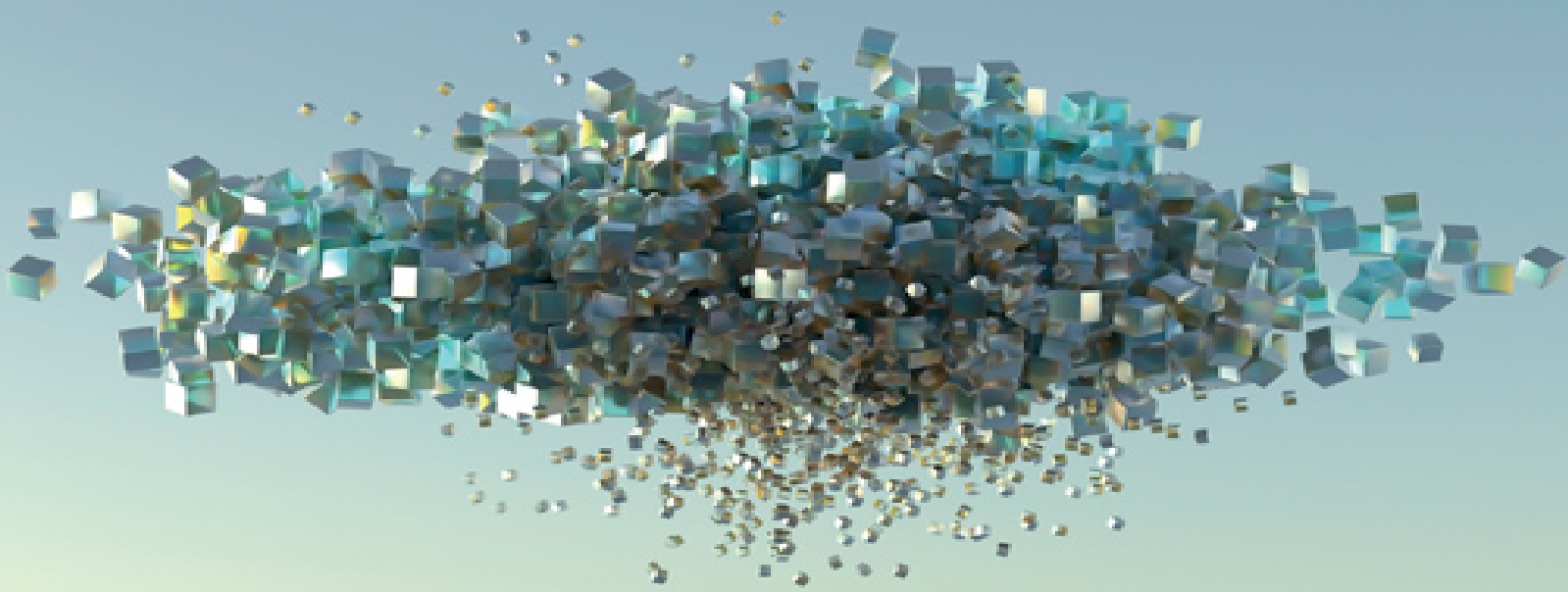

THE POLITICAL PHILOSOPHY OF AI

Mark Coeckelbergh

Table of Contents

- Cover
- Title Page
- Copyright Page
- Acknowledgments
- 1 Introduction
  - Rationale, aims, and approach of this book
  - Structure of the book and overview of its chapters
- 2 Freedom: Manipulation by AI and Robot Slavery
  - Introduction: Historical declarations of liberty and contemporary slavery
  - AI, surveillance, and law enforcement: Taking away negative freedom
  - AI and the steering of human behavior: Circumventing human autonomy
  - Threats to self-realization and emancipation: Exploitation by means of AI and the problem with robot slaves
  - Who decides about AI? Freedom as participation, AI in elections, and freedom of speech
  - Other politically relevant notions of freedom and other values
- 3 Equality and Justice: Bias and Discrimination by AI
  - Introduction: Bias and discrimination as a focus for raising problems concerning equality and justice
  - Why is bias wrong (1)? Equality and justice in standard anglophone liberal political philosophy

