12

Robots in Society

What is covered in this chapter:

- The influence of the media on human–robot interaction (HRI) research.
- Stereotypes of robots in the media.
- Positive and negative visions of HRI.
- Ethical considerations when designing an HRI study.
- Ethical issues of robots that fulfill a user's emotional needs.
- The dilemmas associated with behavior toward robots (e.g., robots' right to be treated in a moral way).
- The issue of job losses as a result of the increasing number of robots in the workforce.

The discussion of robots in society often brings up questions about how we envision robots in the present and future and the social and ethical consequences of using robots in different tasks and contexts. Researchers, the media, and members of the public argue over how robots will affect our perceptions of and interactions with other humans, what the consequences of new robotic technologies will be for labor distribution and relations, and what should be considered socially and ethically appropriate uses of robots. This kind of exploration is crucial to the field of human–robot interaction (HRI) because understanding the societal meaning, significance, and consequences of HRI research will ensure that new robotic technologies fit our common social values and goals. To understand how robots might fit into society, we take a broad view of HRI through the lens of culture and the narratives, values, and practices that provide the context and tools with which people make sense of the world around them and the robots that will be coming to share it.

In this chapter, we look at robots in fiction and film (Section 12.1), two aspects of popular culture that have had particularly strong impacts on how we imagine robotic technology in society. In Section 12.2, we consider ethical concerns about the introduction and use of robots in society to reflect on how our values and priorities should be taken into account while shaping the human–robot interactions of the future. In recent years, there has also been increased focus on considering diversity and inclusion in HRI—in terms of considering more inclusive robot design practices, working with a broader

array of demographic groups in user studies, and considering the potential effects of robot deployment and use with diverse stakeholders in various use cases.

12.1 Robots in popular media

What movies have become popular with audiences or critics recently? Is there a TV series that went viral or an episode that everyone is talking about? Did any of those contain robots, by chance? If so, how were these machines portrayed? Looking into the literature and other media, it becomes clear that robots have always been a "hot topic" for sci-fi writers and avid consumers.

Historically, stories of artificial human beings, such as the Golem in Jewish folklore, have been around for hundreds of years. Karel Čapek was the first author to use the word *robot*, which was featured in his theater play *R.U.R.—Rossum's Universal Robots* that premiered in 1921 (see Figure 12.1). In it, robots take over the world and kill almost all humans. Two robots do, however, start to exhibit emotions for each other, and the last remaining human considers them to be the new Adam and Eve. Isaac Asimov, in turn, coined the term *robotics* as a field of study, as well as the yet-to-be-realized domain of robopsychology, which could be seen as having some overlap with HRI. The notion of robots that befriend humans and aid society is a central focus of the post-WW2 narrative of Osamu Tezuka's Astro Boy series, centering on a robotic boy who lives with a family, has a heart, and works to help his human friends. Some robotics projects, such as the HRP-2 humanoid, bring inspiration from fictional narratives to life—in this case, the robot's functions of aiding people in construction work, moving objects, and other physical tasks are not only inspired by the manga series Patlabor, but its appearance is designed by Yutaka Izubuchi, the mechanical animation designer for the manga and anime series Kaneko et al. (2004).

Figure 12.1 A scene from Čapek's 1921 play *R.U.R.* shows robots rebelling against their human masters.

Now think back to when you first heard about robots. This first encounter with a robot was likely an on-screen encounter. Computer graphics can nowadays visualize almost anything; hence, depictions of robots in movies can be quite fantastical. For example, movies have depicted robots that use antigravity to float around. In reality, there is little use for such robot hardware features. Robots have been portrayed in all types of artistic expressions, such as books, movies, plays, and computer games. Such media portrayals form our perceptions and understanding of robots and can thus bias our views, particularly because these are the only experiences most people have with robots. We are at an interesting point in time where, on the one hand, more and more robots are about to enter our everyday lives, but on the other hand, almost all our public knowledge about robots stems from the media. This gap between the expectations fueled by science fiction and the actual abilities of robots often leads to disappointment when people interact with robots. This is why it is important to look at how robots are portrayed in popular media and to take such portrayals into account when we are designing robots for and presenting them to the public.

