall meetings could rely on such data to target traditionally under-represented groups and factor in their under-representation in the design of local level democratic procedures (Smith, 2010). This is because e-town meetings, by lowering the cost of participation and increasing flows of information, can enhance the organizers' capacity to recruit demographically diverse participants and strengthen the multiplicity of voices in the deliberative process. Given that the civil society is an important autonomous sphere, which is indispensable for enriching and providing important democratic information to the formal legislative process (Habermas, 1994), granting the state the right to the collection of data and stressing its duty to distribute them in a reasonable manner are therefore, an important condition to take advantage of the democratic potential in data to the deepening of democracy.

2.4 Summary

Some types of big data can fit into more than a single approach of the public good, and hence, the conceptual distinction of the three approaches is only for analytical clarity. In reality, we might have multiple grounds to conceive a type of big data as a public good, which legitimizes and demands the state's provision. Furthermore, the state's collection of mass data for public purposes is not a new phenomenon. The modern state's attempt to resolve political problems with the assistance of mass data can be traced back to the mid-nineteenth century (Bigo *et al.*, 2019). As Hacking (2015) puts it, the history of governmental data collection "represented an overt political response by the state" (p. 281); that is, to "find out more about your citizens, cried the conservative enthusiasts and you will ameliorate their conditions, diminish their restlessness and strengthen their character" (Bigo *et al.*, 2019, p. 3). In other words, the state has long relied on data to govern and it does at least partially contribute to the common good. If we accept the claim that big data do have significant democratic potential and there are multiple normative grounds to support the right of the state for big data collection, then the most pressing issue is how to balance the benefits of big data and the protection of individual privacy and freedom. Drawing insights from existing studies, we propose three preliminary principles that ought to guide the political regulations of democratic states' data collection practices.

Before turning to the discussion of the three guiding principles of justice for the regulation of the state's collection and uses of big data, an important question concerning the legitimate scope of data collection needs to be addressed. One might question whether granting the right to the state to collect relevant big data in normal times for crisis management purposes would lead to the conclusion that any data could be collected in normal times as almost any data could be imagined as useful in unknown future crises. It must be noted that we do not uphold the position that a right for the state to collect big data implies an unconstrained right for the state to do this.

In fact, what data collection practices can be deemed reasonable from the citizens' point of view does not and will not have a settled answer. Which political authorities should have access to what kinds of data under which particular purposes are questions that could only be answered through constant and ongoing democratic deliberation and legal contestation. For instance, the controversial Investigatory Power Act 2016 in the UK, which established the legal right for governmental agencies to collect and intercept communication data for security purposes (UK Government, 2016) was ruled by the UK's high court to violate EU law (Cobain, 2018). In light of this ruling, in the following amendment of Act, known as The Data Retention and Acquisition Regulations (2018), a key change was a higher threshold for data access: government authorities can access communication data only in cases where an offense "is capable of being sentenced to imprisonment for a term of 12 months or more" (SI 2018/1123). We envision that three subject matters will need to be addressed through periodic public deliberation:

1. The legitimate scope of the trade-off between privacy and data collection for public good purposes.
2. Which government agencies under what conditions could be granted the right to access what types of data.
3. Whether there should be a limit on the data retention period for specific sensitive data.

3. Guiding political principles and data justice

We propose here three guiding principles of justice for the regulation of the state's *collection* and *uses* of big data. Drawing insights from existing studies on algorithmic transparency, fairness and accountability, the following framework attempts to theorize the political values that should guide the democratic state's data practices.