- Why is bias wrong (2)? Class and identity theories as criticisms of universalist liberal thinking
- Conclusion: AI is not politically neutral

4 Democracy: Echo Chambers and Machine Totalitarianism

- Introduction: AI as a threat to democracy
- AI as a threat to democracy, knowledge, deliberation, and politics itself
  - *Starting with Plato: Democracy, knowledge, and expertise*
  - *Beyond majority rule and representation*
  - *Deliberative and participatory democracy versus agonistic and radical democracy*
  - *Information bubbles, echo chambers, and populism*
  - *More problems: Manipulation, replacement, accountability, and power*
- AI and the origins of totalitarianism: Lessons from Arendt
  - *AI and totalitarianism*
  - *Arendt on the origins of totalitarianism and the banality of evil*

5 Power: Surveillance and (Self-)Disciplining by Data

- Introduction: Power as a topic in political philosophy
- Power and AI: Towards a general conceptual framework
- Marxism: AI as a tool for technocapitalism
- Foucault: How AI subjects us and makes us into subjects
  - *Disciplining and surveillance*

Knowledge, power, and the making and shaping of subjects and selves

Technoperformances, power, and AI

Conclusion and remaining questions

6 What about Non-Humans? Environmental Politics and Posthumanism

Introduction: Beyond a human-centered politics of AI and robotics

Not only humans count, politically: The political status of animals and (non-human) nature

Implications for the politics of AI and robotics

The political significance of the impact of AI on non-humans and natural environments

Political status for AI itself?

7 Conclusion: Political Technologies

What I have done in this book and what we can conclude

What needs to be done next: The question regarding political technologies

References

Index

End User License Agreement

The Political Philosophy of AI

An Introduction

Mark Coeckelbergh

polity

Copyright page

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

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

Acknowledgments

I wish to thank my editor, Mary Savigar, for her support and for guiding this book project to its successful conclusion, Justin Dyer for his careful editing, and Zachary Storms for assisting with the organizational aspects linked to the submission of the manuscript. I also thank the anonymous reviewers for their comments, which helped me to polish the manuscript. I am especially grateful to Eugenia Stamboliev for assisting with the literature search for this book. Finally, I warmly thank my family and friends – nearby and distant – for their support during these two difficult years.

1
Introduction

“I guess the computer got it wrong”: Josef K. in the 21st century

> Someone must have been telling tales about Josef K., for one morning, without having done anything wrong, he was arrested. (Kafka 2009, 5)

This is the first line of *The Trial* by Franz Kafka, originally published in 1925 and widely considered one of the most important novels of the 20th century. The protagonist of the story, Josef K., is arrested and prosecuted, but he does not know why. The reader is also left in the dark about this. Many explorations and encounters follow that only increase the opacity of it all, and after an unfair trial, Josef K. is executed with a butcher’s knife, “like a dog” (165). The story has been interpreted in many ways. One political take is that it shows how oppressive institutions can be and that its descriptions do not only reflect the rising power of modern bureaucracy but also prefigure the horrors of the Nazi regime that took place a decade later: people were arrested without having done anything wrong and sent to camps, facing various forms of suffering and often death. As Adorno put it: Kafka offers a “prophecy of terror and torture that was fulfilled” (Adorno 1983, 259).

Unfortunately, Kafka’s story is still relevant today. Not only because there are still opaque bureaucracies and oppressive regimes, which arrest people without justification and sometimes without trial, or because (as Arendt [1943] and Agamben [1998] already pointed out) refugees are often suffering a similar fate, but also because there is now a new way in which all this can happen,

indeed *has* happened, even in a so-called “advanced” society: one that has to do with technology, in particular with artificial intelligence (AI).

