The Political Philosophy of AI

An Introduction

Mark Coeckelbergh

polity

First published in 2022 by Polity Press

Polity Press
65 Bridge Street
Cambridge CB2 1UR, UK

Polity Press
101 Station Landing
Suite 300
Medford, MA 02155, USA

ISBN-13: 978-1-5095-4853-8
ISBN-13: 978-1-5095-4854-5(pb)

A catalogue record for this book is available from the British Library.

Library of Congress Control Number: 2021941737

Typeset in 10.5 on 12pt Sabon
by Fakenham Prepress Solutions, Fakenham, Norfolk NR21 8NL
Printed and bound in Great Britain by CPI Group (UK) Ltd, Croydon

For further information on Polity, visit our website:
politybooks.com

5

Power: Surveillance and (Self-) Disciplining by Data

Introduction: Power as a topic in political philosophy

One way to talk about politics is to use the concept of power. Power is often seen in negative terms or as a representation of how things really are as opposed to an ideal. For example, power has been invoked in response to those who defend deliberative and participative ideals of liberal democracy. Consider Dewey's ideal of participative democracy again. Critics have argued that this ideal is naïve since it avoids talking about conflict and power. In particular, it is deemed to be too optimistic about the capacity of ordinary citizens to judge and act intelligently and about the chances to reach consensus, thereby ignoring what Hildreth (2009) calls the "darker forces in human nature, including a thirst for power and the willingness to manipulate social relations to one's advantage" (781). Not so long after Dewey, Mills wrote in *The Power Elite* (1956) that American society is ruled by people in corporations, the military, and the government who "are in command of the major hierarchies and organizations of modern society" (4) and have access to the power and wealth available there. Instead of citizens "held in responsible check by a plurality of voluntary associations which connect debating publics with the pinnacles of decisions," as

defenders of participative democracy may imagine things should work, Mills saw a "system of organized irresponsibility" (361) run by an elite. Public problem-solving as Dewey imagined it does not work on a large scale. Politics requires the fight for power, and cannot be modeled on scientific models of problem-solving. Dewey mistakenly ignored how power is distributed in society and how deeply divided societies can be. As we have seen in the previous chapter, this criticism is also in line with Mouffe and Rancière, who propose to instead examine power as dissensus and agonism. And Marxism questions the power distribution between social classes, emphasizing how capital gives power to those who own it. In both cases, power is linked to struggle, which can be used productively under specific conditions.

Another example of power versus ideals, which is directly relevant to AI, is power versus freedom as consent. In the United States and in Europe, clicking that you agree to a particular internet platform's terms of service – including its data processing policy and hence the way AI is put to work – is meant to protect the rights of consumers, including their freedom. However, as Bietti (2020) has argued, this regulatory device fails to account for the unjust background conditions and power structures under which these individual consent acts take place. If power imbalances are "shaping the environment in which a decision to consent is made," then consent is "an empty construct" (315). Power is also seen as a danger to truth (Lukes 2019) and potentially deceptive. Power can be used for coercion, for example in the context of a totalitarian state. But it can also take the form of various forms of manipulation. This threatens reasoning and the development of critical capacities. Thinking is also difficult when you are constantly in a competitive environment where you cannot afford the time to slow down (Berardi 2017). Power is then seen as the enemy of thinking itself.

Yet power is not necessarily bad. An influential and arguably more complex view of power is offered by Foucault. Inspired by Nietzsche, Foucault (1981, 93–4) conceptualized society in terms of power, in particular force relations. But his view differs considerably from Marxism. Instead of analyzing power top down, in terms of centralized sovereignty and the power of rulers or elites, he proposes a bottom-up approach starting from the small mechanisms and operations of power that shape subjects, produce particular kinds of bodies, and pervade the

whole of society. He analyzes these micro-mechanisms of power in prisons and hospitals. Instead of linking power to the head of Leviathan, the central authoritarian sovereign in Hobbes's thinking, Foucault (1980) focuses on the plurality and the body of power: on "the myriad of bodies which are constituted as peripheral subjects as a result of the effects of power" (98), on power's "infinitesimal mechanisms" (99). Power is exercised within the social body "rather than from above it" (39); it "circulates" (98) through the social body (119). Moreover, Foucault is interested in how power "reaches into the very grain of individuals, touches their bodies and inserts itself into their actions and attitudes, their discourses, learning processes and everyday lives" (39). Individuals are not only the points on which power works; instead, they simultaneously exercise and undergo power; they are "the vehicles of power, not its points of application" (98). The individual is the effect of power.