The first political principle is *the principle of transparency and accountability*. The purposes and processes of data collection and utilization must be transparent and open. A central worry about the state's collection of big data is that the processes might infringe on individual citizens' privacy and freedom (Nissenbaum, 2017). Making the processes transparent opens possibilities for not only monitoring the outcomes of data practices but also contesting the outcomes. At its most basic level, the state should make its design and development stages of data science open to the public. One practical challenge, however, is whether such an explanation could be meaningful for data subjects to unpack the logic of algorithmic decision-making (Burrell, 2016). Wachter *et al.* (2018), therefore, advocate for providing "counterfactual explanations" for automated decisions. Counterfactual explanations are statements of "how the world would have to be different for a desirable outcome to occur" (e.g. a loan was denied because of the subject's insufficient income) (p. 844). While this mechanism does not explain the internal workings of machine learning, it could be a possible, complementary tool to offer understandable explanations of data-driven decisions and provide data subjects with grounds to reverse and challenge the outcomes. Nonetheless, this would also require the state to disclose the methods of constructing counterfactual explanations.

The monitoring of the state's uses of big data requires an *active* contestatory civil society where misbehaviors of the state would be publicly exposed. Thus, the principle requires the state to not only be transparent but also to provide a favorable legal infrastructure for activism against data abuse. By way of example, civil society groups have played an important role in raising general awareness of the danger of state surveillance (Dencik *et al.*, 2016; Raley, 2013). In addition, the UK Investigatory Power Bill demonstrated that the civil society could be a valuable input to the legislative process with regard to the state's legitimate scope of data collection. Non-governmental organizations (NGOs) were quite heavily involved in the legislation, as an anti-surveillance activist noted: "[p]reviously NGOs would have fought just to kill a new law and probably been unsuccessful in doing so; now they can say: here is how we can genuinely improve it and have a proper conversation with the Home Office" (Hintz and Brown, 2017, p. 794; Dencik *et al.*, 2016, p. 4). What we are suggesting here is not that the Bill contains no problem, but that this example shows that both productive contestatory and collaborative relationships between the state and NGOs are possible, and the principle of transparency and accountability implies a political duty for the state to actively cultivate both relationships.

The effectiveness of data collection depends to a significant degree on mutual trust. If citizens generally distrust the state, bottom-up data resistance and activism would undermine the accuracy and efficiency of data collection (Beraldo and Milan, 2019; Dencik *et al.*, 2016; Milan and Treré, 2019). Consider the example of contact tracing apps. Digital contact tracing may only be effective when a "critical mass" of citizens uses the apps. However, citizens may boycott such apps when they have privacy concerns about how the data may be (mis)used for political purposes (Lau, 2021). To account for the tensions between public health concerns and citizens' privacy concerns, Vitak and Zimmer (2020) propose that the theory of contextual integrity (Nissenbaum, 2010) is a viable framework for evaluating the appropriateness of information flows in the context of COVID-19. Specifically, the appropriateness of collecting and sharing personal data is contextually dependent on what information is being collected, shared with whom and with purposes. People have varying degrees of institutional trust when considering their willingness to share health data (Hargittai *et al.*, 2020). The state, therefore, should carefully consider citizens' institutional trust and the contexts of data practices.

Additionally, the European Union's General Data Protection Regulation (GDPR) represents one of the recent attempts to build mutual trust through informed data and privacy protection. GDPR requires organizations to inform data subjects about the collection and use of their personal data and provide lawful and valid grounds for such data practices (Information Commissioner's Office, 2021). Accountability is another key principle of data protection, meaning that "data controllers have to be able to demonstrate they are GDPR compliant" by, for example, "maintain[ing] detailed documentation of the data you are collecting, how it is used, where it is stored, which employee is responsible for it" (GDPR.EU, 2021). As such, GDPR represents a possible direction of legal interventions into states' and corporations' data practices. Yet, as Wachter and Mittelstadt (2019) argue, the GDPR focuses primarily on the protection of the stage when data is collected rather than the process of data analysis. Data controllers, in practice, can draw inferences about data subjects, based on not only the data provided by the latter but also third-party predictive assessments (e.g. credit score). As such, inferential analytics could have a low degree of verifiability and affect data subjects' life chances, Wachter and Mittelstadt argue that a right to reasonable inferences should be implemented. This right would necessitate both

(1) "*ex ante* justification to be given by the data controller to establish whether an inference is reasonable" and

(2) an ex post accountability mechanism for "data subjects to challenge unreasonable inferences" (p. 123; italics original).