On a Thursday afternoon in January 2020, Robert Julian-Borchak Williams received a call in his office from the Detroit Police Department: he was asked to come to the police station to be arrested. Since he hadn’t done anything wrong, he didn’t go. An hour later he was arrested on his front lawn, in front of his wife and children, and, according to the *New York Times*: “The police wouldn’t say why” (Hill 2020). Later, in the interrogation room, detectives showed him an image from a surveillance video of a black man shoplifting from an upscale boutique and asked: “Is this you?” Mr. Williams, who was African American, responded: “No, this is not me. You think all black men look alike?” Only much later was he released, and in the end the prosecutor apologized.

What happened? The *New York Times* journalist and the experts she consulted suspect that “his case may be the first known account of an American being wrongfully arrested based on a flawed match from a facial recognition algorithm.” The facial recognition system, using AI in the form of machine learning, is faulty and most likely also biased: it works better for white men than for other demographics. The system thus creates false positives, like in the case of Mr. Williams, and, combined with bad police work, this results in people being arrested for crimes they didn’t commit. “I guess the computer got it wrong,” one of the detectives said. In the 21st-century United States, Josef K. is black and is falsely accused by an algorithm, without explanation.

The moral of the story is not only that computers make mistakes, mistakes that can have severe consequences for particular people and their families; the use of AI can also

worsen existing systemic injustices and inequalities, and in response to cases such as that of Mr. Williams, one could argue that all citizens should have a right to explanation when decisions are made about them. Moreover, this is just *one* of the many ways in which AI can have political significance and impact, sometimes intended but often unintended. This particular case raises questions concerning racism and (in)justice – two timely issues. But there is much more to say about the politics of AI and related technologies.

Rationale, aims, and approach of this book

While there is currently plenty of attention directed to *ethical* issues raised by AI and related technologies such as robotics and automation (Bartneck et al. 2021; Boddington 2017; Bostrom 2014; Coeckelbergh 2020; Dignum 2019; Dubber, Pasquale, and Das 2020; Gunkel 2018; Liao 2020; Lin, Abney, and Jenkins 2017; Nyholm 2020; Wallach and Allen 2009), there is very little work that approaches the topic from a *political-philosophical* angle. This is regrettable, since the topic lends itself perfectly well to such an investigation and leaves valuable intellectual resources from the political-philosophical tradition unused. From their side, most *political philosophers* have left the topic of the politics of AI untouched (exceptions are Benjamin 2019a; Binns 2018; Eubanks 2018; Zimmermann, Di Rosa, and Kim 2020), although in general there is a growing interest in the topic, for example in how algorithms and big data are used in ways that reinforce racism and various forms of inequality and injustice (e.g., Bartoletti 2020; Criado Perez 2019; Noble 2018; O'Neil 2016) and that extract and consume planetary resources (Crawford 2021).

Moreover, while in the current *political context* there is a lot of public attention directed to issues such as freedom, slavery, racism, colonialism, democracy, expertise, power, and climate, often these topics are discussed in a way that makes it seem as if they have little to do with technology and vice versa. AI and robotics are seen as technical subjects, and *if* a link to politics is made, technology is seen as a tool used for political manipulation or surveillance. Usually, the unintended effects remain unaddressed. On the other hand, *developers and scientists* working in the fields of AI, data science, and robotics are often willing to take ethical issues into account in their work, but are not aware of the complex political and societal problems these issues are connected to, let alone of the sophisticated political-philosophical discussions that could be held about the framing and addressing of these problems. Moreover, like most people not familiar with systematic thinking about technology and society, they tend to assume the view that technology itself is neutral and that everything depends on the humans developing and using it.

Questioning such a naïve conception of technology is the speciality of *philosophy of technology*, which in its contemporary form has advanced a non-instrumental understanding of technology: technology is not just a means to reach an end, but also shapes these ends (for an overview of some theories, see Coeckelbergh 2019a). However, when it comes to using philosophical frameworks and conceptual foundations for the normative evaluation of technology, philosophers of technology usually run to ethics (e.g., Gunkel 2014; Vallor 2016). Political philosophy is largely ignored. Only some philosophers make this connection: for example, in the 1980s and 1990s, Winner (1986) and Feenberg (1999), and today, Sattarov (2019) and Sætra (2020). More work is needed on the nexus between philosophy of technology and political philosophy.

