Ethics of *Socially Disruptive Technologies*

An Introduction

Edited by
Ibo van de Poel,
Lily Frank, Julia Hermann,
Jeroen Hopster, Dominic Lenzi,
Sven Nyholm, Behnam Taebi, and
Elena Ziliotti

https://www.openbookpublishers.com/

ISBN Paperback: 978-1-80511-016-3
ISBN Hardback: 978-1-80511-017-0
ISBN Digital (PDF): 978-1-80511-057-6
ISBN Digital ebook (EPUB): 978-1-78374-789-4
ISBN Digital ebook (XML): 978-1-80511-050-7
ISBN Digital ebook (HTML): 978-1-80064-987-3
DOI: 10.11647/OBP.0366

Cover image: Blue Bright Lights, Pixabay, 6th April 2017, https://www.pexels.com/photo/blue-bright-lights-373543/
Cover design: Jeevanjot Kaur Nagpal

3. Social Robots and Society

Lead author: *Sven Nyholm*[1]
Contributing authors: *Cindy Friedman, Michael T. Dale, Anna Puzio, Dina Babushkina, Guido Löhr, Arthur Gwagwa, Bart A. Kamphorst, Giulia Perugia, Wijnand IJsselsteijn*

Advancements in artificial intelligence and (social) robotics raise pertinent questions as to how these technologies may help shape the society of the future. The main aim of the chapter is to consider the social and conceptual disruptions that might be associated with social robots, and humanoid social robots in particular. This chapter starts by comparing the concepts of robots and artificial intelligence and briefly explores the origins of these expressions. It then explains the definition of a social robot, as well as the definition of humanoid robots. A key notion in this context is the idea of anthropomorphism: the human tendency to attribute human qualities, not only to our fellow human beings, but also to parts of nature and to technologies. This tendency to anthropomorphize technologies by responding to and interacting with them as if they have human qualities is one of the reasons

1 SN is the lead author of this chapter. He coordinated the contributions to this chapter and, together with MD, he did the final editing of the chapter. SN wrote the first versions of Sections 3.1. and 3.3. and contributed material to all of the other sections. CF wrote the first version of Section 3.2. and also contributed material to Sections 3.3. and 3.4. MD wrote the first version of Section 3.4 and contributed to Sections 3.2 and 3.3. AP contributed material to all sections. GL contributed to Sections 3.1. and 3.3. DB contributed to Section 3.2. AG contributed to Section 3.4. BK commented on the whole chapter draft and suggested various edits to all sections. GP contributed to Sections 3.1 and 3.2. WI contributed to Section 3.3. All authors and contributors approved the final version.

 https://doi.org/10.11647/OBP.0366.03

why social robots (in particular social robots designed to look and behave like human beings) can be socially disruptive. As is explained in the chapter, while some ethics researchers believe that anthropomorphization is a mistake that can lead to various forms of deception, others — including both ethics researchers and social roboticists — believe it can be useful or fitting to treat robots in anthropomorphizing ways. The chapter explores that disagreement by, among other things, considering recent philosophical debates about whether social robots can be moral patients, that is, whether it can make sense to treat them with moral consideration. Where one stands on this issue will depend either on one's views about whether social robots can have, imitate, or represent morally relevant properties, or on how people relate to social robots in their interactions with them. Lastly, the chapter urges that the ethics of social robots should explore intercultural perspectives, and highlights some recent research on Ubuntu ethics and social robots.

Fig. 3.1 Social Robots. Credit: Menah Wellen

3.1 Introduction

While the expression 'artificial intelligence' comes from computer science, the word 'robot' comes from science fiction. The word was coined by a Czech playwright — Karel Čapek — in his 1920 play *R.U.R.: Rossum's Universal Robots*, which premiered in January of 1921, a little over 100 years before this book was written (Čapek, 1928; Nyholm, 2020). The robots in that play were similar to what many people still imagine when they hear the word 'robot' today: silvery/metallic artificial humans, or entities with a vaguely humanoid form, created to do work for us human beings. The robots in that play work in a factory. Towards the end of the play, the robots want their freedom and they want to know how to create their own robot children, so they do not have to depend on their human creators anymore. As it happens, the word 'robot' derives from the Czech language word 'robota', which roughly means 'forced labor'. The expression 'artificial intelligence', in contrast, was introduced in a 1955 research proposal for a summer workshop that took place at Dartmouth College in Hanover, NH, in 1956 — where the researchers proposed to create technologies that could 'simulate' all aspects of human learning and intelligence that could be precisely described (Gordon and Nyholm, 2021).

The development of robotics and artificial intelligence have both come a long way since 1920 and 1956 respectively, but not, perhaps, as far as many envisioned at several points in between then and now (Russell and Norvig, 2005; Dignum, 2019). These days, philosophers and others who write about or do research on robots typically do not mean artificial humans that work in factories when they use the word 'robot', though that is one of the ideas from science fiction that is still with us today (Gunkel, 2018). In fact, the tech entrepreneur Elon Musk presented a similar vision in August of 2021, when he presented his idea for the 'Tesla Bot' during a publicity event for Tesla. What he presented was the idea of a robot with a humanoid form that would work in Tesla factories, so that humans would not need to do that work anymore — a little bit like the robots in Čapek's play (Nyholm, 2023).

What do researchers who write and do research on robots now mean by the term 'robots'? And what are social robots? Many researchers are reluctant to give precise definitions of what one should understand

by the word 'robot'. There are, they say, so many things that are called 'robots' that it is difficult to articulate what they all have in common; and if we follow some common definitions of what robots are, there are some things that qualify as robots, e.g., smartphones, that do not intuitively seem to be robots (Gunkel, 2018). Nevertheless, when researchers do offer definitions of what they mean by the word 'robot', they usually say something along the following lines: robots are embodied machines with sensors with which they receive information about their environment, and with actuators with which they can respond to their environment, in the service of certain specified tasks (Loh, 2019; Nyholm, 2020).

Researchers sometimes talk about the 'sense, plan, act' paradigm regarding how to understand what a robot is: it is a machine that can *sense* its environment, *plan* what it can do to achieve its task, and then *act* so as to achieve its task (Gunkel, 2018). A Roomba vacuum cleaning robot, for example, senses its environment as it moves around in a room; it detects obstacles (e.g., furniture in its way); and then it takes action so as to be able to continue vacuuming (e.g., moving around the furniture). A Roomba vacuum cleaning robot does not look like a paradigmatic robot out of science fiction. It looks more like a hockey puck or a beetle. But it is a robot by most common definitions of the term. In contrast, it is important to note here that the Roomba (by most accounts) is very limited with respect to its artificial intelligence. The two terms 'artificial intelligence' and 'robots' do not always pick out the same set of things.

A social robot is a robot that is designed to be able to interact with human beings in interpersonal ways (Breazeal, 2003; Darling, 2016). For example, a social robot might respond in a reactive/social way to touch, might have a chat function, or might in other ways respond to human interaction in the way a social being can be expected to. Such a robot does not have to look like a paradigmatic robot out of science fiction either (e.g., like the robots in the classic 1927 film *Metropolis*) but can take different forms. A well-known social robot is the robot seal Paro, which looks like a baby seal and responds to interaction with human beings in a way that appears interactive and soothing to some human beings. To give another example of a social robot from science fiction: R2-D2 from the movie *Star Wars* is a social robot.

Importantly, some social robots take on a humanlike form: a humanoid robot is a robot that is designed to look and behave like a human being

(Zhao, 2006; Friedman, 2022). The advantages and disadvantages of the humanoid form are discussed under the heading *anthropomorphism* (Friedman et al., 2022). Some humanoid robots reproduce the human body and behavior in subtle and stylized ways — as is the case for robots like NAO and Pepper. Other humanoid robots, instead, mimic the human body and behavior in extremely realistic ways — as is the case for robots like Geminoid HI-5 and Erica. These latter humanoid robots, which are conceived as robotic twins of existing (Geminoid HI-5) or imaginary persons (Erica), are called android and gynoid robots depending on whether they resemble a man or a woman. One well-known example of a gynoid robot is the robot Sophia from the company Hanson Robotics. Sophia is well-known, and controversial, for having generated various social responses in people, including being interviewed on popular TV shows (such as *The Tonight Show with Jimmy Fallon*), being invited to speak in front of the UN, and being named an honorary citizen of Saudi Arabia (Nyholm, 2020: 1–3).

