1

WHAT IS TECHNOLOGY (FROM AN ETHICAL POINT OF VIEW)?

1.1 A Hut in the Black Forest

Around 100 years ago, in the Black Forest of Southern Germany, there stood 1.1
a small and simple three-room cabin, to which an eccentric philosopher would retire in order to escape from the modern world. From 1922 onward, he went there to work on philosophical texts about the nature of "being," and he felt deeply inspired by these surroundings. The philosopher in question was Martin Heidegger. He composed some of his most well-known writings in this cabin, which he called "die Hütte" (the hut). While in his Hütte, Heidegger would sometimes dress up in traditional farmers' clothing. And he would wander about in the Black Forest and ponder abstract philosophical questions, far away from bothersome modern technologies and other distractions.

This might seem like a strange place to start this book, and a strange 1.2
choice of philosopher. However, it makes sense to start a book about technology ethics with Heidegger and his hut for a number of reasons. One reason is that one of the most-cited writings about what technology is was written by Heidegger (most likely in the cabin!). Another reason is that it is good to have a reminder early on about the following. Even though we usually think about the latest and most advanced forms of cutting-edge technology when we hear the word "technology," even the simplest and oldest forms of technology are also technologies.

For example, Heidegger would go and get water from a well when he 1.3
visited his hut. (There was no electricity or running water when he first

This is Technology Ethics: An Introduction, First Edition. Sven Nyholm.

started frequenting his cabin.) The well where Heidegger got his water is a technology, which can perform a very important purpose, just as something like a self-driving car or the latest smartphone is a technology. Heidegger himself tended to romanticize the past. He made a sharp distinction between modern technologies (of which he was skeptical) and more traditional technologies (which he happily embraced). But when we think about the general question of what technology is, we should keep in mind that there are many different kinds of technology, both old and new. Old technologies are technologies, just like new ones are.

1.4 If a traditional water well is a technology just as much as a self-driving car or a smartphone is a technology, then what exactly is a technology? This question—"what is a technology?"—is what we will focus on in this first chapter. Since this is a book about technology ethics, it is important to have an idea of what we should understand by "technology" in the first place. The ancient philosopher Aristotle recommended that one should always start by defining one's key terms. That certainly seems like good advice.

1.5 It will also be important to try to define or explain what we mean by "ethics," so that we can put "technology" and "ethics" together to form the idea of "technology ethics." But we will save the question of what *ethics* is until the next chapter and here focus on what *technology* is. When we do so, however, that will already start raising ethical issues even before we get around to trying to say what ethics is, as we will do in the next chapter. One reason for this is that when one tries to explain important concepts, the choices we make about how to define these concepts might prove themselves to be controversial, and raise questions about what is important or valuable.

1.6 For example, consider the questions "what is a human being?" and "what is a person?" On the face of it, these might not seem like inherently ethical questions. But depending on who or what you think qualifies as a human being or person and for what reasons, you might find that before you know it, there will be others who find your views and their implications to be highly controversial. Ethical controversies about abortion, for example, tend to turn partly on who or what counts as a human being or as a person.

1.7 Choices about whose views about something to listen to—e.g., about what "technology" should be taken to be—can also be controversial and raise ethical questions. Heidegger, for example, has in recent times become a highly controversial figure. While he was rector of the University of Heidelberg during the German Nazi era, Heidegger was a member of the Nazi party. That by itself does not necessarily mean that Heidegger was a

Nazi. Anybody in a leading position may have needed to be a member of that party during that era. But the recently discovered "Black Notebooks" (some notebooks of Heidegger's) have revealed that for a while, Heidegger was not only a member of the Nazi party but also convinced of some of the ideological ideas associated with Nazism. Some even think that Heidegger's romanticism and his musings about the simple life close to nature and his hut in the forest are somehow connected with Nazi ideals.

Heidegger scholars claim that Heidegger fairly quickly came to reject the 1.8
ideas of the Nazis. Furthermore, the essay "The Question Concerning Technology" that is of relevance here was written after Heidegger abandoned any sympathies he had with Nazi ideas for a while. Yet it might seem controversial to start a discussion of what technology is with the ideas about this topic from a one-time Nazi. Nevertheless, as noted above, Heidegger's just-mentioned essay is one of the most-often discussed contributions to the main topic of this chapter. So, it is good to be aware of some of the things he says about what technology is, for the sake of context. Moreover, as we will see below, there are also philosophical reasons to start with Heidegger's discussion, despite whatever flaws Heidegger himself might have had as a person, at least during certain stages of his life.[1]

1.2 The Question Concerning Technology: The Instrumental Theory of Technology from Martin Heidegger to Joanna Bryson

The further into Heidegger's "The Question Concerning Technology" one 1.9
gets while reading it, the harder the text becomes to understand. But there are some bits at the beginning of that essay that are fairly straightforward. For example, let us consider this passage:

> Everyone knows the two statements that answer our question. One says: Technology is a means to an end. The other says: Technology is a human activity. The two definitions of technology belong together. For to posit ends and procure and utilize the means to them is a human activity. The manufacture and utilization of equipment, tools, and machines, the manufactured

[1] The Wikipedia entry on "Martin Heidegger" (https://en.wikipedia.org/wiki/Martin_Heidegger) features a picture of Heidegger's hut: https://en.wikipedia.org/wiki/Martin_Heidegger#/media/File:Heideggerrundweg0009.JPG

> and used things themselves, and the needs and ends they serve, all belong to what technology is. The whole complex of these contrivances is technology.
>
> (Heidegger 1977: pp. 4–5)

A couple of lines down, Heidegger adds:

> The current conception of technology, according to which it is a means and a human activity, can therefore be called the instrumental and anthropological definition of technology.
>
> (Heidegger 1977: p. 5)

Notice that what Heidegger does here is in effect to first formulate two different theories of technology, and to then combine them into a hybrid theory. The *instrumental theory of technology* defines technologies as tools or instruments that human beings use as means to their ends. The *anthropological theory of technology* defines technologies as parts of distinctively human activities. The combined theory ("the instrumental and anthropological definition of technology") represents technologies as means to ends within human activities.