As a disclaimer, we have to acknowledge that it was not possible for us to consider every robot mentioned in every book, film, computer game, newspaper article, or play. Still, some valid conclusions might be drawn from the more or less classic examples that will be reiterated in the following discussion.

12.1.1 Robots want to be humans

In many narratives, robots are portrayed as wanting to be like humans, despite their actual superiority to humans, for instance, in terms of strength and computational power. To illustrate, the very desire to become human is the central storyline for Isaac Asimov's *Bicentennial Man*, in which a robot named Andrew Martin is following a lifelong plan to become recognized as a human (Asimov, 1976). The book was used as the basis for a movie of the same name, released in 1999. Besides becoming physically more humanlike, Andrew Martin also fights a legal battle to gain full legal status. He is even prepared to accept mortality to gain full legal status. Other robots, such as the replicant Rachael in the movie *Blade Runner* based on the book by Philip K. Dick, are not even aware of the fact that they are robots (Dick, 2007). The same holds true for some of the humanlike Cylons in the 2004 TV series *Battlestar Galactica*.

On the contrary, a prime example of a robotic character that is aware of its robotic nature is Mr. Data from the TV series *Star Trek: The Next Generation*. Mr. Data is stronger than humans; has more computational power than humans; and does not need sleep, nutrition, or oxygen. Still, this character is set up to have the desire to become more humanlike. The key aspect, however, that actually distinguishes Mr. Data from humans is his lack of emotion. Similarly, in Steven Spielberg's movie *A.I.* (based on Brian Aldiss's short story *Super-Toys Last All Summer Long*), robots also lack emotions (Aldiss, 2001), which prompts Professor Allen Hobby to build the protagonist robot David with the ability to love. Likewise, sci-fi authors have considered emotions to be a feature that all robots would lack. However, in reality, several computational systems to mirror emotions have already been successfully implemented. The computer programs implementing the so-called *OCC model of emotions* (Ortony et al., 1988) are prime examples. Equipping robots with emotions in the attempt to make them human is therefore an archetypal storyline.

A more subtle variation of this narrative concerns the inclusion of a control or setting for honesty and humor, as depicted in the robots from the movie *Interstellar*. The following dialogue between Cooper, the captain of a spaceship, and the TARS robot emerges:

COOPER: Hey, TARS, what's your honesty parameter?

TARS: Ninety percent.

COOPER: Ninety percent?

TARS: Absolute honesty isn't always the most diplomatic nor the safest form of communication with emotional beings.

COOPER: Okay, ninety percent it is.

Although robots might not have emotions themselves, they will be required to interact with humans who do have emotions, and hence it will be necessary for them to process emotions and even adjust their rational behavior accordingly.

The aforementioned classic examples taken from contemporary film are only the tip of the iceberg, but they illustrate humans' steady desire to compare themselves to superhuman entities. A hundred years ago, however, there were already machines that were more powerful than humans, although their power was physical and not mental. These days, we can see the major progress in the area of artificial intelligence (AI). On May 11, 1997, the IBM computer "Deep Blue" won the first chess match against the world champion at the time. In 2011, the IBM computer Watson won as a contestant in the quiz game show *Jeopardy*. In 2017, Google's DeepMind AlphaGo defeated the world's number-one Go player, Ke Jie. In light of this progress, it is easy to imagine how robots in the future might be both strong and intelligent, leaving humans in an inferior position. At the same time, computers and robots are successful in limited task domains, so humans may have the advantage through their ability to adapt and generalize to different tasks and contexts. Fictional narratives let us explore the consequences of these and other possibilities from the safety of our couches.

12.1.2 Robots as a threat to humanity

Another archetypal storyline that continuously reappears in fiction is that of a robotic uprising. In short, humanity builds intelligent and strong robots. The robots decide to take over the world and enslave or kill all humans in order to secure resources for themselves (Barrat, 2015). Karel Čapek's original play, mentioned earlier, already introduced this narrative. Going back to the example of Mr. Data, he has a brother named Lore who possesses an emotion chip. Lore follows the path of not wanting to be like a human but instead wanting to enslave humanity. Other popular examples are *The Terminator* (Cameron, 1984) (see Figure 12.2), the Cylons in *Battlestar Galactica*, the Machines portrayed in the movie *The Matrix*, and the robots portrayed in the 2004 movie *iRobot* (which is based on the book by the same name authored by Isaac Asimov (Asimov, 1991)). Asimov coined the term *Frankenstein complex* to describe the notion that robots would take over the world.