Sophia and Hanson Robotics have been criticized by many technology experts and ethicists: the robot is deceptive, it has been argued, because it is presented as having a much more advanced form of artificial intelligence than it really has (Sharkey, 2018). Another controversial type of humanoid robot is the sex robot: robots created specifically for sexual purposes, but which are sometimes also presented as potential romantic companions for human beings, i.e., as not only being intended for purely sexual purposes (Richardson, 2015; Danaher and McArthur, 2017). The sex robots of today — usually a gynoid robot designed to closely resemble a human woman, though there are also prototypes that look like human men — are fairly rudimentary. But given how fast technological developments can be, it may be reasonably predicted that they and other forms of social robots might become extremely impressive and lifelike within the lifetimes of many of the people who are already alive today (Levy, 2008). We are not there yet, though (Nyholm, 2023).

Of related interest here are disembodied 'bots', such as Amazon Alexa, Siri, or Google assistant, or the chatbots that we interact with via chat windows in our browsers (any kind of customer service chatbots that filters customer complaints and decides whether to escalate an issue to a human). These bots are meant to interact with users through one-dimensional interactions (voice or text), and often maintain the

artificiality of the interaction at the forefront. Even more impressive are the recently developed large language models using so-called transformer technology, like Google's LaMDA or OpenAI's ChatGPT, which specialize in what is presented as a form of 'conversation' with the user. Notably, LaMDA responds to inputs from users in an impressive enough way that one of Google's engineers, Blake Lemoine, famously went to the media to declare that he thought that LaMDA had become a 'sentient' 'person', who should be entitled to rights. To some commentators, the chat transcripts that Lemoine made public were not proof that LaMDA was conscious, but rather proof that these AI technologies will increasingly become able to deceive or at least confuse human users into thinking that they have more advanced properties than they already have (Bryson, 2022). In a sense, this can be seen as technologies that deskill humans with respect to the ability to tell the difference between fellow sentient beings and machines without a 'soul', another thing that Lemoine thought that LaMDA had.

One technology that has received less attention so far, but which is also of interest in this context, is the religious robot: social robots used in religious settings, which are particularly prevalent in non-monotheistic religions and the non-Western world. Religious robots attempt to mimic the spiritual and religious dimensions of being human. They can be used in a variety of ways and take on different functions. Religious robots could accompany religious rituals and ceremonies (e.g., the robot Pepper at funerals or Mindar reciting the Heart Sutra in a Japanese temple), bless (e.g. BlessU2), imitate religious conversations with patients in hospitals, recite Bible passages and religious narrations (e.g. SanTo), or engage in acts that are interpreted to bring luck, and offer protection. Thus, as social robots are increasingly developed, the question arises whether they will be presented as being atheistic, agnostic, or as belonging to a religion and having faith (Puzio, 2023).

Besides these more specific domains of application, social robots are increasingly used in education and healthcare — for instance, to help children learn higher-order thinking skills such as creativity (Elgarf et al., 2022) or to nudge people towards seemingly healthy behavioral outcomes (e.g., losing weight as in Kidd and Breazeal, 2008). As has been seen above, there is a wide range of social robots — either already in existence or in prototype form. In the future, it is to be expected

that social robots will be used in an even wider range of domains of human life. At that point, many new ethical questions will arise about how we should interact with these robots. Yet already today, social robots — perhaps especially social robots with humanoid forms — raise ethical concerns and have the potential to be socially disruptive.

3.2 Impacts and social disruptions

Social robots are both impactful and socially disruptive. They force us to question the meanings of such concepts as sociality, care, relationships, relationality, and community, and more generally the issue of what constitutes social relationships (Zhao, 2006; Turkle, 2020). How is the relationship with a technology different from the relationship of humans to other humans or to animals? What makes relationships valuable, and do they necessarily rely on reciprocity? Below, different ways in which social robots might be socially disruptive or otherwise disruptive are described.

Social robotics researchers are often thinking about ways to improve social interactions between social robots and humans. Indeed, they study what makes humans enjoy interacting with social robots and accept them as social agents (Frennert and Östlund, 2014; Darling, 2016). To gain insights into what it means to engage in social behavior, researchers often turn to important components of human sociality. For example, mimicking other people's behavior is commonly understood to be an important part of human-human relationships, indicative of rapport building (Tickle-Degnen and Rosenthal, 1990). Due to this, some argue that social robots should also be capable of mimicry (Kahn et al., 2006).

Another key aspect to human sociality is reciprocity (Gouldner, 1960; Lorenz et al., 2016). Reciprocity is commonly understood as follows: '[W]e should try to repay, in kind, what another person provided us' (Cialdini, 2009). Or, put more simply: 'If you do something for me, I will do something for you' (Sandoval et al., 2016). Due to its importance in human relationships, robotics researchers have considered to what extent reciprocity should and can be implemented in social robots. Many claim that social robots should be capable of reciprocity (Kahn et al., 2006), pointing to empirical data that reveals that humans enjoy interacting with reciprocating computer programs (Fogg and Nass, 1997). However,

others have pointed out that a seemingly reciprocal relationship between a human and a social robot is a deceptive relationship. Specifically, van Wynsberghe (2022) claims that a robot cannot engage in a truly reciprocal relationship. It is only using reciprocity to become more socially accepted by the human, and thus the relationship is founded on deception. Similarly, Robert Sparrow and Linda Sparrow (2006) argue that a relationship can only be meaningful when it occurs between social entities capable of reciprocal affect and concern.

Social robots are explicitly designed to draw upon people's fundamental social-relational capacities. Specifically, they are designed to draw upon the tendency that human beings have to *anthropomorphize*. The tendency to anthropomorphize is an evolutionary adaptation that people have to attribute human characteristics to that which is not human (Epley et al., 2007; Damiano and Damouchel, 2018). For example, humans tend to see faces in random patterns of objects or shapes (a phenomenon known as pareidolia) and tend to see social meanings in the movements of geometric figures (Heider and Simmel, 1944). When a child talks about her teddy bear being sad, the child is anthropomorphizing the teddy bear.

Anthropomorphization of social robots need not only come in the form of, or as a response to, physical appearance (cf., Sophia the robot or Ai-DA). Disembodied chatbots are examples of social robots or bots that we anthropomorphize, but not by designing them to appear human. Instead, we anthropomorphize them in the sense that we assume that they perform a very human action: they talk! In fact, most large language models of today, like Google's LaMDA or OpenAI's ChatGPT, simulate a conversation, but in fact only output a set of words that they compute as being the most likely to come next after a prompt, based on a huge amount of natural language data. This is clearly a very different kind of linguistic agent compared to a human conversational partner who has intentions, plans, and desires when she talks to you, and who can make commitments and take on obligations (Bender et al., 2021).

Traditionally, the tendency to anthropomorphize robots has been cast in a negative light (Bryson, 2010). It has been viewed as a 'bias, a category mistake, an obstacle to the advancement of knowledge, and as a psychological disposition typical of those who are immature and unenlightened' (Damiano and Damouchel, 2018: 468). However, social

roboticists have seen the tendency to anthropomorphize as less of an obstacle, and more of a tool, which can be utilized to support and improve social exchanges between humans and robots (Gunkel, 2018). Research shows that people perceive computers and virtual characters as social actors (Nass and Moon, 2000). The embodiment and physical movement of robots further amplify this perception (Darling, 2016). As de Graaf (2016) explains, the physical presence of social robots and their capacity to speak and use humanlike gestures or facial expressions encourage people to interact with social robots as if they are human, and not simply a type of technology. Leveraging on this, roboticists have designed social robots to display emotions (e.g., facial expressions of happiness and anger), personality (e.g., introversion and extraversion), and even gender (Paetzel-Prüsmann et al., 2021; Perugia et al., 2022).