1.10 Another thing that Heidegger by implication also does is to highlight that we can think about technology from different points of view. The instrumental theory of technology, for example, fits with what one might call an engineering mindset. Engineers love to identify concrete problems, and to then seek out or design the best tools for solving these problems. From such an engineering point of view, it is very natural to think about technologies as tools used as means to the end of solving concrete problems. The anthropological theory, in contrast, adopts the mindset of an anthropologist. Anthropologists and other social scientists study human practices and activities. So, it is natural for them to portray technologies as parts of human practices or activities.

1.11 Another noteworthy thing about the instrumental theory of technology, as it is usually interpreted, is that it represents technologies as being inherently value-neutral. Think about the saying "guns don't kill people, people do." That slogan has been used by the American National Rifle Association as a response to the charge that having lots of guns in society heightens the risk of violence. The idea behind "guns don't kill people, people do" seems to be that guns are in themselves value-neutral tools. It is people with bad intentions who might use these tools in bad ways. It is also possible to use these value-neutral tools in good ways, according to this part of the instrumental theory of technology.

In general, any tool can be put to good or bad uses, according to this way of seeing things. Accordingly, if bad things happen and technologies are involved, we should not blame the technologies, but the people who use them for their own ends, which might be morally problematic. On the flipside, when technologies are used for good ends, the instrumental theory implies that we should not thank or praise the technologies but the people who create or use them. Technologies can deserve neither praise nor blame, according to this way of thinking. That might seem obviously true. But as we will see below, there are those who take a different view. 1.12

Speaking of how technology can be interpreted from different points of view, below we will consider how we should define or understand technology when we take up a distinctively ethical or perhaps more broadly philosophical point of view. But let us first linger just a little bit longer with the instrumental theory of technology, setting aside the anthropological perspective on technology for a moment. In particular, it is worth noting that even though a purely instrumental theory is often criticized in contemporary technology ethics, there are also prominent defenders of the instrumental theory of technology in current discussions. By considering one example, we can illustrate how the instrumental theory of technology can (1) be used in ethical arguments and (2) be further extended. 1.13

Joanna Bryson is a computer scientist and roboticist who has recently transitioned into the role of a professor of ethics. Among other things, she is the author of a striking essay entitled "Robots Should be Slaves" and a number of follow-up articles that further refine the argument in her original piece.[2] In her characteristically forceful language, Bryson applies an updated version of the instrumental theory to the question of whether we should ever regard robots with artificial intelligence as worthy of rights or moral consideration. 1.14

Bryson often tells an anecdote about how she came to be interested in this topic that helps to illustrate what she is concerned about. When Bryson worked in a robotics lab at MIT early in her career, there was a robot called "Cog" that people in the lab used to talk about in ways that suggested that 1.15

[2] Bryson has expressed some regret about the title of her paper. She now prefers talking about robots as servants we own rather than as slaves we own. But the argument remains the same. You can hear Bryson discuss her ideas in this episode of the "Philosophical Disquisitions" podcast: "Episode #24—Bryson on Why Robots Should Be Slaves": https://philosophicaldisquisitions.blogspot.com/2017/06/episode-24-bryson-on-why-robots-should.html

they owed moral consideration to it. They would say things such as "don't pull the plug or turn Cog off—that would be to kill Cog!" Occasionally, they would say such things without realizing that Cog the robot was actually already unplugged. These people were projecting humanlike qualities onto a robot.

1.16 As Bryson sees things, this way of behaving around robots is a big mistake. Robots—like any other technologies—are tools that we create for our human purposes. Moreover, and here comes the key dimension that Bryson adds to the instrumental theory of technology: robots and other technologies are the property of people. They can be bought and sold. Since robots are property that people can buy and sell and tools we use for human purposes, technologies are like "slaves" or "servants" we own, just like some human beings were once regarded as tools, who could be bought and sold, in some societies of the past.

1.17 In the case of human beings, Bryson thinks that it is wrong and horrible that there were ever slaves anybody could own: tools that others could buy and sell. But that, Bryson thinks, is what all human-created technologies are and should be. One must add "should be" here, because Bryson has the interesting view that it is possible to create robots or other artificially intelligent technologies toward which we could have duties, but that we should avoid doing so.

1.18 If we could create machines that can experience suffering or that are intelligent and sensitive in the ways that human beings and many animals are, then Bryson thinks we would have obligations towards these creations. But we should avoid creating such robots. We should only create technologies that it is okay to treat as tools, which we can buy and sell. Since Bryson adds these ideas about what should and should not be done to the instrumental theory, we can say that she presents a *normative version of the instrumental theory of technology*.

1.19 One of the real-world robots and related events Bryson has reacted strongly to is the robot Sophia and the 2017 event where this robot was given honorary citizenship of the Kingdom of Saudi Arabia. Sophia is a robot that looks like a human being (a "humanoid robot"). The robot has a humanlike face and is able to imitate human speech. The back of the robot's head is transparent, so that one can see that it is a machine. But the robot is put in many distinctively human situations: for example the robot has appeared on talk shows such as *The Tonight Show with Jimmy Fallon*. And many human beings—including highly influential people—have treated the robot as if it were a human person. For instance, the former Chancellor

of Germany, Angela Merkel, once took a "selfie" photograph together with Sophia. Moreover, the robot has been invited to speak in front of high-profile international political bodies, like the United Nations and the Munich Security Council.