This is an academic gap, but also a societal need. If we want to tackle some of the most pressing global and local issues of the 21st century such as climate change, global inequalities, aging, new forms of exclusion, war, authoritarianism, epidemics and pandemics, and so on, each of which is not only politically relevant but also related to technology in various ways, it is important to create a dialogue between thinking about politics and thinking about technology.

This book fills these gaps and responds to this rationale by

- connecting normative questions about AI and robotics to key discussions in political philosophy, using both the history of political philosophy and more recent work;
- addressing controversial issues that are at the center of current political attention, but now linking them to questions regarding AI and robotics;
- showing how this is not just an exercise in applied political philosophy but also leads to interesting insights into the often hidden and deeper political dimension of these contemporary technologies;
- demonstrating how the technologies of AI and robotics have both intended and unintended political effects, which can be helpfully discussed by using political philosophy;
- thereby making original contributions to both philosophy of technology and applied political philosophy.

The book thus uses political philosophy, alongside philosophy of technology and ethics, with the aims (1) to better understand normative issues raised by AI and robotics and (2) to shed light on pressing political issues and the way they are entangled with the use of these new

technologies. I use the term “entangled” here to express the close connection between political issues and issues concerning AI. The idea is that the latter is *already* political. The guiding concept of this book is that AI is not just a technical matter or just about intelligence; it is not neutral in terms of politics and power. AI is *political through and through*. In each chapter, I will show and discuss that political dimension of AI.

Rather than staging a discussion about the politics of AI in general, I will approach this overall theme by zooming in on specific topics that figure in contemporary political philosophy. Each chapter will focus on a particular political-philosophical set of themes: freedom, manipulation, exploitation, and slavery; equality, justice, racism, sexism, and other forms of bias and discrimination; democracy, expertise, participation, and totalitarianism; power, disciplining, surveillance, and self-constitution; animals, the environment, and climate change in relation to posthumanism and transhumanism. Each theme will be discussed in the light of the intended and unintended effects of AI, data science, and related technologies such as robotics.

As the reader will notice, this division in terms of topics and concepts is to some extent artificial; it will become clear that there are many ways in which the concepts, and hence the topics and chapters, interlink and interact. For example, the principle of freedom may be in tension with the principle of equality, and it is impossible to talk about democracy and AI without talking about power. Some of these connections will be made explicit in the course of the book; others are left to the reader. But all chapters show how AI impacts these key political issues and how AI is political.

However, this book is not only about AI but also about political-philosophical thinking itself. These discussions of the politics of AI will not only be exercises in applied philosophy – more specifically applied political philosophy – but will also feed back into the political-philosophical concepts themselves. They show how new technologies put our very notions of freedom, equality, democracy, power, and so on, into question. What do these political principles and political-philosophical concepts mean in the age of AI and robotics?

Structure of the book and overview of its chapters

The book is organized into seven chapters.

In chapter 2, I ask questions related to the political principle of freedom. What does freedom mean when AI offers new ways of making, manipulating, and influencing our decisions? How free are we when we do digital labour for large, powerful corporations? And does the replacement of workers by robots lead to the continuation of slavery thinking? The chapter is structured according to different conceptions of freedom. It discusses the possibilities offered by algorithmic decision-making and influencing by connecting to long-standing discussions about liberty in political philosophy (negative and positive liberty) and nudging theory. It points out how negative liberty can be taken away on the basis of an AI recommendation, questions how libertarian nudging by means of AI really is, and asks critical questions based on Hegel and Marx, showing how the meaning and use of robots risk remaining connected to a history and present of enslavement and capitalist exploitation. The chapter ends with a discussion of AI and freedom as political participation and freedom of speech, which is continued in chapter 4 on democracy.