The magnitude of the potential effects social robotics may eventually have on social imagery, normativity, and human practices has led some researchers, such as Seibt (2016), to discuss the creation of social robots as a form of 'socio-cultural engineering'. For example, creating robots with apparent social skills, and thus making robots more like humans in their behavior, potentially comes hand-in-hand with the opposite tendency: encouraging humans to mimic robotic ways of doing things (Sætra 2022). Accordingly, the field of social robotics challenges socio-cultural sustainability, i.e. our ability to robustly maintain familiar cultural and social norms and practices (Gunkel 2023). The question arises of which of our human beliefs, norms, and practices that are rooted in tradition, culture, and social institutions are worth fighting for, even at the expense of technological innovation. According to Babushkina (2021a), social robotics in effect also brings us face-to-face with a problem of moral sustainability, i.e. 'the preservation of rationally justifiable moral values, norms, and practices' (Babushkina, 2021a: 305).

A reasonable goal in this context is to prevent a situation in which our moral practices change beyond what makes sense to us as human moral agents, rendering some of our interpersonal interactions absurd. Even though it might be difficult to grasp the elusive meaning of 'making sense', it is a fundamental need of a human being in her relationship to the world, be it co-existence with others, interaction with the environment, or experience of her own self. One of the main problems with social robots is that they get introduced as players into

interpersonal relationships, i.e. the relationships that until now were only reserved for humans (e.g., companionship, friendship, parenthood, collegiality: Zhao, 2006). This means that social robots get plugged into various forms of intersubjectivity, apparently assuming the role of a partner in a relationship, but typically effectively failing to perform key functions that are morally required from the partner. What is significant from the moral-psychological point of view, for example, is that robots fail to meet expectations and answer reactive attitudes that we are justified to have towards partners in such relationships. This potentially leads to absurd experiences.

Following Wilks (2010), we can imagine a care robot presented as capable of ensuring the well-being of an elder, including giving her advice about weather-appropriate clothing. One day the companion gives the wrong information and the elder gets sick. You try to complain to the company, but it refers you to a small print where any blameworthiness is denied and users are advised to use the robot at their own risk. Such clashes between interpersonal expectations and robotic reality may create a dilemma: either to rethink moral responsibility so that it can accommodate artificial agents (e.g., Floridi and Sanders 2004; Sullins III 2006; Gogoshin 2021; Babushkina 2022) or limit the extent to which robots should be allowed to take on important roles associated with interpersonal relationships.[2]

Moreover, some have raised concerns that the implementation and use of social robots may negatively impact us should we allow them to crowd out human relationships. We are already seeing something similar occur in Japan, as some men there have shown less interest in starting relationships with human romantic partners, due to the possibility of instead having a 'virtual girlfriend' (Rani, 2013; cf. Nyholm, 2020: Chapter 5). Therefore, the possibility for this to occur with social robots as well is not all that far-fetched.

2 Another example of social robots challenging the fundamental attitudes underlying interpersonal relationships concerns respect. The stronger the need for seamless integration of robots into the interpersonal sphere, the stronger the demand will be for them to be respectful. However, trying to stretch the concept of respectfulness to artificial agents may lead to identification of respect with external behavioral expressions and atrophy of respect as an attitude based on inherent appreciation of human value (Babushkina, 2021b).

Scholars have approached this concern from various angles (Friedman, 2022). Some are worried that the relations we have with social robots may negatively impact our human well-being and quality of life. For example, in the context of care robots for the elderly, these social robots may negatively affect the well-being of the elderly, should they lead to a reduction of human contact, given the importance of human contact for stress reduction and the prevention of cognitive decline (Sparrow and Sparrow, 2006; Sharkey and Sharkey, 2012).

Moreover, Turkle (2011), in her discussion about the 'robotic moment', has voiced the concern that replacing human relations with robotic ones will lead to social isolation, given the illusory nature of human-robot relations. In the context of sex robots, for example, Nyholm and Frank (2019) argue that these robots may block off some people's relations with other people, and that this is something about which we should be concerned, given the premise that human-human relationships are more valuable than human-robot relationships. More generally, Danaher (2019) has argued that in forming relations with robots, people may be less likely to go out into the world and express their moral agency, which may lead to them being reduced to mere moral patients who passively receive the benefits that the technologies bestow.

Many researchers also worry that the relationships people form with social robots may negatively reinforce human stereotypes. In this context, Perugia and Lisy (2022) have noticed how the gender of a humanoid robot transforms the value of the interaction people have with it and might take on normative meanings for human society. For instance, using female robots in service and care-taking scenarios risks reinforcing normative assumptions about gender roles in society (Guidi et al., 2023). They invite roboticists to critically reflect on the ethical implications of gendering humanoid robots, especially considering the highly symbolic value of human-humanoid interactions for human-human relations.

3.3 Conceptual disruption

The way people respond to social robots places these robots in a confusing ontological space in society (Gunkel, 2023). Social robots are, essentially, a technological artifact, yet there is a tendency to perceive them as something more than this (Strasser, 2022). Specifically, social

robots are blurring the line between being alive and being lifelike: we intuitively perceive them as being alive in some sense, although we are aware that they are not (Carpinella et al., 2017; Spatola and Chaminade, 2022).

Moreover, social robots challenge the boundaries between animate and inanimate, human, animal and machine, body and technology. They challenge the understanding of the human being anew. For example, in response to social robotics, we need to ask what emotions are, what constitutes action, what constitutes a relationship with the body. In the context of robotics more generally, questions also arise as to where the boundary between our human body and technology lies. Can technology be understood as part of the human body? Disability studies have shown that wheelchairs or prostheses are also sometimes perceived as part of one's own body. In a similar way, robots can potentially contribute to a broader, more inclusive understanding of the body (Thweatt, 2018; Graham, 1999; Puzio, 2022).

As we have seen in the introduction to this book, the uncertainty about which concepts we should use or apply when interacting with a new technological artifact is a form of conceptual disruption (see also Löhr, 2022). A conceptual disruption occurs if we either have no concepts to classify something or if two or more conflicting concepts seem to apply more or less equally well, such that we have to make a conceptual decision (is it dead or is it alive?). Such decisions are often difficult to make, but since we cannot leave objects uncategorized if we want to talk about them or act in relation with them, we often have no choice but to make a decision eventually.

Social robots can also have disruptive impacts on people's emotional lives. Some people have gone so far as to form deeply emotional social bonds with social robots, due to the perception that they are alive or in the possession of personalities. For example, in Japan, Sony's AIBO robots (which take the form of a dog) were honored with funeral ceremonies, when older models could no longer be updated. Although having 'doggish' behaviors, such as the ability to wag its tail, the AIBO robot also had human-like features, such as the ability to dance and, in later models, speak. Thus, many AIBO owners anthropomorphized these robots and subsequently formed deeply emotional bonds with them. As such, in 2014, when Sony announced that they would no longer support

updates to older models, some AIBO owners perceived this message as a much more somber one: their pet robot dogs would die (Burch, 2018). In this same vein, the philosophers Munn and Weijers (2022) have recently suggested that when people get attached to technologies (such as the chatbot app Replika), this might create novel forms of ethical responsibilities for the tech companies behind these technologies, e.g., not deleting the apps, since this could be seen by some users as being a way of 'killing' their new friend (for further discussion, see Nyholm, 2023: Chapter 9).