Without this right, it may be difficult for data subjects to understand how a specific data-driven decision about themselves is made, and therefore, protect themselves from potential privacy-invasive and discriminatory uses of their personal data.

Alongside legal interventions, it is equally important to avoid romanticizing "open data" or simply use it as a rhetorical strategy to legitimize the state's authority (Currie, 2020). Echoing Zuckerman's (2020) call for digital public infrastructures, we need to envision "decentralized" and "pluralized" networks of data infrastructures. Even though the state has the right to collect data, it does not mean that there is only one system for collecting, aggregating and analyzing citizens' data. Instead, data practices can take place in multiple networks, with distinctive community-based norms and citizen self-governance. Studies have shown that citizens could gather and mobilize data for improving community life (Graeff, 2018; Schrock and Shaffer, 2017) and monitoring local environmental problems (Gabrys *et al.*, 2016) by themselves. For example, citizens could produce what Gabrys *et al.* (2016) called "just good enough data" (p. 2) to document patterns of evidence about air pollution and make collective claims to engage with regulators in Pennsylvania. While such data might not be "big" by size, citizens' data practices reveal the potential for creating alternative data stories and discourses. Such data can then be used for communicating with local governments and holding the authorities accountable. Furthermore, it is important for local governments to take an active step to learn from innovations and practices of these citizen's initiatives and it is also worth considering how the state could offer funding to support these local public networks. The goal is to appreciate diverse identities and values with respect to design and knowledge construction to avoid "border closures, oppressive social control, exclusionary data sets or apps catering solely to the majority" (Milan, 2020, p. 5).

The second political principle is *the principle of fairness*. The state's collection and uses of big data rely on public finance, and the design of what and how data should be collected and distributed is never neutral (Eubanks, 2018). Different designs will result in different social and political groups being benefited. This principle requires the state not only to justify the uses of big data by explaining how it can benefit the public but also to reasonably explain how the design of data collection does not unfairly skew toward advantaged groups and will not result in negative externalities that harm disadvantaged groups. Recent examples about COVID-19 data (D'Ignazio and Klein, 2020; Taylor, 2017) and predictive algorithms in public services (Eubanks, 2018) and in the online economy (Lee, 2018) have shown that social biases could be built into data practices, which, in turn, reinforce social inequalities. What is at stake here is to make explicit choices and assumptions about "algorithmic fairness" in processes of policy design (Mitchell *et al.*, 2021). Problematizing the notion of fairness in machine learning, Selbst *et al.* (2019) contend:

> Fairness and justice are properties of social and legal systems like employment and criminal justice, not properties of the technical tools within. To treat fairness and justice as terms that have meaningful application to technology separate from a social context is therefore to make a category error, or as we posit here, an abstraction error. (p. 59)

To put it simply, fairness has to do with specific policy goals, populations (both mathematically and politically) and outcomes (Mitchell *et al.*, 2021; Selbst *et al.*, 2019). Following our previous discussion of transparency and accountability, it is, therefore, important to incorporate citizens in the process of developing data practices and provide them with epistemic resources to understand and contest the inferences and decisions drawn from data. By acknowledging the limitations of data, it is vital to restrict the scope and outcomes of data practices (Mitchell *et al.*, 2021) *only* for public good purposes. Possible interventions include assessments of whether the technical design and implementation account for localized fairness and privacy concerns in a particular social context (Selbst *et al.*, 2019). Similarly, Wong (2020) suggests that any algorithmic fairness is a political question that requires decision-making processes to be publicly accessible, provide a "reasonable" explanation and allow for revisions.