What do these different views of power imply for the politics of AI? Is AI used by those who manipulate social relations to their advantage and deceive us? And how does it interact with the micro-mechanisms of power that Foucault describes? What kind of individuals, subjects, and bodies are made by means of AI? In this chapter, I ask these questions and apply political and social theory of power to AI. First, I will use a general conceptual framework about power and technology as developed by Sattarov in order to distinguish between various ways AI may impact power. Then I will draw on three theories of power in order to elaborate on some relations between AI and power: Marxism and critical theory, Foucault and Butler, and a performance-oriented approach as proposed in my own work. This will lead to a conclusion about what I will call "artificial power" (the initial title of this book).

Power and AI: Towards a general conceptual framework

The relation between politics and technology is by now a well-known topic in contemporary philosophy of technology. Consider Winner's (1980) work, which shows that technologies can have unintended political consequences, and Feenberg's (1991) critical theory of technology, which is not only inspired

by Marx and critical theory (in particular Marcuse) but also empirically oriented. However, while elsewhere there is much interest in *power*, for example in cultural studies, gender studies, posthumanism, and so on, there have been few systematic philosophical treatments and overviews of the topic in philosophy of technology. In ethics of computing there is work on the power of algorithms (Lash 2007; Yeung 2016), but a systematic framework to think about power and technology was lacking for decades. An exception is Sattarov's *Power and Technology* (2019), which distinguishes between different conceptions of power and applies them to technology. While his contribution is mainly geared towards technology ethics, rather than political philosophy of technology, it is very helpful for the purpose of analyzing the relations between AI and power.

Sattarov distinguishes between four conceptions of power. The first conception, which he calls *episodic*, is about relationships in which one actor exercises power over another, for instance by means of seduction, coercion, or manipulation. The second conception defines power as a *disposition*: as a capacity, ability, or potential. The third, systematic conception, understands power as a property of social and political *institutions*. The fourth conception sees power as *constituting or producing* the social actors themselves (Sattarov 2019, 100). The latter two are thus more structural, whereas the first are about actors and their actions (13).

Following Sattarov, we can map these different conceptions of power onto relations between power and technology. First, technology can (help to) seduce, coerce, force, or manipulate people, and can be used to exercise authority. One could also say that this kind of power is delegated to the technology or – to pick up a concept that is often used in postphenomenology of technology – that the technology mediates. For example, online advertisements can seduce users to visit a website, speed bumps can force drivers to slow down, and technology can also manipulate. Technology can "nudge": it can change the choice architecture so that people are more likely to behave in certain ways, without them being aware of this (see also chapter 2). Second, technology can give power to people in the sense that it increases their abilities and potential for action; it can *em*power. This is also true for humanity in general, as Jonas (1984) has argued: technology has given humanity enormous power.

Consider the concept of the Anthropocene: humanity as a whole has become a kind of geological force (Crutzen 2006). It has acquired a hyper agency that has transformed the entire surface of the earth (see also the next chapter). Third, when it comes to systemic power, we can see how technology can support particular systems and ideologies. For example, from a Marxian view, technology supports the advancement of capitalism. Power here is not about what individuals do; rather, it is embedded in a particular political, economic, or social system, to which technology contributes. For example, mass media shape public opinion. This is also true for social media, which may support a particular political-economic system (e.g., capitalism). Finally, if power is not just something that is possessed or exercised by individuals, or applied to individuals, but is also constitutive of subjects, selves, and identities, as Foucault has argued, then technology can be used to constitute such subjects, selves, and identities. Often there is no intention on the part of technology developers and users to do this, but it may happen nevertheless. For example, social media may shape your identity, even if you are not aware of it.

What does this mean for thinking about power and AI?