The social response of perceiving these robots as being alive or as having a personality (and particularly humanlike) when they are not and do not, can be seen as ethically problematic or disruptive in the sense that human users are being deceived or even manipulated. Some argue that it is unethical to allow ourselves, or to cause others, to be deceived, if we assume that we have a duty to see the world as it is (Sparrow and Sparrow, 2006). In response to this, however, it has been pointed out that an animal using camouflage is a kind of deception, yet we do not find anything morally problematic about that (Sharkey and Sharkey, 2021). Moreover, sometimes deception has positive consequences, such as when baby dolls are introduced to people with dementia to help stimulate memories of a rewarding life role they once had (Mitchell and O'Donnell, 2013). Furthermore, the question arises as to when one should speak of deception as opposed to, say, make-believe. Children are raised with imaginary children's book characters, Disney film characters, and cuddly toys without this being considered deception or ethically reprehensible.

With these nuances in mind, Danaher (2020) argues that a form of deception wherein a robot deceives us into thinking it has a capacity it actually lacks is not necessarily ethically concerning. However, he does contend that deception in which a robot conceals the presence of a capacity which it does actually possess is seriously concerning. In the case of people with dementia — who are more likely to 'be unable to distinguish simulated or mediated reality from actual reality' — while there may be some positive consequences to using baby dolls to trigger certain memories, it does not take away from the fact that such dolls may be conduits of deception (Tummers-Heemels et al., 2021: 19). Thus,

we should allow such instances of deception only 'sparingly, and with integrity and restraint' (Tummers-Heemels et al., 2021: 10).

Others, meanwhile, see robot deception as tolerable and even somewhat inevitable given the functions and purposes of the robots (Wagner and Arkin, 2011; Wagner, 2016). Indeed, just as humans sometimes use deception in their social interactions (such as when it is important to keep information private), it might be useful for a social robot to at least have the capacity to deceive. However until these questions are ultimately settled, it remains the case that conceptual disruption occurs. That is, these robots challenge our ordinary ontological distinctions between persons on the one hand, and things on the other. They seem to occupy some space in between these two extremes, at least with respect to how we intuitively respond to social robots (Strasser, 2022; Gunkel, 2023). Highlighting this form of ontological disruption lays a foundation for an understanding of why, and how, social robots are also potentially morally or, more broadly, conceptually disruptive.

Social robots not only encourage us to rethink our understanding of the human being; they are potentially also fundamentally changing anthropology. Anthropology as a field is increasingly turning away from essentialist conceptions of an imagined 'human nature' towards non-essentialist, dynamic, and fluid understandings of human identity. In particular, movements of thought such as New Materialism and Critical Posthumanism, which have been strongly influenced by the thinking of Donna Haraway among others, are striving to break down old anthropological concepts and dichotomies (of animate-inanimate, human-animal, human-machine, nature-culture/technology, woman-man). Haraway (1985) influentially discussed the ontological, epistemological, and political figure of the cyborg, which as a 'cybernetic organism' has a hybrid, fluid, and dynamic identity. The cyborg is neither unambiguously human, animal nor machine, thus refusing any categorization and classification and therefore maintaining subversive potential to resist any reontologization by humans.[3] Critical

3 The expression 'reontologization' here refers to the attempt to redefine what something is — i.e. to put it into a new or slightly different category in response to some new technological development or scientific discovery. Posthumanists tend to resist limiting definitions of what it is to be human, because they think that being human is open-ended, partly due to our 'cyborg'-like nature that is related to how we merge with the technologies we use.

Posthumanism and New Materialism thus reflect anew on notions of human, body, life, nature, etc. They draw attention to the fact that technologies such as social robots blur and question the above-mentioned boundaries and also seek to redraw these boundaries (Puzio, 2022).

The conceptual disruption of ontological concepts and categories caused by social robots also potentially creates a disruption of moral concepts and values, given the view that what an entity is, or is perceived as being, usually determines its moral status. Specifically, there may be a disruption in the context of our moral relations with social robots. Luciano Floridi (2013: 135–36) notes that 'moral situations involve at least two interacting components — the initiator of the action or the agent and the receiver of this action or the patient'. As Floridi sees things, robots can be moral agents but not moral patients. However, many authors who discuss the ethics of human-robot interaction disagree (for an overview, see Nyholm, 2021). They think that social robots can be both moral agents and moral patients. Moreover, the question arises as to what agency means and what it requires. For example, does agency presuppose consciousness? Some roboticists and philosophers — e.g. Asada (2019) and Metzinger (2013) — take seriously that it might be possible to create conscious robots. The well-known and influential philosopher of mind David Chalmers has even recently taken seriously the possibility that large language models might at some point become conscious.[4] However, this is controversial, and it also poses the difficulty that consciousness cannot easily be defined (Coeckelbergh, 2010a; Gunkel, 2018).

The different views about whether and why social robots can potentially be seen as moral patients can be divided into four broad classes, the first three of which relate the patiency of robots to their properties. These views can all be explained with reference to the following set of questions (Nyholm, 2023). The first question is: can social robots *have* morally relevant properties or abilities? Notably, most authors discussing this question are skeptical about the idea of current robots having morally relevant properties/abilities such as sentience or rationality/intelligence. However, some authors (e.g. Bryson, 2010;

4 In a presentation at New York University, Chalmers (2022) discusses the topic 'Are large language models sentient?'. Video available here: https://youtu.be/-BcuCmf00_Y

Metzinger, 2013; Schwitzgebel and Garza, 2015) think that it is possible to create social robots that could be conscious or have feelings and intelligence like human beings, and that such future robots should be treated with moral consideration.

Another question is whether robots can *imitate* or *simulate* morally relevant properties or abilities. This is perhaps more realistic. Danaher (2020), for example, focuses on this idea, and argues that if robots consistently behave like human beings with moral status behave, we should treat these robots with moral consideration, independently of whether we can establish whether anything is going on within their 'minds' (Coeckelbergh, 2010b). While Véliz (2021) argues that technologies can neither be moral agents nor moral patients because they are 'moral zombies' without consciousness or feelings, Danaher argues that what matters is instead whether they consistently behave as if they do. This is a kind of ethical Turing test, one could say.

Yet another question is whether social robots could *symbolize* or *represent* morally important properties or abilities. This expects even less of technology. Sparrow (2017; 2021) argues that robots and our interaction with robots represent various different morally important ideas, which means that how we treat, and interact with, robots is not morally neutral. In particular, Sparrow thinks that how we interact with robots — and how robots are made to appear to us — can represent various things that are highly problematic from an ethical point of view. Like Richardson (2015), Sparrow (2017) discusses sex robots as a key example of this, and they both think that human interaction with sex robots will almost inevitably represent morally problematic ideas — such as tropes associated with so-called rape culture. According to Sparrow (2021), while our interaction with robots could represent negative moral ideas, it is much harder — if not impossible — for human interaction with robots to represent or symbolize morally good ideas. Treating a robot 'well' cannot, Sparrow thinks, reflect well on a person, whereas treating a robot in a 'cruel' way (e.g. kicking a robot dog) can reflect poorly on us and our moral character.

A further type of view — which seeks to turn the idea of focusing on the properties or abilities of the robots on its head — says that the question we should be asking is not whether robots have, imitate, or symbolize morally relevant properties/abilities. We should instead be

asking whether people *relate* to, or are disposed to relate to, (certain forms of) robots in ways that seem to treat the robots with moral consideration, and that welcome them into the moral community. Coeckelbergh (2010a) and Gunkel (2018) call this the 'relational' view of the moral status of robots. Chris Wareham (2021) defends a version of that view which appeals to the Ubuntu idea that 'we become persons through other persons'. According to Wareham, social robots can become persons through other persons, just like humans can: if the social robots are treated like persons and are welcomed into the moral community. Loh (2022) argues that a post-human perspective on human-technology relations favors this kind of relational view. According to Loh (2019), when somebody tends to treat a robot like a moral patient, a friend, or even a romantic partner, this is not a 'shortcoming' but a 'capability', which can be celebrated as part of human diversity.[5] Others, like Müller (2021), think that such views are deeply misguided. According to Müller, while we might wrong the owner of a social robot if we 'mistreat' their social robot (which the owner might presumably be attached to), we cannot wrong the social robot itself any more than we can wrong a pencil or a toaster — though here too we might wrong their owners if the owners are very attached to those.