Figure 12.2 The Terminator. (Source: Dick Thomas Johnson)

This archetype builds on two assumptions. First, robots resemble humans. The robots depicted in the aforementioned movies and shows have been designed to look, think, and act like their creators. However, they exceed their creators in intelligence and power. Second, once they interact with the now "inferior" human species, robots dehumanize their subordinates, a theme

familiar in examples from human history as well. Many colonial powers declared indigenous populations as nonhumans in an attempt to vindicate the atrocities committed toward them. Accordingly, because robots resemble humans, they will also enslave and kill humans. However, this rationale is overly simplistic. The issue of a perceived threat to distinctiveness is also addressed in the psychological literature (Ferrari et al., 2016). If you want to learn more about the psychology of feeling threatened by robots, then consider reading the work of Złotowski et al. (2017).

The movie *Ex Machina* (Garland, 2014) combines the archetypes just discussed (robots pretending to be human and robots taking over) with an interesting twist. Human protagonist Caleb falls in love with robot protagonist Ava, who, unbeknownst to him, has been designed to be his dream woman. The two grow an apparent emotional attachment, and Ava begs Caleb to help her escape from the lab where she is kept. However, after Caleb does so, she reveals that she manipulated him in order to escape, then leaves him trapped in the same lab with no possibility of escape. Although the movie adheres to the archetype that emotions displayed by robots are not real and that robots are hostile toward humans, it gives both paradigms a twist because Ava's behavior originates from her (very human) outrage at being exploited and kept prisoner.

12.1.3 Superior robots being good

Several science-fiction authors have already proposed future scenarios in which superior robots quietly influence human society. In Isaac Asimov's *Prelude to Foundation*, he describes a robotic first minister, Eto Demerzel (a.k.a. R. Daneel Olivaw), who keeps the empire on the right track (Asimov, 1988). Interestingly, he hides his robotic nature. He is a very humanlike robot in appearance but resorts to various strategies to blend in. For example, he eats food, despite the fact that he cannot digest it. He collects it in a pouch that can be emptied later. Here we have a scenario in which a superior being works to help human society behind the scenes.

The notion of robots being evil and humans being good is most persistent in Western culture. Robots are extremely popular in the Japanese media, and there we can observe a different relationship between humans and robots: robots, such as Astro Boy and Doraemon, are good-natured characters that help humans in their daily lives. This more positive spin on the social uses and consequences of robots is often seen as being partially responsible for the large number of personal and home robots being developed in Japan and their perceived higher acceptance there than in Western societies.

12.1.4 Similarity between humans and robots

The common theme between all these science-fiction narratives concerns the fact that all of them explore the question of to what degree humans and robots are alike. From a conceptual point of view, robots are typically portrayed by emphasizing either their similarities to humans or lack thereof in terms

Table 12.1 Topics of HRI in media portrayals.

		Mind	
		Similar	Different
Body	Similar	Type I	Type II
	Different	Type III	Type IV

of the robot's body and mind (see Table 12.1). Dixon supports this view by stating that artists explore the deep-seated fears and fascinations associated with machine embodiment in relation to two distinct themes: the humanization of machines and the dehumanization of humans (Dixon, 2004; Haslam, 2006).

These four types of topics can, of course, be mixed. If we take the example of Mr. Data, at the superficial level, he looks very much like a human, which sets our expectations accordingly (Type II). It then appears dramatic and surprising if Mr. Data can enter the vacuum of space without being damaged. In the movie *Prometheus*, the android David is wearing a space suit when walking on an alien planet. Wearing this suit does not serve a functional purpose because David does not require air. The dialogue unfolds as follows:

CHARLIE HOLLOWAY: David, why are you wearing a suit, man?