Chapter 3 asks: what are the (usually unintended) political effects of AI and robotics in terms of equality and justice? Does the automation and digitalization enabled by robotics increase inequalities in society? Does automated decision-making by AI lead to unjust discrimination, sexism, and racism, as Benjamin (2019a), Noble (2018), and Criado Perez (2019) have argued, and, if so, why? Is the gendering of robots problematic, and how? What is the meaning of justice and fairness used in these discussions? This chapter puts the debates about automation and discrimination by AI

and robotics in the context of classical political-philosophical discussions about (in)equality and (in)justice as fairness in the liberal-philosophical tradition (e.g., Rawls, Hayek), but also connects to Marxism, critical feminism, and anti-racist and anti-colonial thinking. It raises questions concerning the tension between conceptions of universal justice versus justice based on group identity and positive discrimination, and discusses issues regarding inter-generational justice and global justice. The chapter ends with the thesis that AI algorithms are never politically neutral.

In chapter 4, I discuss the impacts of AI on democracy. AI can be used to manipulate voters and elections. Does surveillance by AI destroy democracy? Does it serve capitalism, as Zuboff (2019) has argued? And are we on our way to a kind of “data fascism” and “data colonialism”? What do we mean by democracy, anyway? This chapter puts the discussions about democracy and AI in the context of democracy theory, discussions about the role of expertise in politics, and work on the conditions for totalitarianism. First, it shows that while it is easy to see how AI can threaten democracy, it is much harder to make explicit what kind of democracy we want and what the role of technology is and should be in democracy. The chapter outlines tensions between Platonic-technocratic conceptions of politics and ideals of participative and deliberative democracy (Dewey and Habermas), which in turn have their critics (Mouffe and Rancière). It connects this discussion to issues such as information bubbles, echo chambers, and AI-powered populism. Second, the chapter argues that the problem of totalitarianism through technology points to deeper and long-standing problems in modern society such as loneliness (Arendt) and lack of trust. Ethical discussions, insofar as they focus on harm to individuals, neglect this broader societal and historical

dimension. The chapter ends by pointing to the danger of what Arendt (2006) called “the banality of evil” when AI is used as a tool for corporate manipulation and bureaucratic management of people.

Chapter 5 discusses AI and power. How can AI be used for disciplining and self-disciplining? How does it impact on knowledge and shift and shape existing power relations: between humans and machines but also between humans and even within humans? Who benefits from this? To raise these questions, the chapter connects back to discussions about democracy, surveillance, and surveillance capitalism, but also introduces Foucault’s complex view of power that highlights the micro-mechanisms of power at the level of institutions, human relationships, and bodies. First, the chapter develops a conceptual framework with which to think about relations between power and AI. Then it draws on three theories of power in order to elaborate on some of these relations: Marxism and critical theory, Foucault and Butler, and a performance-oriented approach. This enables me to shed light on the seductions and manipulations of and by AI, the exploitation and self-exploitation that it produces and its capitalist context, and the history of data science in terms of marking, classifying, and surveilling people. But it also points to ways in which AI may empower people and – through social media – play a role in the constitution of self and subjectivity. Moreover, it is argued that, by seeing what AI and humans do here in terms of technoperformances, we can point to the increasingly leading and more-than-instrumental role that technology plays in organizing the ways we move, act, and feel. I show that these exercises of (techno)power always have an active and social dimension, which involves both AI and humans.

In chapter 6, I introduce questions concerning non-humans. Like most ethics of AI, classic political discussions are human-centered, but this can and has been questioned in at