Furthermore, the question arises whether this topic of moral agency, moral patiency, and the moral community is at all an appropriate and important question or whether discussion of this set of issues instead distracts people away from more urgent questions robot ethics should focus on instead (Birhane and Van Dijk, 2020). Gunkel (2023) points out that the debate shows that the right questions have to be asked, and that some authors might be asking the wrong questions or formulating their questions in misleading ways. Nevertheless, the very fact that such a varied debate about the moral patiency of social robots exists is indicative of the social and conceptual disruptiveness of the technology itself. Much as social robots create conceptual disruption with regard to our

5 Yet another way to approach moral patiency of social robots is through the concept of derivative vulnerabilities proposed by Babushkina and Votsis (2021). Their idea is that an artificial agent may be seen as acquiring a derivative right to persist depending on the degree of pairing with the user. This may happen when a computer device merges with the cognition of the user to such extent that they form a hybrid personhood, creating vulnerabilities, and mutual dependency of the user and the artificial agent.

uncertainty of how to ontologically classify them, so too the debate about the moral patiency of social robots shows that there is uncertainty about whether the concept of moral patiency is even applicable here (Löhr, 2022), especially since most technologies do not prompt such discussion. Moreover, we could also question whether (if we do indeed apply such a concept to social robots) the very meaning of what it is to be a moral patient may change, and whether it could alter the ways in which we apply the concept to ourselves. Could it even alter the way in which we perceive ourselves as moral patients in the world (Sætra, 2022)?

3.4 Looking ahead

In this final section, we briefly zoom out and look to the future. While a lot of interesting research has been carried out, there are still many opportunities when it comes to the future of social robots and their potential role(s) in society. Gaps need to be filled in, theories need to be further developed, and more diverse perspectives need to be taken into consideration. We are excited about the future, but we also urge caution, and in this last section we highlight some of the directions we see the field heading. We also make some brief recommendations about especially promising areas of new research.

Notably, in the future, it is to be expected that social robots will be used in an even wider range of domains of human life. This has implications not just for their technical design (i.e. their physical architecture and cognitive design) but also for the sociotechnical systems that underpin the various further potential contexts for social robots, as well as the ecosystems in which they will be deployed. On the technical side, there is likely to be increased convergence between social robotics and other developments in AI, such as generative AI, i.e. forms of AI that can generate new content out of the data they have been trained on, such as the large language model technologies discussed earlier.[6]

6 Regarding technical developments in robotics more generally, an interesting example here is how the COVID-19 pandemic generated interest in the potential of urban robotics and automation to manage and police physical distancing and quarantine in China (Chen et al., 2020). For discussion of development in drones, driverless vehicles, and service robots, see Macrorie et al. (2019) and While et al. (2020).

A key ethical question in relation to the potential introduction of social robots into more and more contexts is whether there are some contexts/domains where it is more problematic to make use of social robots than in others, and where it is better to avoid introducing social robots. In general, many new ethical questions will be raised about how we should interact with these robots in various settings, along with distributions of responsibilities as the robots become equipped with more advanced capacities and capabilities, and new hybrid intelligence systems are born, bringing further implications for sociotechnical systems design, across cultures (and generations).

Such developments have further implications. There is no guarantee that our traditional ethical norms related to human-human interaction will always carry over naturally to the ethics of human-robot interaction in all domains where social robots might come to be utilized (Nyholm, 2021). We may need to extend or update our current ethical frameworks in order to be able to tackle the new ethical issues that arise within new forms of human-robot interaction. Moreover, in addition to building on and extending traditional ethical frameworks from Western philosophy, we also see an increasing need for engaging with non-Western perspectives. Excitingly, some discussions are already taking such perspectives into account, such as those surrounding moral character.

In particular, there is a question about how the increasing prevalence of robots in human social relations could impact human moral character. For example, Friedman (2022) has contributed to this discussion by taking an ubuntu approach to the topic. Ubuntu places emphasis upon the importance of interdependent human relations, and, specifically on having other-regarding traits or characteristics within the context of these interdependent relationships (such as by exhibiting a concern for human equality, reciprocity, or solidarity). Such relations are important because they help us become 'fully human'. The notion of becoming 'fully human' is important because in Ubuntu philosophy we are not only biologically human, but must strive to become better, more moral versions of ourselves, in order to become fully human. Therefore, being fully human means being particularly moral in character. If robots crowd out human relations, this is morally concerning because we cannot plausibly experience an interdependent relationship with a robot,

wherein other-regarding traits (such as human equality, reciprocity, or solidarity) are fully exhibited. Therefore, we cannot become 'fully human' i.e., better moral versions of ourselves, through relations with robots alone. Or so Friedman argues. This is concerning because should robot relations crowd out human relations, we would be interacting with human beings much less and, therefore, have less opportunity to develop our moral character in this way.

In addition to the Ubuntu approach, the dominant Western approach to robot ethics could also draw inspiration from Asian cultures, in particular, in South Korea, China, and Japan where many people place the perceptions of AI and robots at different points along the spectrum ranging from 'tool to partner' (Gal, 2020).

An interesting case, for example, is Japan, which has the highest percentage of industrial robots in the world (Kitano, 2015). The adoption of robots in Japan is partly based on Japanese Animism, 'Rinri' (in English, 'the Ethics'), in the context of Japanese modernization. Under this approach, the focus is on the harmonization of society, with each individual person forming a responsibility and accountability to that community. Within this culture, according to one interpretation, robots identify with their proprietor, and through such responsibility are just as accountable as their proprietor for the harmonization of Japanese society (Kitano, 2015). Conceptually, the Japanese approach could also be seen as a form of post-humanization — a distinct variant of posthumanism — which erases sharp human/non-human boundaries (Gladden, 2019: 8). In terms of social implications, under its Society 5.0 vision, Japan is promoting the integration of robots into society, and this is expected to contribute to society by presenting solutions to social problems, such as the labor shortages caused by the low birthrate and aging society, to enable every person to play a significant role by utilizing their own abilities (Japan Advisory Board on Artificial Intelligence and Human Society, 2017).

In general, how robots are received in society, whether they are accepted and how they are dealt with depends very much on cultural factors, which is why multicultural approaches to robots are important. Religions and other forms of worldviews also play an important role as cultural influences, as they shape value systems, understandings of nature and creation, as well as attitudes towards non-human entities,

and thus also affect attitudes towards technology. There are major differences in the attitudes of religions towards technology, especially between the monotheistic religions and non-monotheistic religions which are historically more open towards a diverse range of attitudes towards objects and technologies (Puzio, 2023).

One area where we see room for further expansion is the discussion surrounding our obligations to robots. Notably, and partly due to their potential for significant social and conceptual disruption, Bryson (2010) warns against developing any kinds of robots that we would have obligations towards. Indeed, Bryson argued that we should only design robots that can be used as tools, to the benefit of humans. Following her lead, Nyholm (2020) contends that we should avoid creating humanoid robots in particular, other than if there is some very clear and morally significant benefit associated with certain forms of humanoid robots, such as in therapy. This will help us avoid running into moral dilemmas about how we should and should not relate to and treat robots.

How we think about our obligations to robots and what this means for the development of social robots will prove to have a significant impact on society at large. As such, we want to make sure that the benefits of developing social robots that we have obligations to outweigh the risks and costs. If we do not, we might end up putting ourselves into moral situations that we are not capable of dealing with, or develop technologies that we lose control over (for further discussion, see Nyholm, 2022).