For somebody like Bryson, who adopts a normative version of the instru- 1.20
mental theory of technology, the just-described events are highly problematic from an ethical point of view. In an interview, Bryson commented on this in the following way:

> What is this about? It's about having a supposed equal you can turn on and off. How does it affect people if they think you can have a citizen that you can buy?[3]

Again, on the instrumental theory of technology, a robot and any other technology is a tool, which is value-neutral in itself, and that you can buy or sell. Since Bryson subscribes to that view, she thinks that the way people behave around Sophia the robot is highly problematic. However, not everyone agrees with the instrumental theory of technology. Let us now explore some other ideas about what technology is or can be.

1.3 "Post-Phenomenology" and the Mediation Theory of Technology

As was mentioned above, we can adopt different mindsets when we reflect 1.21
on what technology is or should be taken to be. One thing we can do is to ask what we should think of technology as being from an *ethical* point of view. For most of the rest of this chapter, that is the point of view from which we will consider what technology is. In other words, our question will be: "when we think about technology from the point of view of ethical theorizing, what should we then understand technology as being?" Taking up this ethical point of view on technology, many authors have criticized the purely instrumental theory of technology.

---

[3] James Vincent (2017) "Pretending to Give a Robot Citizenship Helps No One," *The Verve*, https://www.theverge.com/2017/10/30/16552006/robot-rights-citizenship-saudi-arabia-sophia

1.22 Some sociologists and philosophers who consider what technology is, from an ethical point of view—and who criticize the instrumental theory—do so in a way that fits with what Heidegger calls the anthropological way of understanding technology. That is, they are looking at and analyzing human practices involving technologies. They are interested in how the technologies we have around us affect us. And they argue that within many human practices, technologies are not always merely tools that are completely value-neutral in nature. Technologies, they argue, can also play other roles in our lives. One such group of thinkers are the members of the so-called post-phenomenological school of thought.

1.23 The term "post-phenomenology" is not only hard to pronounce. It may also be hard to understand for those who have not heard it before. So, let's try to break this idea down into parts. The first thing to know is that "phenomenology" was a philosophical movement that started around the time that Heidegger was writing his philosophy in his hut. This movement was premised on the idea that in reflecting on human life, we should not start with abstract philosophical theories, but with the phenomena that we encounter in the world. In our philosophizing, we should be investigating how reality and how we ourselves appear to us from the point of view of how we experience things as human beings. Post-phenomenology takes this idea a step further and asks how technologies shape how we experience reality, ourselves, and what we are able to do or aim for.

1.24 Members of this school of thought—such as Don Ihde, Bruno Latour, and, more recently, Peter-Paul Verbeek—are interested in how technologies "mediate" the way we experience things and what we are able to do.[4] Technologies, they say, are a medium in between us and what we perceive and what we can do. There are two different ideas here. One has to do with inputs, the other with outputs. Let us consider them one after the other.

1.25 One key idea is that technologies shape how we perceive or experience the world and even ourselves. To use a simple example, if we are wearing glasses, our visual experiences and perceptions are going to be different than if we remove our glasses. Or, to use another example, if you visit the Louvre in Paris and you view the Mona Lisa through the camera on your phone, as many tourists who visit the Louvre do these days, then

[4] This short video, by Verbeek, introduces the mediation theory of technology: "Animation: Explaining Technological Mediation": https://www.youtube.com/watch?v=FVhrLwBNbvU

your perception of the Mona Lisa will be mediated or shaped by your phone. It will be different from what it would have been if you simply look straight at the painting without your phone in between it and your eyes. More generally, what we pay attention to as we go about our lives is shaped by the technologies we use. For example, these days, what news (or fake news!) people hear about and pay attention to tends to be strongly influenced by algorithms that are part of the social media that they use.

The second key idea is that technologies shape what we are able to do and 1.26
how we are able to do it. Relatedly, technologies shape what we see ourselves as being able to do. They also sometimes "suggest" certain actions to us. For example, these days, we can take weekend trips to faraway places because we have airplanes that can quickly take us to a distant country. So, the idea of taking a weekend trip to some place far away might occur to a person, because the technologies available to us make this possible. This would not have been possible, nor anything that would have occurred to anybody to do, before these technologies were available. Speaking of social media—to return to the example at the end of the previous paragraph—that is also an example of a technology that might shape how we act, for example how we communicate with other people.

A lot of people would never think of, or even dare to say certain things to 1.27
other people face-to-face that they think of saying and dare to say to them on social media. This can be both good and bad: good because people might dare to speak out against something, bad because people might hurl abuse at each other. This is another example of how technology shapes what we do and what we think of doing.

Actually, there is an idea in Heidegger's "The Question Concerning 1.28
Technology" essay that is very much in line with this post-phenomenological idea that technologies shape what we can do and how we perceive the world. Heidegger talks about how new technologies—e.g., a water powerplant—might change how we view something like a river. Suddenly, the river is presented by the modern water powerplant as a source of electricity, or as a means to an end to which it could not have been viewed as a means before.

Heidegger takes this idea to the extreme. He argues that the more mod- 1.29
ern technologies we surround ourselves with, the more nature will appear to us to be a large set of means to human ends as opposed to an end in itself. While a river might have appeared to us as a beautiful piece of nature before, a new technology might suddenly present it as a means to the end of producing electricity. As a result, the way we perceive the river might then have been changed by the technology.

1.30 This general idea is picked up by post-phenomenologists when they criticize the part of the instrumental theory of technology that portrays technologies as being in themselves value-neutral. Members of this school of thought argue that technologies are not purely value-neutral for the reason that technologies influence how we value things around us. Peter-Paul Verbeek, for example, is fond of using the example of ultrasound images of fetuses. Verbeek claims that such images present fetuses in a new light (different from the mere knowledge that there is a fetus in the womb). In particular, those images present the fetus as a patient, whose health status the parents need to worry about and make decisions about.