least two ways. First, are humans the only ones who count, politically? What are the consequences of AI for non-humans? And is AI a threat or an opportunity for dealing with climate change, or both? Second, can AI systems and robots themselves have political status, for example citizenship? Posthumanists question the traditional anthropocentric view of politics. Moreover, transhumanists have argued that humans will be superseded by superintelligent artificial agents. What are the political implications if a superintelligence takes over? Is this the end of human freedom, justice, and democracy? Opening up resources from animal rights and environmental theory (Singer, Cochrane, Garner, Rowlands, Donaldson and Kymlicka, Callicott, Rolston, Leopold, etc.), posthumanism (Haraway, Wolfe, Braidotti, Massumi, Latour, etc.), ethics of AI and robotics (Floridi, Bostrom, Gunkel, Coeckelbergh, etc.), and transhumanism (Bostrom, Kurzweil, Moravec, Hughes, etc.), this chapter explores conceptions of AI politics that go beyond the human. It argues that such a politics would require a rethinking of notions such as freedom, justice, and democracy to include non-humans, and would raise new questions for AI and robotics. The chapter ends with the claim that a non-anthropocentric politics of AI reshapes both terms of the human–AI relation: humans are not only de-powered and *em*powered by AI, but also give AI its power.

The concluding chapter summarizes the book and concludes that (1) the issues we currently care about in political and societal discussions such as freedom, racism, justice, and democracy take on a new urgency and relevance in the light of technological developments such as AI and robotics; and that (2) conceptualizing the politics of AI and robotics is not a matter of simply applying existing notions from political philosophy and political theory, but invites us to interrogate the very notions

themselves (freedom, equality, justice, democracy, etc.) and to ask interesting questions about the nature and future of politics and about ourselves as humans. The chapter also argues that, given the close entanglement of technology with societal, environmental, and existential-psychological changes and transformations, political philosophy in the 21st century can no longer evade what Heidegger (1977) called “the question concerning technology.” The chapter then outlines some further next steps that need to be taken in this domain. We need more philosophers working in this area and more research on the nexus of political philosophy/philosophy of technology, hopefully leading to a further “thinking together” (*zusammendenken*) of politics and technology. We also need more thinking about how to render the politics of AI more participatory, public, democratic, inclusive, and sensitive to global contexts and cultural differences. The book ends with the question: what *political technologies* do we need for shaping that future?

2
Freedom: Manipulation by AI and Robot Slavery

Introduction: Historical declarations of liberty and contemporary slavery

Freedom or liberty (I will use these terms interchangeably) is considered one of the most important political principles in liberal democracies, whose constitutions aim to protect basic liberties of citizens. For example, the First Amendment of the US Constitution, adopted in 1791 as part of the Bill of Rights, protects individual freedoms such as freedom of religion, freedom of speech, and freedom of assembly. Germany's constitution or Basic Law (*Grundgesetz*), adopted in 1949, states that the freedom of the person is inviolable (Article 2). Historically, the French Declaration of the Rights of Man and of the Citizen of 1789 is very influential. It is rooted in Enlightenment thinking (Rousseau and Montesquieu) and was developed at the time of the French Revolution in consultation with Thomas Jefferson: one of the founders of the United States and the principal author of the 1776 US Declaration of Independence, which already proclaimed in its preamble that "all men are created equal" and that they have "unalienable Rights," including "Life, Liberty and the pursuit of Happiness." Article I of the French Declaration says that "Men are born and remain free and equal in rights." While this Declaration still excluded women and did not forbid slavery, it was part of a history of declarations of rights and civil liberties that started in 1215 with Magna Carta (*Magna Carta Libertatum* or the great

charter of freedoms) and ended with the Universal Declaration of Human Rights (UDHR), adopted by the United Nations General Assembly in December 1948, which states that “All human beings are born free and equal in dignity and rights” (Article 1) and that “No one shall be held in slavery or servitude” (Article 4) (UN 1948).

Yet in many countries in the world, people still suffer from, and protest against, oppressive and authoritarian regimes that threaten or violate their liberty. Often protest has lethal consequences: consider, for example, how political opposition is treated in contemporary Turkey, Belarus, Russia, China, and Myanmar. And while slavery is illegal, new forms of slavery continue today. The International Labour Organization estimates that globally there are more than 40 million people in some form of forced labor or forced sexual exploitation, for example in domestic work or in the sex industry (ILO 2017). It occurs within countries and via trafficking. Women and children are especially affected. It happens in North Korea, Eritrea, Burundi, the Central African Republic, Afghanistan, Pakistan, and Iran, but also persists in countries such as the US and the UK. According to the Global Slavery Index, in 2018 there were an estimated 403,000 people working under forced labor conditions in the US (Walk Free Foundation 2018, 180). Countries in the West also import goods and services that risk having involved modern slavery at the site of production.