DAVID: I beg your pardon?

CHARLIE HOLLOWAY: You don't breathe, remember? So, why wear the suit?

DAVID: I was designed like this, because you people are more comfortable interacting with your own kind. If I didn't wear the suit, it would defeat the purpose.

Again, the human embodiment sets our expectations, and when a difference from humans is displayed, it surprises the audience. Godfried-Willem Raes takes a different approach with his robot orchestra. He emphasizes the equality of robots and humans in his theatrical performances (Type I). He argues:

> If these robots conceal nothing, it is fairly self-evident that when their functioning is made dependent on human input and interaction, this human input is also provided naked. The naked human in confrontation with the naked machine reveals the simple fact that humans, too, are actually machines, albeit fundamentally more refined and efficient machines than our musical robots.

An example of Type III could be Johnny Five from the 1986 movie *Short Circuit*. Although Johnny Five has a distinctively robotic body, he does express human emotions, which suggests that his mind is similar to that of humans.

12.1.5 Narratives of robotic science

Ben Goldacre has pointed out how the media promotes the public misunderstanding of science (Goldacre, 2008). Two narratives that the media frequently uses are science-scare stories and wacky science stories.

The performance of autonomous vehicles, which can also be considered a form of HRI, is currently the target of immense scrutiny. The crash statistics provided by Tesla, Waymo, and others indicate that they are performing better than humans. Tesla, for example,[1] showed that driving using the vehicle's autopilot feature reduces the probability of crashes dramatically.

One question that almost all reporters ask when interviewing HRI researchers focuses on when robots will actually take over the world. The goal, then, is to write a story that scares the public and hence attracts attention. A story entitled "Robots Are Harmless and Almost Useless" is very unlikely to get published. But that is what most HRI projects come down to at this point in time. The question of whether and when robots will take over the world addresses our inner fears and fascinations involving interacting with robots. It reflects the ambivalent attitudes we might hold toward robots—on the one hand, robots are viewed as an asset and support in everyday life, but on the other hand, the prospect of a hybrid society appears threatening to many because they fear losing their jobs or finding their privacy breached, for example.

We may ask ourselves why the ambivalent portrayals of robots are so persistent in the media. The most obvious answer is that many storylines call for a "conflict" to make a storyline more interesting. A (science-) fictional world in which everybody is happily living ever after is unlikely to capture the attention of a broader audience. Pitching evil robots against good humans not only serves the purpose of creating ambivalence but also triggers an "in-group" effect (Ferrari et al., 2016; Złotowski et al., 2017). Humans often show the tendency to defend our species against "out-group" robots. This division can then be challenged by introducing robots that are indistinguishable from humans, such as in the TV shows *Battlestar Galactica* and *Westworld*. This creates great uncertainty, which in turn creates tension. Notable exceptions from the gloomy visions in the media are the TV series *Futurama* by Matt Groening and the movie *Robot and Frank* by Jake Schreier, both of which depict a vision of the future in which humans and robots live peacefully side by side, even becoming friends. In the movie *Her*, the protagonist Theodore, played by Joaquin Phoenix, falls in love with his AI mobile phone Samantha (Jonze, 2013).

On the other hand, media representations of robot technologies can be biased in the sense that they fit the wacky science narrative. This narrative resonates with pop science, is less prevalent, and serves to entertain rather than to report scientific progress (Berghuis, 2017).

Because the interest in all technologies that feature AI is still growing, many HRI researchers are invited for interviews. This offers a great opportunity for them to showcase their work, but at the same time, media coverage also carries considerable risk. To illustrate, a reporter might intend to write a scare story or even a wacky science story, without always giving that goal away. In light of the extensive media attention that HRI researchers commonly get, it

[1] See www.tesla.com/VehicleSafetyReport

might be advisable to participate in media training sessions before engaging with journalists. Such training sessions are offered at many universities and research institutes, and taking part in such training can minimize misrepresentations and detrimental outcomes of encounters with journalists who want to cover social robots and AI. As a general guideline for talking to the media, it appears advisable to stick to the research that was actually performed and avoid engaging in wild speculations about topics that were not covered in the study at hand. It might be helpful to clarify before an actual interview which questions will be asked and to request to view a manuscript draft prior to publication. Thereby, misunderstandings or misrepresentations of the science involved can be corrected prior to publication.