1.31 Another type of example of how technologies are not value-neutral tools according to post-phenomenologists is often illustrated with the help of an idea that Don Ihde discusses. The idea is that technologies sometimes contain intended or unintended "scripts." That is to say, they may tell us what to do and perhaps even force us to act in certain ways. For example, the French sociologist and philosopher Bruno Latour talks about speed bumps as telling drivers to slow down, paper cups as communicating to users that they should be thrown away after use, and the beeping noises cars make when we do not put on our seatbelts as forcing us to wear our seatbelts.

1.32 Members of this post-phenomenological movement also criticize the instrumental theory of technology by arguing that the instrumental theory tends to falsely represent humans and technologies as being wholly separate from each other when it portrays humans as acting on the basis of ends and technologies as being means used to these ends. This is not correct, many post-phenomenologists argue. It is not correct because human beings and technologies form "assemblages" or units that can do things that humans alone, without the technologies, neither could nor would do on their own.

1.33 Verbeek, for example, takes issue with the above-mentioned idea that "guns don't kill people, people do." According to Verbeek, a man with a gun easily becomes what Americans tend to call a "gunman," i.e., somebody with a gun and a disposition to shoot. The man and his gun form a unit, as Verbeek sees things. And this unit can do things, and may become inclined to do things, that the man cannot or would not do without his gun.[5]

---

[5] You can hear Verbeek discuss these ideas in more detail in this episode of the "New Books in Philosophy" podcast: "Peter-Paul Verbeek: Moralizing Technology: Understanding and Designing the Morality of Things": https://newbooksnetwork.com/peter-paul-verbeek-moralizing-technology-understanding-and-designing-the-morality-of-things-university-of-chicago-press-2011

Notably, the science and technology researcher Donna Haraway suggests 1.34
that is helpful to think of us human beings using the metaphor of cyborgs. There is, as Haraway sees things, no sharp distinction between humans and technologies. Many of us have heard the expression "you are what you eat." Haraway's view seems to be something along the lines of "you are, in part, the technologies you use." The idea that human beings might merge with technologies, or that we have already done so, is something we will have occasion to return to later in this book. Post-phenomenologists are not the only ones fascinated by that idea. As we will see in Chapter 10, many others are fascinated by it as well.

To summarize the key points in this section: according to post- 1.35
phenomenologists, technologies are not value-neutral tools that are wholly separate from human beings, who set ends and use technologies as means to their ends. Technologies are more than that. Technologies shape how we perceive the world, what we are able to do, and what we value and how we value it. Humans and the technologies we use are not wholly separate from each other. They form units that are able to do things, and that are also disposed to do things, that humans without these technologies would neither be able nor inclined to do. This can be seen as an updated and refined version of the anthropological theory of technology since this is a theory about what roles technologies play in human practices. These interesting ideas from the post-phenomenologists will be popping up here and there throughout this book. But let us set them aside for now, and return to the instrumental theory of technology once more.

1.4 Technologies Conceived of as Being More Than Mere Means or Instruments

Recall how Joanna Bryson talks about how robots and other technologies 1.36
are, and should be conceived of as, tools that we own, which we can buy or sell, and which should be treated as means to human ends. This brings to mind some ideas that the illustrious Enlightenment philosopher Immanuel Kant discusses in his classic, very influential ethical treatise *Groundwork for the Metaphysics of Morals*. What Bryson says about how we should treat technologies can be seen as an inversion, or the opposite, of what Kant says about how we should treat human persons. Since Bryson talks about technologies and Kant talks about persons, their views can be seen as complementing each other.

1.37 In that just-mentioned book, Kant divides up the world into two separate categories: what he calls "persons" and "things." *Persons* can think and act, be rational, and make moral decisions. They have an absolute value and dignity, and should be treated with care and respect. Persons, in Kant's phrase, should be regarded as "ends-in-themselves." They can be responsible for their actions, and are members of the moral community. Everything else in the world, Kant argues, belongs to the category of *things*. Anything that is not a person only ever has a relative value—namely, a value relative to the desires or wishes of persons.

1.38 As Kant sees things, all things can be treated as mere means to the ends of persons. Persons, in contrast, should always be treated as ends and never as mere means. Kant also relates this to our distinctive humanity. When he does so, Kant formulates the following well-known and very popular moral principle: "so act that you always treat the humanity in each person, never merely as a means, but always at the same time as an end-in-itself." This is sometimes called the "formula of humanity" or, alternatively, Kant's principle of "persons as ends-in-themselves."[6]

1.39 We can restate Bryson's normative version of the instrumental theory of technology in these Kantian terms in the following way: unlike human beings, all technologies are and should be regarded as things, and not persons. They should always be treated as mere means and never as ends-in-themselves. Moreover, we should avoid creating any technologies that would ever be anything other than mere things that we can treat as mere means. We should never create technologies that are or appear to be persons, which are ends-in-themselves. Bryson's normative instrumental theory of technology, in other words, can be seen as an interesting inversion of Kant's principle of humanity.

1.40 It makes sense to bring up this Kantian terminology and these Kantian ideas here, not only because they are interesting in themselves but also because they give us a philosophical vocabulary with which we can reinterpret the instrumental theory of technology. It also makes sense to bring up

[6] Here is a mini lecture on Kant's humanity formula from the "Cogito" channel: "Kant's Second Categorical Imperative": https://www.youtube.com/watch?v=0Cl0FG-xUKA.You can listen to a longer discussion about Kant's ethics on this episode of BBC's radio program "In Our Time" here: "Kant's Categorical Imperative": https://www.bbc.co.uk/sounds/play/b0952zl3

these things because Kant's philosophy, just like Heidegger's, has had a lot of influence within technology ethics, as we will see later on.