With this in mind, we think that further research needs to be done in creating and developing a more moderate approach. That is to say, we do not think society should limit research on social robotics in the way Bryson (2010) seems to suggest, but we also want to make sure we tread carefully, with awareness of potential dangers and social disruptions. Thus, we call on researchers to come up with more suggestions on how to develop social robotics research in a responsible yet forward-looking way. For instance, there could be more of an emphasis on developing warning systems for social robots, which alert people to the particular capabilities of each robot (Frank and Nyholm, 2017). This would enable people to understand how best to approach and treat the robot, without needing to wrestle (quite as much) with moral and relational issues.

Beyond more design-oriented solutions, however, in order to appreciate the ethical disruption social robots set upon us and identify meaningful ways forward, we need to foster transdisciplinary research. Only by doing this can we encompass and fruitfully blend cross-disciplinary perspectives on social robots from diverse fields of knowledge, such as philosophy, anthropology, social science, psychology, design, computer science, and robotics, as well as the future individual users who will be the most affected by the introduction of social robots in society.

Further listening

Readers who would like to learn more about the topics discussed in this chapter might be interested in listening to these episodes of the ESDiT podcast (https://anchor.fm/esdit):

Cindy Friedman on 'Social robots': https://anchor.fm/esdit/episodes/Cindy-Friedman-on-Social-Robots-e19jnjc

Sven Nyholm on 'A new control problem? Humanoid robots, artificial intelligence, and the value of control': https://anchor.fm/esdit/episodes/Sven-Nyholm-on-A-new-control-problem--Humanoid-robots--artificial-intelligence--and-the-value-of-control-e1thcu1

Dina Babushkina on 'Disruption, technology, and the question of (artificial) identity': https://anchor.fm/esdit/episodes/Dina-Babushkina-on-Disruption--technology-and-the-question-of-artificial-identity-e1jstvm

References

Asada, Minoru. 2019. 'Artificial pain may induce empathy, morality, and ethics in the conscious mind of robots', *Philosophies*, 4(3): 38, https://doi.org/10.3390/philosophies4030038

Babushkina, Dina. 2021a. 'Robots to Blame?', in *Culturally Sustainable Social Robotics: Proceedings of Robophilosophy Conference 2020*, ed. by Marco Nørskov, Johanna Seibt, and Oliver Santiago Quick (Amsterdam: IOS Press), 305–15, https://doi.org/10.3233/FAIA200927

——. 2021b. 'What does it mean for a robot to be respectful?', *Techné*, 26(1): 1–30, https://doi.org/10.5840/techne2022523158

——. 2022. 'Are we justified to attribute a mistake in diagnosis to an AI diagnostic system?', *AI & Ethics*, 3, https://doi.org/10.1007/s43681-022-00189-x

Babushkina, Dina, and Athanasios Votsis. 2021. 'Disruption, technology and the question of (Artificial) Identity', *AI & Ethics*, 2: 611–22, https://doi.org/10.1007/s43681-021-00110-y

Behdadi, Dorna, and Christian Munthe. 2020. 'A normative approach to artificial moral agency', *Minds and Machines*, 30: 195–218, https://doi.org/10.1007/s11023-020-09525-8

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. 'On the dangers of stochastic parrots: Can language models be too big? 🦜', *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–23, https://doi.org/10.1145/3442188.3445922

Birhane, Abeba, and Jelle van Dijk. 2020. 'Robot rights? Let's talk about human welfare instead', *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 207–13, https://doi.org/10.1145/3375627.3375855

Breazeal, Cynthia. 2003. 'Toward sociable robots', *Robotics and autonomous systems*, 42(3–4): 167–75, https://doi.org/10.1016/S0921-8890(02)00373-1

Bryson, Joanna. 2010. 'Robots should be slaves', in *Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues*, ed. by Yorick Wilks (Amsterdam: John Benjamins Publishing Company), 63–74, https://doi.org/10.1075/nlp.8.11bry

——. 2022. 'One day, AI will seem as human as anyone. What then?', *Wired*, https://www.wired.com/story/lamda-sentience-psychology-ethics-policy/

Burch, James. 2018. 'Beloved robot dogs honored in funeral ceremony', *National Geographic*, https://www.nationalgeographic.com/travel/article/in-japan--a-buddhist-funeral-service-for-robot-dogs

Čapek, Karel. 1928. *R.U.R. (Rossum's Universal Robots): A Play in Three Acts and an Epilogue* (London: Humphrey Milford; Oxford University Press)

Chalmers, David. 2022. 'Are large language models sentient?', *NYU Mind, Ethics, and Policy Program*, https://youtu.be/-BcuCmf00_Y

Chen, Bei, Simon Marvin, and Aidan While. 2020. 'Containing COVID-19 in China: AI and the robotic restructuring of future cities', *Dialogues in Human Geography*, 10: 238–41, https://doi.org/10.1177/2043820620934267

Carpinella, Colleen, Alisa Wyman, Michael Perez, and Steven Stroessner. 2017. 'The Robotic Social Attributes Scale (RoSAS): Development and validation', *HRI '17: Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 254–62, https://doi.org/10.1145/2909824.3020208

Cialdini, Robert. 2009. *Influence: Science and Practice* (Boston: Pearson Education)

Coeckelbergh, Mark. 2010a. 'Robot rights? Towards a social-relational justification of moral justification', *Ethics and Information Technology*, 12: 209–21, https://doi.org/10.1007/s10676-010-9235-5

——. 2010b. 'Moral appearances: Emotions, robots, and human morality', *Ethics and Information Technology*, 12: 235–41, https://doi.org/10.1007/s10676-010-9221-y

Damiano, Luisa, and Paul Damouchel. 2018. 'Anthropomorphism in human-robot co-evolution', *Frontiers in Psychology*, 9: 468, https://doi.org/10.3389/fpsyg.2018.00468

Danaher, John. 2019. 'The rise of the robots and the crisis of moral patiency', *AI & Society*, 34: 129–36, https://doi.org/10.1007/s00146-017-0773-9

——. 2020. 'Robot betrayal: A guide to the ethics of robotic deception', *Ethics & Information Technology*, 22(2): 117–28, https://doi.org/10.1007/s10676-019-09520-3

Danaher, John, and Neil McArthur. 2017. *Robot Sex and Consent: Social and Ethical Implications* (Cambridge: MIT Press)

Darling, Kate. 2016. 'Extending legal protection to social robots: The effects of anthropomorphism, empathy, and violent behavior towards robotic objects', in *Robot Law*, ed. by Ryan Calo, A. Michael Froomkin, and Ian Kerr (Cheltenham: Edward Elgar Publishing), 213–31, https://doi.org/10.4337/9781783476732.00017

De Graaf, Maartje. 2016. 'An ethical evaluation of human-robot relationships', *International Journal of Social Robotics*, 8: 589–98, https://doi.org/10.1007/s12369-016-0368-5

Devlin, Kate. 2018. *Turned On: Science, Sex, and Robots* (London: Bloomsbury)

Dignum, Virginia. 2019. *Responsible Artificial Intelligence* (Berlin: Springer), https://doi.org/10.1007/978-3-030-30371-6

Elgarf, Maha, Natalia Calvo-Barajas, Patricia Alves-Oliveira, Giulia Perugia, Ginevra Castellano, Chirstopher Peters, and Ana Paiva. 2022. '"And then what happens?" Promoting children's verbal creativity using a robot', *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 71–79, https://doi.org/10.1109/HRI53351.2022.9889408

Epley, Nicholas, Adam Waytz, and John Cacioppo. 2007. 'On seeing human: A three-factor theory of anthropomorphism', *Psychological Review*, 114(4): 864–86, https://psycnet.apa.org/doi/10.1037/0033-295X.114.4.864

Floridi, Luciano, and Jeff Sanders. 2004. 'On the morality of artificial agents', *Minds and Machines*, 14(3): 349–79, https://doi.org/10.1023/B:MIND.0000035461.63578.9d