HRI researchers cannot shy away from representations of robots in the media, fictional or otherwise, and the elicitation of associated hopes and fears that create ambivalence toward robots (Stapels and Eyssel, 2022). In actual HRI research, we invite people to engage with robots, and every single person who interacts with a robot does so with certain attitudes, ambivalence, or hopes and expectations of what the robot can and cannot do. Many of these expectations are grounded in science fiction and potentially biased reports in the media rather than the annals of scientific research.

12.2 Ethics in HRI

Is it okay to develop and sell a sex robot, which is always willing to do what you want and will stay forever young and fit? Would you have your parents be taken care of by a carebot instead of a human nurse?

Roboticists and philosophers alike have long been concerned with such ethical issues in robotics, coining a shared domain of scholarship called *roboethics*. More recently, a group of HRI scholars formulated five ethical rules, which they call their Principles of Robotics, to raise broader awareness about the role of ethics in HRI.[2] Ethical rules have also been a subject of discussion in popular literature, particularly the well-known "Three Laws of Robotics" (see the accompanying text box). Moreover, work by Fosch-Villaronga et al. (2020) outlines the ethical, legal, and social (ELS) implications that emerge when reflecting on HRI. A recent overview by Wullenkord and Eyssel (2020) outlines the various overarching challenges associated with social robots and HRI in a diverse set of contexts.

Figure 12.3 Isaac Asimov (January 2, 1920–April 6, 1992).

Isaac Asimov (January 2, 1920–April 6, 1992; see Figure 12.3) proposed three rules of robotics that would safeguard humanity from malevolent robots:

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.

[2] See https://doi.org/10.1080/09540091.2016.1271400

2. A robot must obey the orders given to it by human beings except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.

Although Asimov's work is highly visible in the public media, it has been criticized by philosophers, and it is clear even from the stories that the three rules are not a practical guide to satisfying ethical requirements for designing robots. Asimov eventually added a zeroth law:

0. A robot may not harm humanity or, by inaction, allow humanity to come to harm.

This clearly marks the relevance of debating issues such as the ubiquitous deployment of robots in future society; their use in home and care contexts; the implications of developing autonomous weapons systems and autonomous cars; or, giving it a seemingly positive touch, the development of robots for attachment, love, or sex.

These days, many robotics research projects envision robots as conducting acts on behalf of humans, such as killing others; doing "dull, dirty, and dangerous" tasks; or serving to fulfill humans' need for psychological closeness and sexuality. Some of these projects are even funded by government agencies. At the same time, there are clear counter-movements, such as the Campaign Against Killer Robots.[3] As responsible researchers, we have to consider the ethical implications of what we envision and the steps we take to approach these visions of the future (Sparrow, 2011). In the following subsections, we discuss some of the common topics of ethical concern in HRI research.

12.2.1 Robots in research

As a student beginning to get hands-on experience with empirical research in HRI, you might plan to conduct a study with a robot that acts seemingly autonomously. Even here, ethics has to be considered because you might choose to deceive your participants by controlling your robot using the Wizard-of-Oz approach. You thereby make the participants believe that the robot has certain functions, whereas in reality, you control the robot's behavior in the background. The problem with this approach is that the deception concerning the robot's skills raises and biases users' hopes and expectations about the robot's abilities. This may manipulate them into thinking that robotic technology is more advanced than it actually is (Riek, 2012).