The third political principle is *the principle of democratic legitimacy*. In a democratic society, the state's collection and uses of big data can only be legitimate when they are democratically authorized. Given that today's governments are increasingly reliant on big data for governance (Desrosières, 2002), it is even more urgent to avoid the state becomes a technocracy (Habermas, 2015) in which political problems are deemed the area belongs to political experts who are capable of understanding and harnessing the power of big data. An ability to *see* processes of data collection is not equated with an ability to know how they work and should be regulated (Ananny and Crawford, 2018; Kemper and Kolkman, 2019). In the case of the UK Government's 2050 Energy Calculator, Kemper and Kolkman (2019) argue that open-sourcing the technical documentation and algorithmic model might signify a transparent process to legitimize data practices, which, in turn, could also lead non-experts to adopt a less critical assessment of the system. Therefore, the principle requires not only democratic authorization but also the massive nurturing of critical audiences' data literacy. A democratic people cannot hold the state accountable to its data abuse and cannot meaningfully authorize the state's collection and uses of big data without understanding what big data is, how it operates and what the ethical implications of data are.

Following Sander (2020), we conceptualize critical big data literacy as "citizens' *awareness*, *understanding* and *critical* reflection of big data practices and their risks and implications, as well as the ability to *implement this knowledge* for a more empowered interest usage" (p. 14; original emphasis; Fotopoulou, 2020). Using snowball sampling, Sander (2020) identified about 40 English-language data literacy tools (e.g. "tactical technology collective" for citizens and civil society organizations), but found that many of these tools might "aim at already interested individuals with a pronounced prior knowledge on issues related to big data [. . .] or those planning to teach about big data" (pp. 9–10). Similarly, there are existing skills-focused educational projects that teach governments and international organizations about the management of open data such as the World Bank's Open Data Essentials (Fotopoulou, 2020). Yet, as Fotopoulou (2020) reminds us, critical data literacy training must account for these organizations' institutional and economic needs.

Overall, it is worth highlighting that the principle of democratic legitimacy is distinct from the principle of transparency and accountability. The latter principle concerns mainly with whether there are enough checks and balances to hold the state accountable to its data practices and whether the state's handling of data is transparent so that accountability mechanisms could be more effectively enforced. The policy implication of this principle is chiefly about strengthening the legal and material resources available to data activists and civil society groups. The principle of democratic legitimacy, however, asks whether the state's data practices have been meaningfully authorized by general citizens, not just civil society groups and activists. The policy implication, therefore, requires a larger scope of effort because such principle essentially points to the need for the state to actively nurture the data literacy and awareness of general citizens to enable a meaningful democratic authorization.

4. Conclusion

In the growing trend of datafication, when the state's ability to deliver goods and services hinges increasingly on its ability to leverage data (Pistor, 2020), the question of whether the state could have a normative right to collect data and what are the constraints over the state's data practices become important political questions (van Dijck, 2014; Lyon, 2014). This article, by drawing on major normative political theories of the public good, argues that market failure, basic rights protection and deepening democracy can be normative grounds to justify the state's right to data collection. This framework is intended to *restrict* rather than expand the state’s data practices because the framework suggests that the state’s collection and uses of big data could be justifiable only when they are for public good purposes. The article further argues that the public good framework entails at least three guiding political principles regulating the state’s data practices, including

(1) The principle of transparency and accountability, which stresses the importance of transparent data practices and the role of civil society to the accountability of the state.

(2) The principle of fairness, which emphasizes the significance of the distributive effects of the state’s data practices.

(3) The principle of democratic legitimacy, which highlights the fact that democratic participation in data-related policies requires a certain degree of data literacy, and it further implies that the state has a duty to actively cultivate such citizen ability especially in the age of datafication, where social interactions are increasingly mediated through data.

The public good framework and three guiding principles of justice represent a first step toward conceptualizing the moral and ethical dimensions of the state’s data governance and practices in a democratic context. Future research should account for specific historical, cultural and social contexts where data practices take place to avoid data universalism (Milan and Treré, 2019). Future research should also consider what concrete institutional arrangements can help to apply the three principles in practice.

Notes

1. One notable move in the scholarly field is the annual association for computing machinery (ACM) conference on fairness, accountability and transparency since 2018 (ACM FAccT after 2019; formerly known as fairness, accountability, and transparency* in 2018).