Moreover, it is also worth bringing up these ideas from Kant and their relation to Bryson's updated version of the instrumental theory of technology because of certain striking recent developments within the ethics of technology. Recently, several technology ethicists have defended ideas that clash starkly with pretty much all of the ideas just mentioned in the paragraphs above. In particular, some technology ethics researchers have recently defended one or more of the following ideas: 1.41

- Some technologies can think and act, and be rational and make moral decisions.
- Some technologies can be persons, as opposed to mere things.
- Some technologies should be treated as ends (i.e., be shown moral consideration or given rights), and not as mere means.
- Some technologies can be responsible for what they do, just like human persons can be responsible for what they do.

Such ideas can seem counterintuitive. They are perhaps shocking and even disturbing to those who accept any version of the instrumental theory of technology. But as noted, these ideas are gaining a foothold within contemporary technology ethics. To see why some philosophers of technology are beginning to take them seriously, we can now consider some examples of how some people interact with new forms of technologies like robots and other technologies equipped with artificial intelligence. We will save a full discussion on these controversial ideas until later chapters. But it makes sense to briefly also look at them in this first chapter, since we are here considering the question of what we should understand by the idea of technology. 1.42

1.5 Technologies Regarded as Moral Agents

We have already briefly considered the example of how some of Bryson's colleagues talked about the robot "Cog" (e.g., they were hesitant about pulling the plug or turning off the robot). And we have also considered how some people have been treating Sophia the robot (e.g., giving honorary citizenship to the robot, inviting it to appear on talk shows, taking selfie photos with it, or allowing it to speak in front of important political bodies). These 1.43

are examples of people treating robots in ways that suggest that they regard them, or that they are willing to treat them, as being more than mere means or things that are categorically different from persons. Many more examples can be given. And more examples will be given throughout the book. But here are just a few more examples of how people think about technologies in ways that clash with the purely instrumental theory.

1.44 Many people working on questions in technology ethics are interested in what are sometimes called functionally "autonomous" machines. That is, machines that for some period of time are able to operate on their own, performing specific tasks, without direct human steering. Such technologies are occasionally put into situations where they seem to need to be able to make moral decisions. Notably, many philosophers and other researchers think that it is possible to create machines that can make moral decisions. A widely discussed example is that of the self-driving car. Another is the example of military robots or what are sometimes called autonomous weapons systems. These technologies operate in contexts that are dangerous and where human lives are at stake.

1.45 We can imagine that a self-driving car might need to "decide" whether to go left or right when a fork in the road occurs. The breaks of the car may have stopped working. The problem might be that each option involves a threat to human beings. On the road on the left, there might be five people, who would not be able to get away from the road in time and thereby avoid being hit by the car. On the road on the right, there might be one person, who would also not be able to get away from the road in time to avoid being hit by the car. The person riding in the car may have fainted, and therefore not be able to offer any input regarding what to do. So, it might seem that this self-driving car, whose breaks are currently not working, needs to make a "moral decision." Should it go left, and potentially injure or even kill five people? Or should it go right and thereby save the five, but potentially injure or kill one person? This is widely regarded as a type of situation in which a machine—in this case, a self-driving car—needs to be able to make a moral decision and then act on it.

1.46 An interdisciplinary research field called "machine ethics" is devoted to investigating whether it is possible to create moral machines and if so, how this can and should be done. According to this way of thinking, the technologies in question are not simply regarded as value-neutral tools that are used by human beings to achieve their ends. Rather, the technologies are regarded as what philosophers call "moral agents." That is, they are regarded as entities or beings capable of making moral decisions and acting on those decisions.

Most ethicists of technology who take this idea seriously do not think 1.47
that the technologies in question would be moral agents of the sorts that we human beings are. Rather, the technologies—the self-driving cars, military robots, or whatever it might be—would be some other kind of moral agent. They might be able to make moral decisions and act on them, but perhaps they would not be morally responsible for their decisions in the ways that human beings can be. This is a common position. It is defended, for example, by the well-known Italian technology ethicist and information philosopher Luciano Floridi.[7]

As Floridi sees things, any entity that can act in some sense and that can 1.48
act in situations that are morally significant is a moral agent. But this is not enough, Floridi thinks, for regarding the entity as a morally responsible agent who can be blamed or praised for its actions or decisions. This idea of artificial agents that are not able to be responsible but that are nevertheless making morally sensitive decisions is sometimes thought to give rise to worries about so-called responsibility gaps. This expression refers to the idea that a machine, e.g. a military robot, might autonomously make a morally important decision and do something (e.g., kill a human being), but then not be able to be held responsible for the outcome. There may also not be any human beings who can be held morally responsible for what the robot did. At the same time, it might seem as if somebody should be held responsible for what happened. Hence a gap in responsibility apparently arises. Worries about such responsibility gaps are widely discussed within technology ethics. And most ethicists of technology do not think that these problems can be solved by holding the machines themselves responsible.

There are others, however, who take the more extreme view that some 1.49
technologies can not only make moral decisions, but sometimes also be responsible for those decisions in some sense. The philosopher Daniel Tigard, for example, has argued in a series of academic articles that there are many different "faces of moral responsibility" (that is, many different aspects of what is involved in being morally responsible). According to Tigard, some advanced technologies, such as some artificially intelligent robots or other autonomous systems, can be responsible in certain, at least

[7] Here is an interview with Floridi, from "The Dissenter": "#329: Luciano Floridi: Information, Knowledge, Science, and AI": https://www.youtube.com/watch?v=jgmdEk7rTxU. Floridi discusses his views about artificial moral agents and responsibility just after the 38:03 mark.

limited, ways.[8] A view such as Tigard's is very far removed from a purely instrumental view of technology, on which all technologies are value-neutral tools or means to human ends. If we think that some technologies can be responsible for what they do or decisions they make, our understanding of technology is very different from the sort of view Bryson advocates.