Another critical example to consider might be the use of robots as persuasive communicators within your research project. Previous research on persuasive technology has shown that robots can be used to manipulate people into changing not only their attitudes but also their behaviors (Brandstetter

[3] See www.stopkillerrobots.org/

et al., 2017). Examples of behaviors that have been successfully influenced include health-related habits, such as exercising or maintaining a healthy diet (Kiesler et al., 2008). Even if it might benefit people to change their health-related habits, such as smoking less and exercising more, instrumentalizing social robots for this purpose poses ethical concerns if they exploit the social bond with the user and influence the user without the user's explicit consent and conscious knowledge or understanding of how he or she is being influenced. The notion of robot deception and manipulation is not easy to disentangle because these constructs remain ill-defined and distinct but related. Moreover, generally speaking, deception is a characteristic that marks empirical experimental research with humans and robots alike, in order not to reveal the true nature of the research questions at hand. This overlaps with the deceptive nature of robots and their capabilities—which may lead naive users to believe that robots indeed possess intentions, emotions, a mind, or other essentially human qualities (see Chapter 8).

12.2.2 Robots to fulfill emotional needs

Robotic care

Imagine your grandmother has been given a robot companion by a group of researchers. They tell her that this new technological friend will stay with her in her home for the next three weeks. She interacts with the robot every day for these three weeks, and over time, she becomes quite attached to it. The robot invites her to do activities like memory games on a regular basis. It asks her how she is doing and whether she slept well; it keeps her company, and it never argues with her. She is delighted with her new companion, and life is good. That is, until the researchers come back and ask her to complete some questionnaires before packing up the robot and taking it away. The dull routine of the elderly care center creeps back, and she feels even more lonely than before.

This brief scenario gives a glimpse of the psychological experience of getting attached—not only to people but also to objects like robots. HRI researchers have shown how easily people grow attached to a robot, even when it only briefly enters their everyday lives (Šabanović et al., 2014; Forlizzi and DiSalvo, 2006; Chang and Šabanović, 2015; Kidd and Breazeal, 2008). The emotional and social consequences of withdrawing this source of attention and "artificial affection" clearly need to be considered when running case studies with a social robot that has to be returned at the end of the study.

Relatedly, Steil et al. (2019) have proposed an ethical perspective reflecting the challenges associated with the use of robots in medical domains, which usually involve vulnerable populations like children, the elderly, or persons with cognitive or physical impairments.

Other studies, however, have demonstrated the beneficial effect of deploying small-scale robots, such as the therapeutic robot Paro (Wada and Shibata, 2007; Shibata, 2012) or the robot dog Aibo (Broekens et al., 2009). These

robots are not able to do any tedious manual labor, but they can provide companionship. Given the high workload that caretakers are burdened with, any relief, even small, is likely welcomed.

Manzeschke (2019) reflects on ethics in care contexts, with a particular focus on taking into account the different types of human–robot relationships. For instance, the robot is viewed as a mere tool, the robot is deemed a tool with social capabilities, or the robot is interpreted as a relationship partner. Above and beyond, Sparrow and Sparrow (2006) offer an interesting perspective on robotic care that has become a classic in the literature. They argue that even when a robotic caregiver can be developed that is capable of providing superb emotional and physical care, it would still be unethical to outsource care to machines. The reason for this is that a relationship can only be meaningful when it is between two entities that are capable of experiencing reciprocal affect and concern; an imitation of caring, however perfect, should never substitute the real product. This kind of relationship may also be detrimental to the value of upholding a person's dignity. This brings us to the ethics of developing a deeper emotional attachment to a robot (Law et al., 2022).

Emotional attachment to robots

Affection toward robots can go deeper and beyond the care setting. Humans may start to favor robot companions over humans. Imagine a social robot that can truly mimic friendship and emotional support, such as the android Klara in Kazuo Ishiguro's novel *Klara and the Sun*. This "ideal robotic friend" comes with all the perks of a human friend, never complains, and learns never to annoy its owner. Slowly, people could come to prefer these robotic companions over their human peers, who would not be able to measure up to the high standards that robotic friends provide. Would such a future be desirable? What would be the broader societal consequences of supporting the development of human–robot relationships?

Even though users may project all kinds of human traits onto a robot, the robot is not able to experience those traits in the same way humans do, and therefore, the authenticity of the expression can be doubted (Turkle, 2017). Still, robots are sometimes specifically designed to express social cues to deliberately facilitate bonding with them. The authenticity of feelings is normally important in human–human interaction, and we do not know how humans will react to robots that express themselves based on calculations rather than the sensation of emotions.