2. Kohn’s (2021) article also mentions a fourth approach, which she names the solidaristic approach to the public good. This article leaves aside the solidaristic approach as it has yet to become one of the dominant approaches in political theory and this approach, at least in the context of her article, is primarily used for demonstrating how the three major approaches fail to capture physical public space as a public good, which is a question beyond the scope of this article.

References

Acquisti, A. (2014), “The economics and behavioral economics of privacy”, in Lane, J., Stodden, V., Bender, S. and Nissenbaum, H. (Eds), *Privacy, Big Data, and the Public Good: Frameworks for Engagement*, Cambridge University Press, New York, NY, pp. 76-95.

Akman, I., Yazici, A., Mishra, A. and Arifoglu, A. (2005), “E-Government: a global view and an empirical evaluation of some attributes of citizens”, *Government Information Quarterly*, Vol. 22 No. 2, pp. 239-257.

Ananny, M. and Crawford, K. (2018), “Seeing without knowing: limitations of the transparency ideal and its application to algorithmic accountability”, *New Media and Society*, Vol. 20 No. 3, pp. 973-989.

Andejevic, M. (2014), “The big data divide”, *International Journal of Communication*, Vol. 8, pp. 1673-1689.

Barocas, S. and Selbst, A.D. (2016), “Big data’s disparate impact”, *California Law Review*, Vol. 104 No. 3, pp. 671-732.

Batina, R.G. and Ihori, T. (2005), *Public Goods: Theories and Evidence*, Springer, Berlin.

Beraldo, D. and Milan, S. (2019), "From data politics to the contentious politics of data", *Big Data and Society*, Vol. 6 No. 2, available at: https://doi.org/10.1177/2053951719885967 (accessed 21 February 2021).

Bigo, D., Isin, E. and Ruppert, E. (2019), "Data politics", in Bigo, D., Isin, E. and Ruppert, E. (Eds), *Data Politics: Worlds, Subjects, Rights*, Routledge, London, pp. 1-18.

boyd, d. and Crawford, K. (2012), "Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon", *Information, Communication and Society*, Vol. 15 No. 5, pp. 662-679.

Burrell, J. (2016), "How the machine 'thinks': understanding opacity in machine learning algorithms", *Big Data and Society*, Vol. 3 No. 1, available at: https://doi.org/10.1177/2053951715622512 (accessed 21 February 2021).

Claassen, R. (2018), *Capabilities in a Just Society: A Theory of Navigational Agency*, Cambridge University Press, New York, NY.

Cobain, I. (2018), "UK has six months to rewrite snooper's charter, high court rules", The Guardian, 27 April.

Couldry, N. and Mejias, U.A. (2019), *The Costs of Connection: How Data is Colonizing Human Life and Appropriating It for Capitalism*, Standard University Press, Standard, CA.

Currie, M. (2020), "Data as performance – showcasing cities through open data maps", *Big Data and Society*, Vol. 7 No. 1, available at: https://doi.org/10.1177/2053951720907953 (accessed 21 February 2021).

D'Ignazio, C. and Klein, L.F. (2020), "Seven intersectional feminist principles for equitable and actionable COVID-19 data", *Big Data and Society*, available at: https://doi.org/10.1177/2053951720942544 (accessed 21 February 2021).

Deibert, R.J. (2020), "We've become dependent on a technological ecosystem that is highly invasive and prone to serial abuse", The Globe and Mail, 23 November.

Dencik, L., Hintz, A. and Cable, J. (2016), "Towards data justice? The ambiguity of anti-surveillance resistance in political activism", *Big Data and Society*, Vol. 3 No. 2, available at: https://doi.org/10.1177/2053951716679678 (accessed 20 November 2020).

Desrosières, A. (2002), *The Politics of Large Numbers: A History of Statistical Reasoning*, Harvard University Press, Cambridge, MA.