Floridi, Luciano. 2013. *The Ethics of Information* (Oxford: Oxford University Press)

Fogg, Brian, and Clifford Nass. 1997. 'How users reciprocate to computers: An experiment that demonstrates behavior change', *CHI '97 Extended Abstracts on Human Factors in Computing Systems*, 331–32, https://doi.org/10.1145/1120212.1120419

Frank, Lily, and Sven Nyholm. 2017. 'Robot sex and consent: Is consent to sex between a robot and a human conceivable, possible, and desirable?', *Artificial Intelligence and Law*, 25(3): 305–23, https://doi.org/10.1007/s10506-017-9212-y

Frennert, Susanne, and Britt Östlund. 2014. 'Review: Seven matters of concern of social robots and older people', *International Journal of Social Robotics*, 6(2): 299–310, https://doi.org/10.1007/s12369-013-0225-8

Friedman, Cindy. 2022. 'Ethical concerns with replacing human relations with humanoid robots: An Ubuntu perspective', *AI & Ethics*, 3 https://doi.org/10.1007/s43681-022-00186-0

Friedman, Cindy, Sven Nyholm, and Lily Frank. 2022. 'Emotional embodiment in humanoid sex and love robots', in *Social Robotics and the Good Life: The Normative Side of Forming Emotional Bonds with Robots*, ed. by Janina Loh amd Wulf Loh (Bielefeld: transcript), 233–56

Gal, Danit. 2020. 'Perspectives and approaches in AI ethics: East Asia', in *Oxford Handbook of Ethics of Artificial Intelligence*, ed. by Markus Dubber, Frank Pasquale, Sunit Das (Oxford: Oxford University Press) 607–24, https://doi.org/10.1093/oxfordhb/9780190067397.013.39

Gladden, Mathew. 2019. 'Who will be the members of society 5.0? Towards an anthropology of technologically posthumanized future societies', *Social Science*, 8(5): 148–86, https://doi.org/10.3390/socsci8050148

Gogoshin, Dane Leigh. 2021. 'Robot responsibility and moral community', *Frontiers in Robotics and AI*, 8(768092), https://doi.org/10.3389/frobt.2021.768092

Gordon, John-Stewart, and Sven Nyholm. 2021. 'The ethics of artificial intelligence', *Internet Encyclopedia of Philosophy*, https://iep.utm.edu/ethics-of-artificial-intelligence/

Gouldner, Alvin. 1960. 'The norm of reciprocity: A preliminary statement', *American Sociological Review*, 25(2): 161–78, https://doi.org/10.2307/2092623

Graham, Elaine. 1999. 'Words made flesh: Women, embodiment and practical theology', *Feminist Theology*, 7(21): 109–21, https://doi.org/10.1177/096673509900002108

Guidi, Stefano, Latisha Boor, Laura van der Bij, Robin Foppen, Okke Rikmenspoel, and Giulia Perugia. 2023. 'Ambivalent stereotypes towards gendered robots: The (im)mutability of bias towards female and neutral robots', *Social Robotics: 14th International Conference, ICSR 2022*, Florence, Italy, December 13–16, 2022, Proceedings, Part II (Cham: Springer), 615–26, https://doi.org/10.1007/978-3-031-24670-8_54

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

——. 2023. *Person, Robot, Thing* (Cambridge: MIT Press)

Haraway, Donna. 1985. 'A cyborg manifesto', *Socialist Review*, 80: 65–108.

Heider, Fritz, and Marianne Simmel. 1944. 'An experimental study of apparent behavior', *The American Journal of Psychology*, 57(2), 243–59, https://doi.org/10.2307/1416950

Japan Advisory Board on Artificial Intelligence and Human Society. 2017. *Report on Artificial Intelligence and Human Society*, http://ai-elsi.org/wp-content/uploads/2017/05/JSAI-Ethical-Guidelines-1.pdf

Kahn, Peter. Hiroshi Ishiguro, Batya Friedman, Takayuki Kanda, Nathan G. Freier, Rachel L. Severson, and Jessica Miller. 2006. 'What is a human? Toward psychological benchmarks in the field of human–robot interaction', *Interaction Studies*, 8(3): 364–71, https://doi.org/10.1075/is.8.3.04kah

Kidd, Cory, and Cynthia Breazeal. 2008. 'Robots at home: Understanding long-term human-robot interaction', *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 3230–35, https://doi.org/10.1109/IROS.2008.4651113

Kitano, Naho. 2015. 'Animism, Rinri, modernization; The base of Japanese Robotics', School of Social Sciences, Waseda University, http://documents.mx/documents/kitano-animism-rinri-modernization-the-base-of-japanese-robots.html

Levy, David. 2008. *Love and Sex with Robots: The Evolution of Human-Robot Relationships* (New York: Harper Perennial), https://doi.org/10.1109/MTS.2008.930875

Loh, Janina. 2019. *Roboterethik: Eine Einführung* (Stuttgart: Suhrkamp)

——. 2022. 'Posthumanism and ethics', in *Palgrave Handbook of Critical Posthumanism*, ed. by Stefan Herbrechter, Ivan Callus, Manuela Rossini, Marija Grech, Megen de Bruin-Molé, and Christopher John Müller (London: Palgrave Macmillan), https://doi.org/10.1007/978-3-030-42681-1_34-1

Löhr, Guido. 2022. 'Linguistic interventions and the ethics of conceptual disruption', *Ethical Theory and Moral Practice*, 25: 835–49, https://doi.org/10.1007/s10677-022-10321-9

Lorenz, Tamara, Astrid Weiss, and Sandra Hirche. 2016. 'Synchrony and reciprocity: Key mechanisms for social companion robots in therapy and care', *International Journal of Social Robotics*, 8(1): 125–43, https://doi.org/10.1007/s12369-015-0325-8

Macrorie, Rachel, Simon Marvin, and Aidan While. 2021. 'Robotics and automation in the city: A research agenda', *Urban Geography*, 42(2): 197–217, https://doi.org/10.1080/02723638.2019.1698868

Metzinger, Thomas. 2013. 'Two principles for robot ethics' in *Robotik und Gesetzgebung*, ed. by Eric Hilgendorf and Jan-Philipp Günther (Baden-Baden: Nomos), 263–302, https://doi.org/10.5771/9783845242200-263

Mitchell, Gary, and Hugh O'Donnell. 2013. 'The therapeutic use of doll therapy in dementia', *British Journal of Nursing*, 22(6), 329–34, https://doi.org/10.12968/bjon.2013.22.6.329

Müller, Vincent. 2021. 'Is it time for robot rights? Moral status in artificial entities', *Ethics and Information Technology*, 23(3): 579–87, https://doi.org/10.1007/s10676-021-09596-w

Munn, Nick, and Dan Weijers. 2022. 'Corporate responsibility for the termination of digital friends', *AI & Society*, https://doi.org/10.1007/s00146-021-01276-z

Nass, Clifford, and Youngme Moon. 2000. 'Machines and mindlessness: Social responses to computers', *Journal of Social Issues*, 56: 81–103, https://doi.org/10.1111/0022-4537.00153

Nyholm, Sven. 2020. *Humans and Robots: Ethics, Agency, and Anthropomorphism* (London: Rowman & Littlefield)

——. 2021. 'The ethics of human-robot interaction and traditional moral theories', in *The Oxford Handbook of Digital Ethics*, ed. by Carissa Véliz (Oxford: Oxford University Press), https://doi.org/10.1093/oxfordhb/9780198857815.013.3

——. 2022. 'A new control problem? Humanoid robots, artificial intelligence, and the value of control', *AI & Ethics*, https://doi.org/10.1007/s43681-022-00231-y

——. 2023. *This is Technology Ethics: An Introduction* (New York: Wiley-Blackwell)

Nyholm, Sven, and Lily Frank. 2019. 'It loves me, it loves me not: Is it morally problematic to design sex robots that appear to "love" their owners?', *Techné*, 23: 402–24, https://doi.org/10.5840/techne2019122110