First, AI can seduce, coerce, or manipulate, for example via social media and recommender systems. Like algorithms in general (Sattarov 2019, 100), AI can be designed to change the attitude and behaviors of users. Without using coercion, it can function as a "persuasive technology" (Fogg 2003) by seducing and manipulating people. Music recommender systems such as Spotify or sites like Amazon aim to steer the listening or buying behavior of people by nudging them through changing the decision environment (see also the previous chapters), for example by suggesting that other people with a similar taste in books have bought book x and book y. And the order of Facebook posts is decided by an algorithm, which can, for example, influence the feelings of users through processes of "contagion" (Papacharissi 2015). Individuals are clustered into groups based on similar interests and behavior, which may reproduce social stereotypes and reaffirm old power structures (Bartoletti 2020). People are also manipulated by means of dynamic pricing and other "personalization" techniques, exploiting individual decision-making vulnerabilities, including well-known biases (Susser, Roessler, and Nissenbaum 2019, 12).

As in all forms of manipulation, people are influenced to act in a certain way without them being aware of this influencing. As we have seen, such covert influencing of individuals' decision-making threatens freedom, understood as individual autonomy. To the extent that this happens, we are no longer in control of our choices, and do not even understand the underlying mechanisms of how this happens. While the modern conception of autonomy, according to which we are or should be atomistic and rationalistic individuals, is not adequate and has been criticized inside and outside mainstream Western philosophy (see, e.g., discussions about relational autonomy in Christman 2004 and Westlund 2009), even as social and relational beings we want some control over our decisions and our lives, and we do not want to be manipulated. In terms of power, the mentioned forms of seduction and manipulation by AI shift the power balance (even more) to those who collect, own, and monetize our data. Moreover, particular groups in society (e.g., racist groups) may try to gain power by manipulating people on social media.

Second, AI may empower by increasing people's individual capacities. Consider, for example, natural language processing that helps translate and thus opens up new possibilities for individuals (as well as creating problems, e.g., deskilling and threats to privacy). But AI also increases the potential for exercising power over others, humans and non-humans, and in the end it increases humanity's power over the natural environment and the earth. Consider, for instance, search engines and social media, which may empower individuals who previously did not have access to this amount and bandwidth of information, and who might not have had a voice in classic media. But at the same time those search engines and the companies who offer them are given a lot of power: they shape information flows and hence play what is known as a gatekeeper role. In addition, the companies and their algorithms use personalization: they "filter information per individual," which introduces human and technical biases (Bozdag 2013, 1). This gatekeeping role and these biases have implications for democracy and diversity (Granka 2010). As we already saw in chapter 3, in terms of power, AI serves here the interests of some people rather than others. AI can also be used at the level of the state to enable surveillance and its authoritarian use. It offers governments and their intelligence agencies power in the sense of new instruments and capacities for surveillance,

which can lead to enhanced oppression and even totalitarianism. Sometimes states and private companies team up to increase those capacities, as in China and the US. The corporate tech sector knows a lot about the lives of citizens (Couldry and Meijas 2019, 13). Even liberal democracies are installing facial recognition systems, use predictive policing, and employ AI tools at their borders. There is the risk that what Sætra (2020, 4) calls a new form of "algorithmic governance" will order "human action in general." Moreover, AI also empowers humanity as a whole, which may have consequences for non-humans such as animals and for natural environments. If, in the context of the Anthropocene, AI further increases humanity's ability to intervene in, and transform, nature, then it further supports an ongoing shift in terms of power: from non-humans to humans. Consider AI that helps to extract natural resources from the earth and the energy consumption by AI technology (see chapter 6), which in turn also requires the use of natural resources. That AI affords power to humans may be empowering at the individual level, but may have vast consequences for non-human nature, given the increased Baconian powers of humanity to mine and transform the earth: scientific knowledge and technologies are used to control nature. In the next chapter, I will say more about these non-human and earthly aspects of the politics and power of AI.