Dworkin, R. (2002), *Sovereign Virtue: The Theory and Practice of Equality*, Harvard University Press, Cambridge, MA.

Eubanks, V. (2018), *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, St. Martin's Press, New York, NY.

Fourcade, M. and Healy, K. (2017), "Seeing like a market", *Socio-Economic Review*, Vol. 15 No. 1, pp. 9-29.

Fotopoulou, A. (2020), "Conceptualising critical data literacies for civil society organisations: agency, care, and social responsibility", *Information, Communication and Society*, available at: https://doi.org/10.1080/1369118X.2020.1716041 (accessed 21 February 2021).

Gabrys, J., Pritchard, H. and Barratt, B. (2016), "Just good enough data: figuring data citizenships through air pollution sensing and data stories", *Big Data and Society*, Vol. 3 No. 2, available at: https://doi.org/10.1177/2053951716679677 (accessed 21 February 2021).

GDPR.EU (2021), "What is GDPR, the EU's new data protection law?", available at: https://gdpr.eu/what-is-gdpr/ (accessed 20 November 2020).

Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S. and Brilliant, L. (2009), "Detecting influenza epidemics using search engine query data", *Nature*, Vol. 457 No. 7232, pp. 1012-1014.

Goerge, R.M. (2014), "Data for the public good: challenges and barriers in the context of cities", in Lane, J., Stodden, V., Bender, S. and Nissenbaum, H. (Eds), *Privacy, Big Data, and the Public Good: Frameworks for Engagement*, Cambridge University Press, New York, NY, pp. 153-172.

Graeff, E. (2018), "Evaluating civic technology design for citizen empowerment", unpublished doctoral dissertation, MIT, Cambridge, MA.

Habermas, J. (1994), "Three normative models of democracy", *Constellations*, Vol. 1 No. 1, pp. 1-10.

Habermas, J. (2015), *The Lure of Technocracy*, Polity, Cambridge.

Hacking, I. (2015), "Biopower and the avalanche of printed numbers", in Cisney, V.W. and Morar, N. (Eds), *Biopower: Foucault and Beyond*, University of Chicago Press, Chicago, IL, pp. 65-80.

Hargittai, E., Redmiles, E.M., Vitak, J. and Zimmer, M. (2020), "Americans' willingness to adopt a COVID-19 tracking app: the role of app distributor", *First Monday*, available at: https://firstmonday.org/ojs/index.php/fm/article/download/11095/9985 (accessed 21 February 2021).

Heeks, R. and Shekhar, S. (2019), "Datafication, development and marginalised urban communities: an applied data justice framework", *Information, Communication and Society*, Vol. 22 No. 7, pp. 992-1011.

Hintz, A. and Brown, I. (2017), "Enabling digital citizenship? The reshaping of surveillance policy after Snowden", *International Journal of Communication*, Vol. 11, pp. 782-801.

Ience, M. and Vayena, E. (2020), "On the responsible use of digital data to tackle the COVID-19 pandemic", *Nature Medicine*, Vol. 26, pp. 464-464.

Information Commissioner's Office (2021), "Principle (a): lawfulness, fairness and transparency", available at: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/lawfulness-fairness-and-transparency/ (accessed 21 February 2021).

Kemper, J. and Kolkman, D. (2019), "Transparency to whom? No algorithmic accountability without a critical audience", *Information, Communication and Society*, Vol. 22 No. 14, pp. 2081-2096.

Kitchin, R. (2014), "Big data, new epistemologies and paradigm shifts", *Big Data and Society*, Vol. 1 No. 1, available at: https://doi.org/10.1177/2053951714528481 (accessed 21 February 2021).

Kohn, M. (2021), "Public goods and social justice", *Perspectives on Politics*, Vol. 18 No. 4, pp. 1104-1117.

Lane, J., Stodden, V., Bender, S. and Nissenbaum, H. (Eds) (2014), *Privacy, Big Data, and the Public Good: Frameworks for Engagement*, Cambridge University Press, New York, NY.