1.50 We have just considered technologies in the role of what philosophers like to call "moral agents." There are also examples where technologies seem to occupy a role that philosophers call that of "moral patients." A moral patient is an entity or being towards which it is possible to have obligations or duties, or that it is possible to treat more or less well. In short, a moral patient is a being or entity with moral status. It is an entity or being that might have rights of some important sort. When Sophia was made into an honorary citizen of Saudi Arabia, for example, this might have seemed to bestow certain rights and some important form of social status onto Sophia, since citizens are typically regarded as having certain rights and an important status. But let us now also consider another example to illustrate the idea of robots as potential moral patients.

1.6 Technologies Regarded as Moral Patients

1.51 In February of 2015, the news network CNN published an article with the headline "Is it cruel to kick a robot dog?"[9] This article was about some interesting reactions to a video that had been posted by a tech company called Boston Dynamics. The video featured a robot that looks a little bit like a dog, and which is very good at keeping its balance. The robot, called "Spot," is shown running on a treadmill and walking up some stairs, doing all this without falling over. Later in the same video, and again in order to illustrate how good the robot is at keeping its balance, one of the Boston Dynamics engineers is shown kicking Spot. Sure enough, Spot is able to keep its

[8] You can hear Tigard discuss this topic on this episode of the "Philosophical Disquisitions" podcast: "79: Is There a Techno-Responsibility Gap?" https://philosophicaldisquisitions.blogspot.com/2020/08/79-is-there-techno-responsibility-gap.html

[9] Phoebe Park (2015) "Is It Cruel to Kick a Robot Dog?" *CNN Edition:* https://edition.cnn.com/2015/02/13/tech/spot-robot-dog-google/index.html

balance and does not fall over, even after a pretty solid kick from the engineer. While Spot was able to stabilize itself after being kicked, many viewers of this video lost their cool when watching Spot being kicked. CNN reported that viewers reacted to the video with comments such as: "Kicking a dog, even a robot dog, just seems so wrong," "Poor Spot!," and "Spot getting kicked is creepy."

1.52 One thing that is fascinating about these reactions, from a philosophical point of view, is that the people who reacted in this way to this video seemed to regard Spot the robot dog as more than a mere tool, which we can treat in any way we might want for the sake of our ends. The people who reacted in this way seemed to be responding in a way that suggests that they are willing to regard this robot as having some form of moral standing. Their reactions portrayed Spot as a moral patient, i.e., as a being or entity toward which one can act wrongly and that should be treated with some moral consideration.

1.53 But not only people who post comments to online videos adopt that stance towards technologies. Some ethicists of technology have started to take seriously the idea that some technologies—such as some robots—should be regarded as moral patients, to which we should give moral consideration. There are even a few scholars who are taking seriously the idea of what they call "robot rights." Some of them even go as far as to discuss the idea of "robotic persons."

1.54 For example, Mark Coeckelbergh and David Gunkel have developed what they call a "relational" theory of moral status ascriptions. What they mean by this is that we should consider the question of who or what deserves to be treated with moral consideration, or be given rights, as being a question of what sorts of relationships human beings can have with those entities. Just like non-human animals can become part of human communities and thereby acquire special moral status and even rights (e.g., as pets or farm animals do), so it could also be that some technologies (e.g., some robots) might become part of human communities. It might then make sense to extend moral consideration, and give rights, to them. Or so Coeckelbergh and Gunkel argue.

1.55 If we move away from the instrumental theory of technology and we adopt this relational way of thinking about human-technology interaction, we would take what Coeckelbergh and Gunkel call a "relational turn" in our thinking about technology. On such a view, what technologies themselves are able to do is seen as less important than what relations there might be between these technologies and the humans interacting with them. Sophia

the robot, for example, might not have human emotions or be a sensitive or intelligent being. But a robot like Sophia might nevertheless become part of human society and thereby become something akin to a person with moral standing and even certain rights.[10]

1.56 Another type of theory that has been suggested to defend the idea of extending moral status to robots is what the Irish legal academic and moral philosopher John Danaher calls "ethical behaviorism." According to this idea, if a technology consistently behaves like a human being or an animal that has moral standing behaves, then we should treat that technology with the same degree of moral consideration that we show for a human being or an animal. For example, if a piece of technology behaves in a way that indicates that it is suffering (for instance, screaming in apparent pain), then since we should avoid causing unnecessary suffering to human beings and animals, we should also refrain from whatever actions are causing the technology to behave as if it is suffering.[11]

1.57 Danaher has even gone so far as to extend this idea of ethical behaviorism to the topic of friendship. He argues that if a technology—e.g., a robot, an avatar, or a chatbot—can consistently behave in the way that a friend behaves, then we can be friends with these technologies, just like we can have human beings as our friends. More on that in Chapter 9.

1.58 These are all examples of ordinary people and researchers who think that some technologies are not to be regarded as only being tools or instruments that are value-neutral in themselves and that we can always use as mere means for whatever ends we might have. These are examples of non-instrumental ways of regarding some technologies, and of ways of attributing a humanlike status to some types of technology. In other words, they are examples of how, from an ethical point of view, it is possible to understand what technologies can be in a wide and perhaps surprising range of different ways, some of which are very far removed from the purely instrumental theory of technology.