But what does liberty mean, exactly, and what does political liberty mean in the light of developments in AI and robotics? To answer these questions, let us look at a number of threats to freedom, or, rather, threats to *different kinds of freedoms.* Let us examine some key conceptions of freedom developed by political philosophers: negative freedom, freedom as autonomy, freedom as self-

realization and emancipation, freedom as political participation, and freedom of speech.

AI, surveillance, and law enforcement: Taking away negative freedom

As we have seen in the introduction, AI can be used in law enforcement. It can also be used in border policing and airport security. Across the world, facial recognition technology and other biometric technologies such as fingerprints and iris scans are being employed in airports and other border crossing sites. As well as incurring the risk of bias and discrimination (see the next chapter) and threats to privacy (UNCRI and INTERPOL 2019), this can lead to all kinds of interventions that infringe on a person's *freedom*, including arrest and imprisonment. If an error is made by the AI technology (e.g., miscategorizing a person, not recognizing a face), individuals may be falsely arrested, denied asylum, publicly accused, and so on. A "small" margin of error may impact thousands of travelers (Israel 2020). Similarly, so-called *predictive policing*, which uses machine learning to "predict" crime, may lead to unjustified liberty-depriving judicial decisions, in addition to (again) discrimination. More generally, it may lead to "Kafkaesque" situations: opaque processes of decision-making and arbitrary, unjustified, and unexplained decisions, significantly affecting the lives of defendants and threatening the rule of law (Radavoi 2020, 111–13; see also Hildebrandt 2015).

The kind of freedom that is at risk here is what political philosophers call "negative liberty." Berlin famously defined negative liberty as freedom from interference. It concerns the question: "What is the area within which the subject – a

person or a group of persons – is or should be left to do or be what she is able to do or be, without interference from other persons?" (Berlin 1997, 194). Negative freedom is thus the absence of interference, coercion, or obstruction by others or the state. This is the kind of freedom that is at stake when AI is used to identify people who pose a security risk, who are said to have no right to migration or asylum, or who have committed a crime. The freedom that is threatened is a freedom of non-interference.

In the light of surveillance technologies, one could extend this conception of freedom to the freedom of not being at *risk* of interference. This negative freedom is at stake when AI technology is used for surveillance to keep people in a state of enslavement or exploitation. The technology creates invisible chains and ever-watching non-human eyes. The camera or the robot is always there. As has often been observed, this situation resembles what Bentham and, later, Foucault called the Panopticon: prisoners are watched, but they cannot see the watchers (see also chapter 5 on power). Physical restraint or direct supervision, as in earlier forms of imprisonment or slavery, is no longer necessary; it suffices that the technology is there to monitor people. It does not even have to function, technically speaking. Compare this with the speed camera: whether it actually functions or not, it already influences – in particular, *disciplines* – human behavior. And this is part of the very design of the camera. Knowing that you are being watched all the time, or could be watched all the time, is enough to discipline you. It is sufficient that there is a risk of interference; this creates the fear that one's negative freedom will be taken away. This can be used in prisons and camps, but also in work situations in order to monitor the performance of employees. Often surveillance is hidden. We do not see the algorithms, the data, and those who use these data. Bloom (2019) speaks, somewhat

misleadingly, of “virtual power” because of this hidden aspect. But the power is real.