Lau, J. (2021), "A trust deficit is hindering Hong Kong's COVID-19 response", The Diplomat, 22 February.

Lee, N.T. (2018), "Detecting racial bias in algorithms and machine learning", *Journal of Information, Communication and Ethics in Society*, Vol. 16 No. 3, pp. 252-260.

Lepri, B., Oliver, N., Letouze, E., Pentland, A. and Vinck, P. (2018), "Fair, transparent, accountable algorithmic decision-making processes: the premise, the proposed solutions, and the open challenges", *Philosophy and Technology*, Vol. 31 No. 4, pp. 611-627.

Lyon, D. (2014), "Surveillance, Snowden, and big data: capacities, consequences, critique", *Big Data and Society*, Vol. 1 No. 2, available at: https://doi.org/10.1177/2053951714541861 (accessed 21 February 2021).

Milan, S. (2020), "Techno-solutionism and the standard human in the making of the COVID-19 pandemic", *Big Data and Society*, Vol. 7 No. 2, available at: https://doi.org/10.1177/2053951720966781 (accessed 21 February 2021).

Milan, S. and Treré, E. (2019), "Big data from the South(s): beyond data universalism", *Television and New Media*, Vol. 20 No. 4, pp. 319-335.

Mitchell, S., Potash, E., Barocas, S., D'Amour, A. and Lum, K. (2021), "Algorithmic fairness: choices, assumptions, and definitions", *Annual Review of Statistics and Its Application*, Vol. 8 No. 1, pp. 12.1-12.23.

Nissenbaum, H. (2010), *Privacy in Context: Technology, Policy, and the Integrity of Social Life*, Stanford University Press, Stanford, CA.

Nissenbaum, H. (2017), "Deregulating collection: must privacy give way to use regulation?", available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3092282 (accessed 21 February 2021).

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

Pistor, K. (2020), "Statehood in the digital age", *Constellations*, Vol. 27 No. 1, pp. 3-18.

Pollman, E. and Barry, J. (2017), "Regulatory entrepreneurship", *Southern California Law Review*, Vol. 90 No. 3, pp. 383-448.

Poom, A., Järv, O., Zook, M. and Toivonen, T. (2020), "COVID-19 is spatial: ensuring that mobile big data is used for social good", *Big Data and Society*, Vol. 7 No. 2, available at: https://doi.org/10.1177/2053951720952088 (accessed 21 February 2021).

Raley, R. (2013), "Dataveillance and countervailance", in Gitelman, L. (Ed.), *"Raw Data" is an Oxymoron*, MIT Press, Cambridge, MA, pp. 121-146.

Rawls, J. (2001), *Justice as Fairness: A Restatement*, Harvard University Press, Cambridge, MA.

Rogerson, S., Miller, K.W., Winter, J.S. and Larson, D. (2017), "Information systems ethics – challenges and opportunities", *Journal of Information, Communication and Ethics in Society*, Vol. 17 No. 1, pp. 87-97.

Sander, I. (2020), "What is critical big data literacy and how can it be implemented?", *Internet Policy Review*, Vol. 9 No. 2, pp. 1-22.

Schrock, A. and Shaffer, G. (2017), "Data ideologies of an interested public: a study of grassroots open government data intermediaries", *Big Data and Society*, Vol. 4 No. 1, available at: https://doi.org/10.1177/2053951717690750 (accessed 21 February 2021).

Schumpeter, J.A. (2008), *Capitalism, Socialism, and Democracy*, Harperperennial, New York, NY.

Selbst, A.D., boyd, d., Friedler, S.A., Venkatasubramanian, S. and Vertesi, J. (2019), "Fairness and abstraction in sociotechnical systems", *FAT* '19: Proceedings of the Conference on Fairness, Accountability, and Transparency*, pp. 59-68.

Shi, Q. and Abdel-Aty, M. (2015), "Big data applications in real-time traffic operations and safety monitoring and improvement on urban expressways", *Transportation Research Part C: Emerging Technologies*, Vol. 58, pp. 380-394.