---

[10] You can listen to Gunkel talk about his ideas on the "Dolores Project" podcast here: "Robot Rights w/David Gunkel": https://anchor.fm/joshua-k-smith6/episodes/Robot-Rights-w-David-Gunkel-e12viq7

[11] You can listen to an audio essay by Danaher on "ethical behaviorism" here: "Assessing the Moral Status of Robots: A Shorter Defence of Ethical Behaviourism": https://philosophicaldisquisitions.blogspot.com/2019/10/assessing-moral-status-of-robots.html

The ideas and examples surveyed in this section are so controversial— 1.59
and so interesting!—that we cannot just leave them hanging like this. We will discuss all of them in much more detail later on in the book. One whole chapter will be devoted to whether technologies can be moral agents (Chapter 7). Another chapter will be devoted to whether technologies can be moral patients (Chapter 8), and yet another to whether technologies can be our friends or have other significant relationships with human beings (Chapter 9). For now, the main point of briefly introducing these examples has simply been to illustrate that there is a wide range of ways of regarding what technologies can be and what roles they can have within our lives. The instrumental theory of technology provides one, but not the only possible perspective on what roles technologies can have.[12]

1.7 Some of the Key Types of Technologies That Will Be Discussed at Greater Length in Later Chapters of the Book

Up until this point, we have focused on the very general question of what 1.60
technology is or what it could be. We have seen that depending on what mindset one adopts when one thinks about technology, one can take radically different views about what technologies are or can be. For example, if one puts on one's engineering hat, one is likely to view technologies as tools that we design and use to solve specific problems, whereas if one puts on one's anthropologist hat, one might think of technologies as parts of human practices or activities. And if we think about what technologies are from an ethical point of view, we have seen that there are even those who think that some technologies can be moral agents or moral patients, or have the status as some form of persons, or perhaps even become our friends.

So far, the discussion has been on a pretty general level. We have asked 1.61
what "technology" refers to in general terms. At this point, some might respond—not without some justification!—that it might be better to try to explain or define certain specific classes of technology that are particularly interesting from an ethical point of view. So that is the last thing this first chapter will do.

[12] For references to the work mentioned in this and previous section, see the annotated bibliographies of Chapters 7–9.

1.62 We can start with the concept of a *robot*, a concept that has already been used a number of times above. What is a robot? Some—such as the robot ethicist David Gunkel—are unwilling to define what a robot is. There are so many different kinds of machines that are classified as "robots," and some of them may seem to have little to do with each other. But if pressed, researchers like Gunkel are likely to define what a robot is with reference to what is sometimes called the "sense, think/plan, act" paradigm. According to this way of seeing things, a robot is a machine with sensors and actuators, which is able to take in information about its environment, process that information, and take some form of action in response to this information about its environment—and which does so in the service of specific tasks that the robot is supposed to perform.

1.63 For example, a vacuum cleaning robot like the "Roomba" robot is able to detect whether there are big pieces of furniture in its path. And it is able to use this information from its sensors to steer itself in a different direction, to thereby be able to carry out its task of vacuuming the room it is operating in. When we think of robots, we do not only think of robots like the Roomba robot, however. We might also think of paradigmatic robots out of science fiction, that is, robots that look a little bit like metallic or silvery beings with a roughly humanlike form, which may move and talk in a "robotlike" way.

1.64 If functional robots like Roomba or lawn-moving robots, or logistics robots or military robots, etc., are on one end of a spectrum, and paradigmatic robots out of science fiction are at the middle of the spectrum, we can place so-called humanoid robots at the other extreme end of the spectrum. These are robots created to look or act like human beings. Sophia the robot is one example. Another example is a robot created by the Japanese robotics researcher Hiroshi Ishiguro. This robot is a replica of Ishiguro himself, which looks and, to some extent, behaves like Ishiguro. In general, then, a robot is a machine that is embodied, that can sense its environment, and that can respond to the environment based on the information it takes in, so as to perform some task or set of tasks.

1.65 When robots or other technologies are able to imitate or replicate human behaviors in a seemingly intelligent way, they are sometimes said to have "artificial intelligence" or "AI," in short. In other words, a technology or technological system is said to possess *artificial intelligence* when it is able to perform tasks that humans (or animals) use their intelligence to perform. In the most widely used textbook about AI, written by Stuart Russell and Peter Norvig, a two-fold distinction is made, which is worth repeating here. On the one hand, Russell and Norvig distinguish between technologies that

are meant to either *behave* or *think* in intelligent (or intelligent-seeming) ways. On the other hand, they distinguish between technologies attempting to reach a *human level* in their performance of some task and technologies designed to reach an *optimally rational level* in their performance of their tasks. According to this way of thinking, there could be machines behaving at a human level, thinking at a human level, or perhaps behaving or thinking at a more optimally rational level than humans are able to.

We will have occasion to discuss different forms of robots and AI throughout the book. But one last introductory observation might be of interest here. Namely, while the term "robot" was introduced in science fiction, the term "artificial intelligence" comes from a scientific context. The word "robot" comes from a 1920 stage play by the author Karel Čapek, in which artificial humans ("robots") are created to work in a factory. In contrast, the term "artificial intelligence" was introduced by a team of scientists, in a research proposal from 1955. The scientists set themselves the goal of creating machines that would be able to perform tasks that we human beings need our intelligence to perform. 1.66

In short, then, the term "robot" comes from fiction, whereas the term "artificial intelligence" comes from science. But in both cases, our ideas about these things tend to be influenced and colored both by science and by science fiction, and sometimes primarily by the latter. So, it is good to be careful and to be aware of how strongly influenced we might be by ideas from fiction rather than from science when we think about technology ethics. This might not be a bad thing. Science fiction has produced many interesting ideas and concepts. But it can mean that our expectations of what robots and AI can do or be might sometimes be overly optimistic, if not outright unrealistic. 1.67

Another kind of technology that will be of interest in some of the chapters to come include *(self-)tracking and data-logging technologies*, which might also be so-called *behavior change technologies*. These are technologies that are used to track, log, or otherwise gather and make use of data and information that might be generated by our actions or behaviors, whether online or offline. Some of these technologies might be designed to influence or change the ways we behave, by stimulating additional motivation to act in some way, for example by creating incentives or other prompts to change our behavior. Technologies can do so, for example, by *gamifying* certain behaviors or activities. This means that they turn something that is ordinarily not conceived of as a game into a game or something game-like. In other words, game-like elements are introduced into some otherwise not necessarily game-like activity or domain. 1.68

1.69 A running app that one might have on one's phone can explain all three of these ideas. Such an app is likely to track information/data about how much, how often, where, and how fast the user is running. This information, and information about how other people run, might then be used in a way that gamifies the user's running: there might be points given, achievements declared, or a leaderboard where the user can see how they are doing in comparison to other runners. This will make exercising by running more like a game. And it might influence and motivate the user to run more often, run longer distances, or run faster.