Third, AI can support neoliberal versions of capitalism, authoritarianism, and other systems and ideologies. Software and hardware systems related to AI "form part of the broader social, economic, and political institutional reality" (Sattarov 2019, 102), and this includes socio-economic systems and ideologies. Those larger systems affect the development of technology, for example by creating a context of investment in AI, but technology may also help to maintain those systems. For example, Dyer-Witheford, Kjøsen, and Steinhoff (2019) claim that AI is an instrument of capital, and therefore entails exploitation and the concentration of power in the hands of the owners of high tech – who are in turn concentrated in particular countries and regions such as the US (Nemitz 2018). Thus, AI is not just technological but also creates or maintains a particular social order, here capitalism and neoliberalism. Consider also again Zuboff's claim about surveillance capitalism: the point is not only that a particular technology is problematic; AI and big data help to

create, maintain, and expand an entire socio-economic system in which capital is accumulated (by some) through technologies that harvest and sell data (of the many), thus exploiting human nature and reaching into the intimate sphere. Even our emotions are monitored and monetized (McStay 2018). The same AI technology can be used to support totalitarian regimes or to maintain oppressive political systems and their corresponding narratives and images (e.g., a racist utopia), although in principle AI could also offer opportunities for supporting democracy – much depends on how we think of democracy (see chapter 4) and indeed about politics.

Most researchers in AI and the politics of AI support a democratic and fair way of coding. Some believe we need more limits and regulation. Oppressive effects are not always and not usually intended, although sometimes AI is used on purpose to promote a racist and nationalist politics. Yet as we have seen in chapter 3 and 4, there are also problematic non-intended effects. AI may support racist and neo-colonial political cultures and systems by introducing bias against particular individuals and groups, or help to create the conditions for authoritarianism or totalitarianism. Consider again Noble's (2018) argument that (search) algorithms and classification systems may "reinforce oppressive social relationships" (1). An example of such "algorithmic oppression" (4) was the case of Google Photos that tagged African Americans as "apes" and "animals" – a problem Google could not really fix (Simonite 2018). However, whether a particular use or outcome of AI is biased or unjust is not always as clear as in this case, and also depends on one's conceptions of justice and equality (see chapter 3). In any case, decisions, thoughts, actions, and emotions can also be controlled on purpose in order to support a particular political system. In the case of totalitarianism, AI may support the system's unlimited reach into the minds and hearts of people.

Fourth, AI can play a role in the constitution of the self and the formation of the subject, even if we are not aware of it. The point here is not only that AI manipulates us and intervenes deeply at the personal level in the sense that it can help to deduce the thoughts and feelings of people – making inferences about their inner states based on observable behaviors such as facial expressions and musical preferences, which are then used for predication and monetization by surveillance capitalism – but

also that AI contributes to the shaping of how we understand and experience ourselves. While what Rouvroy (2013) calls "algorithmic governmentality" bypasses "any encounter with human reflexive subjects" (144), does not allow room for human judgments and explicit evaluations of our beliefs and ourselves, and leads to exploitation of relations between individuals (Stiegler 2019), this does not mean that there is no effect on our self(-knowledge). What kind of perception and knowledge of the self does AI help to create? For example, do we start to understand ourselves as producers and collections of data for sale? Do we quantify ourselves and our lives as we track ourselves and are tracked by others? Do we think that we have "data doubles" (Lyon 2014), digital models of ourselves – even if AI does not store a digital model of the user (Matzner 2019)? Do we acquire and communicate a networked sense of self (Papacharissi 2011)? What sort of identity and subjectivity is enabled by AI?

Asking such questions goes beyond an instrumentalist understanding of human–technology relations. The self and human subjectivity are not external to information technologies such as AI; instead, "digital technology does something to human subjectivity itself" (Matzner 2019, 109). AI technologies have impacted the way we perceive and act in the world, leading to new forms of subjectivity (118). And there are different forms of subjectivity connected to AI. For example, based on the kinds of subjects we are and the kind of communities we belong to, we will react differently to a particular AI-based security system. If someone is not recognized by the system, this may be perceived as threatening by one person, given personal previous experiences and tensions in a particular social context (e.g., racism that has impacted that person and community), whereas another person from another background might have fewer problems with it. In Matzner's (2019) words: "[S]pecific applications of AI connect in quite different manners to pre-existing socio-technical situations and the respective forms of subjectivity" (109). AI technology will enable different relations to different subjectivities because we are situated subjects (118). In line with Foucault's view, this means that the power of AI is not just about (top-down) manipulations, capacities, and systems; it is also about concrete, situated experiences and mechanisms of power, shaped by the technology. One could also say (as I will do at the end of this chapter): as living, moving, and situated beings,