Shue, H. (1996), *Basic Rights: Subsistence, Affluence, and U.S. Foreign Policy*, (2nd ed.), Princeton University Press, Princeton, NJ.

Smith, G. (2010), *Democratic Innovations: Designing Institutions for Citizen Participation*, Cambridge University Press, New York, NY.

Taylor, L. (2017), "What is data justice? The case for connecting digital rights and freedom globally", *Big Data and Society*, Vol. 4 No. 2, available at: https://doi.org/10.1177/2053951717736335 (accessed 21 February 2021).

The Data Retention and Acquisition Regulations (2018), (SI 2018//123), available at: www.legislation.gov.uk/uksi/2018/1123/made/data.htm (accessed 21 February 021).

UK Government (2016), "Factsheet – communications data: investigatory powers bill", available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/473747/Factsheet-Communications_Data_General.pdf (accessed 21 February 2021).

van Dijck, J. (2014), "Datafication, dataism and dataveillance: big data between scientific paradigm and ideology", *Surveillance and Society*, Vol. 12 No. 2, pp. 197-208.

Veljković, N., Bogdanović-Dinić, S. and Stoimenov, L. (2014), "Benching marking open government: an open data perspective", *Government Information Quarterly*, Vol. 31 No. 2, pp. 287-290.

Vitak, J. and Zimmer, M. (2020), "More than just privacy: using contextual integrity to evaluate the long-term risks from COVID-19 surveillance technologies", *Social Media + Society*, available at: https://doi.org/10.1177/2056305120948250 (accessed 21 February 2021).

Wang, J., Ng, C.Y. and Brook, R.H. (2020), "Response to COVID-19 in Taiwan: big data analytics, new technology, and proactive testing", *JAMA*, Vol. 323 No. 14, pp. 1341-1342.

Wachter, S. and Mittelstadt, B. (2019), "A right to reasonable inferences: re-thinking data protection law in the age of big data and AI", *Columbia Business Law Review*, Vol. 2, pp. 494-620.

Wachter, S., Mittelstadt, B. and Russell, C. (2018), "Counterfactual explanations without opening the black box: automated decisions and the GDPR", *Harvard Journal of Law and Technology*, Vol. 31 No. 2, pp. 841-887.

Wieringa, M. (2020), "What to account for when accounting for algorithms: a systematic literature review on algorithmic accountability", *FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 1-18.

Wong, P.H. (2020), "Democratizing algorithmic fairness", *Philosophy and Technology*, Vol. 33 No. 2, pp. 225-244.

Zingales, L. (2017), "Towards a political theory of the firm", *Journal of Economic Perspectives*, Vol. 31 No. 3, pp. 113-130.

Zuboff, S. (2019), *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, Public Affairs, New York, NY.

Zuckerman, E. (2020), "What is digital public infrastructure", center for journalism and liberty", available at: https://static1.squarespace.com/static/5efcb64b1cf16e4c487b2f61/t/5fb41b6aac578321b0c50717/1605639019414/zuckerman-digital-infrastructure-cjl-nov2020.pdf (accessed 21 February 2021).

About the authors

Chi Kwok is an Assistant Professor (corresponding author) in the Department of Political Science at Lingnan University. Prior to joining Lingnan University, he was a postdoctoral researcher at Utrecht University. His current research examines theories of the corporation and the political legitimacy of corporate power. His work has appeared in journals, such as, among others, *Philosophy and Social Criticism*, *Review of Social Economy*, *Information, Communication and Society* and *China Perspectives*. Chi Kwok is the corresponding author and can be contacted at: c.kwok@uu.nl or chikwok@ln.edu.hk

Ngai Keung Chan is an Assistant Professor at the School of Journalism and Communication at The Chinese University of Hong Kong. His research examines the intersection of platform governance, algorithms, and service work.. His work has appeared in such journals as *New Media and Society*, *Surveillance and Society* and *Space and Culture* and *Media and Communication*.