1.70 *Information and communication technologies* will also be important throughout the book, as well as in any discussion about technology ethics. Broadly, this includes any forms of technologies we use to seek information or to communicate with those around us, such as search engines, social media platforms, or the devices we use for such purposes, such as computers or smartphones. Importantly, some information and communication technologies might be partly gamified, and they might track behavior, collect data about us, and so on. In general, from an ethical point of view, the idea of *data-collecting technologies*—i.e., technologies that collect data about people for different purposes—is a controversial form of technologies. They are ethically controversial since users may not know what will happen to their data, and since having their data might give others (such as big tech companies) some control or power over them.[13]

1.71 We will also occasionally be discussing *medical technologies* (i.e., technologies used for medical purposes or perhaps broadly having to do with health) in distinction to technologies for other domains of life. Some technologies may start as medical technologies—or, say, as *military technologies*—and then start also being used in other domains, for other purposes than those originally intended. This is a common occurrence. Ethicists of technology often worry about the potential for what is sometimes called the "dual use" of technologies: they may have been designed and intended for one domain (e.g., some medical application), but then be co-opted for use in some other domain as well (e.g., some questionable military use), which the inventors or designers of the technology may not be fully on-board

[13] In an episode of the "Philosophy 247" podcast, you can hear the philosopher Carissa Véliz being interviewed about ethical issues related to data-collecting technologies: "The Future of Privacy": https://philosophy247.org/podcasts/the-future-of-privacy/

with. This is related to worries about unintended consequences of the use of the technologies, another topic that we will have plenty of occasions to return to in the chapters that follow.

Another form of technology that is worth mentioning here as a fascinating kind of technology are *brain stimulation technologies* or *brain implants*. These are technologies used to target particular brain regions, such as, for example, the sort of deep brain stimulation technologies that is used in one form of treatment of Parkinson's disease and other neurologically based afflictions. This is an example of a kind of technology that might have been developed for medical use, but that might conceivably later be put to other uses as well, instead of the ones for which the technologies were originally designed. 1.72

These just-mentioned different types of technologies—robots, AI, brain implants, and so on—are all quite different than the traditional forms of technologies that Heidegger philosophized about and had in his hut, which we visited at the beginning of the chapter. But, as was already noted at the beginning of the chapter, we should not give in to the temptation of only regarding these kinds of high-tech emerging technologies as technologies. As noted above, older technologies are also technologies, just like newer ones are. 1.73

Actually, Heidegger makes another point about technologies that is worth mentioning as we bring the discussion in this chapter to a close. He notes that when technologies are working well and we have become accustomed to them, they sort of fade into the background and become almost transparent to us. We do not really notice them. Instead, as the post-phenomenologist philosophers note, we start to perceive reality through our technologies and act through our technologies as well. It is typically only when technologies stop working properly or when new technologies are introduced and when we are not yet used to them that we notice them and that they are in the foreground of our experience. We quickly come to take technologies for granted. This is something worth keeping in mind as we discuss ethical questions related to different technologies. 1.74

As noted earlier on, what we will do in the next chapter is to consider the question of what ethics is or what it should be taken to be. It turned out in this chapter that explaining what technology is isn't as straightforward as one might think. Even if Heidegger was of the opinion that everyone will agree with the instrumental and anthropological definition of technology, we have seen that there can be quite a bit of disagreement about how one should think about what technology is. We will see in the next chapter that 1.75

the same applies to ethics. It is one of those concepts that when we try to define it, things quickly become contentious, and disagreements arise.

1.76 In the case of ethics, this is to be expected. Ethics is a normative enterprise. It has to do with what is good and bad, right and wrong, meritorious or shameful, and so on. And it is in the nature of everything normative that people tend to disagree about it. So, it is with ethics in general. And so it is with technology ethics in particular. This is part of what makes it interesting.

Annotated Bibliography

Sarah Bakewell (2017). *At the Existentialist Café: Freedom, Being, and Apricot Cocktails with Jean-Paul Sartre, Simone de Beauvoir, Albert Camus, Martin Heidegger and Others*, London: Other Press.

An accessible introduction to the ideas of, among others, Martin Heidegger and the phenomenologists. Contains more information about Heidegger's hut, as well as about his controversial time as a member of the Nazi party.

Joanna Bryson (2010). "Robots Should be Slaves," in *Close Engagements with Artificial Companions*, edited by Yorick Wilks, Amsterdam: John Benjamins Publishing Company, pp. 63–74.

An oft-quoted contemporary defense of the instrumental theory of technology, with a particular focus on robots.

David Gunkel (2018). *Robot Rights*, Cambridge, MA: The MIT Press.

A deep-dive into the literature about non-instrumental perspectives on robots. Discusses the idea of treating robots as persons and giving them moral consideration.

Martin Heidegger (1977). *The Question Concerning Technology, and Other Essays*, New York: Garland Publishing.

Not exactly an easy read, but the most accessible of Heidegger's writings, and a bit of a classic within the philosophy of technology.

Peter-Paul Verbeek (2011). *Moralizing Technology: Understanding and Designing the Morality of Things*, Chicago: Chicago University Press.

A defense of the post-phenomenological perspective, including the mediation theory of technology, by one of its leading exponents.