AI surveillance is not only used in law enforcement and by governments, or in corporate environments and work contexts; it is also employed in the private sphere. For example, on social media there is not only “vertical” surveillance (by the state and by the social media company) but also peer surveillance or “horizontal” surveillance: social media users watch each other, mediated by algorithms. And there is *sousveillance* (Mann, Nolan, and Wellman 2002): people use portable devices to record what is happening. This is problematic for various reasons, but one reason is that it threatens freedom. Here this could mean the negative freedom to have privacy, understood as freedom from interference in the personal sphere. Privacy is usually seen as a basic right in a liberal, that is, free society. But this may be in danger in a society in which we are asked to embrace a culture of sharing. As Véliz (2020) puts it: “Liberalism asks that nothing more should be subjected to public scrutiny than what is necessary to protect individuals and cultivate a wholesome collective life. A culture of exposure requires that everything be shared and subjected to public inspection” (110). Full transparency thus threatens liberal societies, and big tech plays an important role in this. Using social media, we voluntarily create digital dossiers about ourselves, with all kinds of personal and detailed information that we willingly share, without any governmental Big Brother forcing us to give it or having to do the painstaking work to acquire it in covert ways. Instead, tech companies openly and shamelessly take the data. Platforms such as Facebook are an authoritarian regime’s but also a capitalist’s wet dream. People create dossiers and track *themselves*, for example for social purposes (meeting) but also health monitoring.

Moreover, such information can be and has been used against people for law enforcement. For example, based on analysis of data from her Fitbit device, an activity and health tracker, US police charged a woman with making a false report about rape (Kleeman 2015). Fitbit data were also used in a US murder case (BBC 2018). Data from social network sites and phones can be used for predictive policing, which may have consequences for personal liberty. Yet even if there are no threats to freedom from interference, the problem is also situated at the societal level and impacts different kinds of freedoms, such as freedom as autonomy (see the next section). As Solove (2004) puts it: "[I]t is a problem that implicates the type of society we are becoming, the way we think, our place in the larger social order, and our ability to exercise meaningful control over our lives" (35).

That being said, when it comes to threats to negative liberty by means of technology, the issue can get very physical. Robots can be used to physically restrain people, for example for security or law enforcement purposes, but also for "people's own good" and safety. Consider the situation when a young child or an elderly person with cognitive impairments risks crossing a dangerous road without watching or risks falling from a window: in such cases a machine could be used to restrain the person by, for example, preventing that person from leaving a room or leaving the house. This is a form of paternalism (more in the next section) that restricts negative liberty by means of surveillance followed by a physical form of interference. Sharkey and Sharkey (2012) even see in the use of robots to restrict the activities of the elderly "a slippery slope towards authoritarian robotics." Such a scenario concerning monitoring and restraining humans through AI and robotics technology seems more realistic than the distant, science-fiction scenario of superintelligent AI

taking over power – which may also lead to taking away liberty.

Anyone using AI or robotics to restrict the negative liberty of people has to justify why it is necessary at all to violate such a basic kind of freedom. As Mill (1963) argued in the mid-19th century, when it comes to coercion, the burden of proof should be on those who contend for a restriction or prohibition, not on the people defending their negative liberty. In the case of privacy violations, law enforcement, or paternalistic restriction of movement, the onus is on the one who restricts to show that there is a considerable risk of harm (Mill) or that there is another principle (e.g., justice) that is more important than liberty – in general or in the particular case. And justifying such uses and interventions becomes even harder when the technology makes mistakes (the false match case in the introduction) or when the technology itself causes harm. For example, facial recognition may lead to unjustified arrest and imprisonment, or a robot may cause injury when and while restraining someone. Furthermore, beyond utilitarian and, more generally, consequentialist frameworks, one could emphasize rights to liberty from a deontological point of view, for example the rights to liberty enshrined in national and international declarations.

Yet considering these cases when the technology has (unintended) harmful effects, it becomes clear that there is more at stake than liberty alone. There are tensions and trade-offs between liberty and other political principles and values. Negative freedom is very important, but there may also be other political and ethical principles that are very important and that (should) play a role in a particular case. It is not always clear which principle should prevail. For example, whereas it may be crystal clear that it is justified to restrain the negative freedom of a small child in order to prevent a particular harm (e.g., falling out of a window), it