Paetzel-Prüsmann, Maike, Giulia Perugia, and Ginevra Castellano. 2021. 'The influence of robot personality on the development of uncanny feelings', *Computers in Human Behavior*, 120, 106756, https://doi.org/10.1016/j.chb.2021.106756

Perugia, Giulia, Stefano Guidi, Margherita Bicchi, and Oronzo Parlangeli. 2022. 'The shape of our bias: Perceived age and gender in the humanoid robots of the ABOT database', *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 110–19, https://doi.org/10.1109/HRI53351.2022.9889366

Perugia, Giulia, and Dominika Lisy. 2022. 'Robot's gendering trouble: A scoping review of gendering humanoid robots and its effects on HRI', arXiv preprint arXiv:2207.01130, https://doi.org/10.48550/arXiv.2207.01130

Puzio, Anna. 2022. 'Über-Menschen', in *Philosophische Auseinandersetzung mit der Anthropologie des Transhumanismus* (Bielefeld: transcript Verlag), https://doi.org/10.14361/9783839463055

Puzio, Anna. 2023.: 'Robot theology: On theological engagement with robotics and religious robots', in *Alexa, Wie Hast Du's Mit der Religion? Theologische Zugänge zu Technik und Künstlicher Intelligenz*, ed. by Anna Puzio, Nicole Kunkel, and Hendrik Klinge (Darmstadt: wbg academic), 95–114.

Rani, Anita. 2013. 'The Japanese men who prefer virtual girlfriends to sex', *BBC News Magazine*, http://www.bbc.com/news/magazine-24614830

Richardson, Kathleen. 2015. 'The asymmetric "relationship": Parallels between prostitution and the development of sex robots', *SIGCAS Computers & Society*, 45(3): 290–93, https://doi.org/10.1145/2874239.2874281

Russell, Stuart, and Peter Norvig. 2005. *Artificial Intelligence: A Modern Approach* (Hoboken: Prentice Hall)

Sætra, Henrik. 2022. 'Robotomorphy: Becoming our creations', *AI & Ethics*, 2(1): 5–13, https://doi.org/10.1007/s43681-021-00092-x

Sandoval, Eduardo Benítez, Jürgen Brandstetter, Mohammad Obaid, and Christoph Bartneck. 2016. 'Reciprocity in human-robot interaction: A quantitative approach through the prisoner's dilemma and the ultimatum game', *International Journal of Social Robotics*, 8(2): 303–17, https://doi.org/10.1007/s12369-015-0323-x

Seibt, Johanna. 2016. 'Integrative social robotics—a new method paradigm to solve the description problem and the regulation problem?', in *What Social Robots Can and Should Do. Proceedings of Robophilosophy 2016*, ed. by Johanna Seibt, Marco Nørskov, and Søren Schack Andersen (Amsterdam: IOS Press), 104–15, https://doi.org/10.3233/978-1-61499-708-5-104

Schwitzgebel, Eric, and Mara Garza. 2015. 'A defense of the rights of artificial intelligences', *Midwest Studies in Philosophy*, 39(1): 98–119, https://doi.org/10.1111/misp.12032

Sharkey, Noel. 2018. 'Mama Mia, it's Sophia: A show robot or dangerous platform to mislead?', *Forbes*, https://www.forbes.com/sites/noelsharkey/2018/11/17/mama-mia-its-sophia-a-show-robot-or-dangerous-platform-to-mislead/

Sharkey, Amanda, and Noel Sharkey. 2012. 'Granny and the robots: Ethical issues in robot care for the elderly', *Ethics and Information Technology*, 14: 27–40, https://doi.org/10.1007/s10676-010-9234-6

Sharkey, Amanda, and Noel Sharkey. 2021. 'We need to talk about deception in social robotics!', *Ethics and Information Technology*, 23: 309–16, https://doi.org/10.1007/s10676-010-9234-6

Sparrow, Robert, and Linda Sparrow. 2006. 'In the hands of machines? The future of aged care', *Minds and Machines*, 16: 141–61, https://doi.org/10.1007/s11023-006-9030-6

Sparrow, Robert. 2017. 'Robots, rape, and representation', *International Journal of Social Robotics*, 9(4): 465–77, https://doi.org/10.1007/s12369-017-0413-z

——. 2021. 'Virtue and vice in our relationships with robots: Is there an asymmetry and how might it be explained?', *International Journal of Social Robotics*, 13(1): 23–29, https://doi.org/10.1007/s12369-020-00631-2

Spatola, Nicolas, and Thierry Chaminade. 2022. 'Cognitive load increases anthropomorphism of humanoid robots. The automatic path of anthropomorphism', *International Journal of Human-Computer Studies*, 167: 102884, https://doi.org/10.1016/j.ijhcs.2022.102884

Strasser, Anna. 2022. 'Distributed responsibility in human-machine interactions', *AI & Ethics*, 2: 523–32, https://doi.org/10.1007/s43681-021-00109-5

Sullins III, John. P. 2006. 'When is a robot a moral agent?', *The International Review of Information Ethics* 6: 23–30, https://doi.org/10.29173/irie136

Thweatt, Jennifer. 2018. 'Cyborg-Christus: Transhumanismus und die Heiligkeit des Körpers', in *Designobjekt Mensch. Die Agenda des Transhumanismus auf dem Prüfstand*, ed. by Benedikt Göcke and Frank Meier-Hamidi (Freiburg: Herder), 363–76

Tickle-Degnen, Linda, and Robert Rosenthal. 1990. 'The nature of rapport and its nonverbal correlates', *Psychological Inquiry*, 1(4): 285–93, https://doi.org/10.1207/s15327965pli0104_1

Tummers-Heemels, Ans, Rens Brankaert, and Wijnand Ijsselsteijn. 2021. 'Between benevolent lies and harmful deception: Reflecting on ethical challenges in dementia care technology', *Annual Review of CyberTherapy and Telemedicine*, 19: 15–20.

Turkle, Sherry. 2011. *Alone Together: Why We Expect More from Technology and Less from Each Other* (New York: Basic Books)

——. 2020. 'A nascent robotics culture: New complicities for companionship', in *Machine Ethics and Robot Ethics*, ed. by Wendell Wallach and Peter Asaro (London: Routledge, 107–16), https://doi.org/10.4324/9781003074991

Véliz, Carissa. 2021. 'Moral zombies: Why algorithms are not moral agents', *AI & Society*, 36: 487–97, https://doi.org/10.1007/s00146-021-01189-x

Van Wynsberghe, Aimee. 2022. 'Social robots and the risks to reciprocity', *AI & Society*, 37: 479–85, https://doi.org/10.1007/s00146-021-01207-y

Wagner, Alan, and Ronald Arkin. 2011. 'Acting deceptively: Providing robots with the capacity for deception', *International Journal of Social Robotics*, 3(1): 5–26, https://doi.org/10.1007/s12369-010-0073-8

Wagner, Alan. 2016. 'Lies and deception: Robots that use falsehood as a social strategy', in *Robots That Talk and Listen: Technology and Social Impact*, ed. by Judith Markowitz (Berlin: de Gruyter), 203–25, https://doi.org/10.1515/9781614514404

Wareham, Christopher. 2021. 'Artificial intelligence and African conceptions of personhood', *Ethics and Information Technology*, 23(2): 127–36, https://doi.org/10.1007/s10676-020-09541-3

Wilks, Yorick. 2010. 'Introducing artificial companions', in *Close Engagements with Artificial Companions*, ed. by Yorick Wilks (Amsterdam: John Benjamins Publishing Co), 11–22. https://doi.org/10.1075/nlp.8

Zhao, Shanyang. 2006. 'Humanoid social robots as a medium of communication', *New Media & Society*, 8(3): 401–19, https://doi.org/10.1177/1461444806061951