Going beyond human–robot friendship, there are individuals who feel closeness and intimacy toward robots. The broader question is whether promoting human–robot emotional bonds is desirable (Borenstein and Arkin, 2019). After all, we have to realize that the emotional relationship between humans and robots might be asymmetrical. Humans might nevertheless be quite satisfied with the robot exhibiting sympathetic responses, whether the robot has a humanlike sensation of attachment or not.

Ethical implications of persuasion through robots

Language develops dynamically, and every participant in discourse influences its development simply through its usage. New words appear (e.g., "to google"), others change their meaning (e.g., *gay*), and yet other words fall out of usage altogether. We can use Siri, Cortana, or Bixby to control our phones, homes, or shopping tours.

Familiarity alone will influence our attitudes toward concepts, political ideas, and products; this is called the *mere exposure effect* (Zajonc, 1968). The more often people hear a word, the more positive their attitude toward this word becomes. One day, it will make a great difference if your smart-shopping robot proposes to purchase "Pepsi" compared to offering a "Coca-Cola." The question really is who gets to decide what words our artificial counterparts use.

Robots have the ability to synchronize their vocabulary through the internet in seconds. Even the mass media cannot compete with this level of consistent usage of selected words (Brandstetter et al., 2017). Because of its ability to communicate in humanlike ways, a robot can be a convincing, persuasive communicator.

This comes with negative implications, though: without us even noticing, computers and robots can influence what words we use and how we feel about them. This can and probably is happening already, and we need to develop media and language competency to be able to withstand attempts to influence our views. With the ever more personalized and intimate relationships that we form with technologies, we are increasingly vulnerable. We probably already spend more time with our phones than with our partners and friends.

Furthermore, to our knowledge, there are no regulations or policies in place at this point in time to supervise how large information technology companies, such as Google, Amazon, or Facebook, influence the usage of language, although there is concern about "fake news" and the difficulty of telling fact from fiction in online contexts. It might also be a better approach to regulate the development of our language only to the degree that it should be left to its natural flow of change. With powerful tools at our fingertips, we need to ensure that no company or government can influence our language without our consent and that the robots we design do not become just one additional persuasive and misleading technology.

Generalizing abusive behavior toward robots

Being recognized as a social interaction partner comes with a downside: not all social behaviors aimed at you are positive. In a few field experiments with autonomous robots that were left unsupervised in public spaces, people were observed attempting to intimidate and bully robots (Brscić et al., 2015; Salvini et al., 2010). It is noteworthy that the type of aggression that people displayed seemed to resemble human–human abuse, such as kicking, slapping, insulting, and refusing to move out of the way after the robot politely asked. Abuse that would be more meaningful for machines, such as unplugging them or cutting their wires, was not observed.

Robots normally do not experience any pain or humiliation, hence, the human actually faces greater danger than the robot when, for example, slapping the robot because the human might hurt his or her hand. But there are more issues to consider than just the bully's bodily integrity. It has been argued that bullying a robot is a moral offense—even though nobody gets hurt, responding with violence is still considered wrong and should therefore be discouraged (Whitby, 2008). In addition, scholars have argued that if this behavior is perceived as acceptable, it might generalize to other social agents, such as animals and humans (Whitby, 2008; De Angeli, 2009). This transfer of negative behavior from a humanlike agent to actual humans is argued to also happen in other domains, such as violent computer games (Sparrow, 2017; Darling, 2012), and has been a topic of discussion for quite a while. Further research on this topic is still needed.

A related issue is that interactions with a robot may raise expectations regarding the behavior of other humans. This has been argued to be particularly dangerous in the domain of sex. A robot could easily be designed to seem to desire intercourse at any time and to readily and fully comply with any wishes of the user without having any desires or demands of its own. This could change what people consider normal or appropriate behavior from an intimate partner.

This issue becomes even more problematic if the robot is specifically designed for sexual behaviors that would be considered wrong if it involved human partners. For example, child-shaped sex robots could be designed to fit the desires of pedophiles, or sex robots could be programmed to explicitly not consent to or even struggle against sex in order for users to play out their rape fantasies. These robot behavior designs have been deemed ethically inappropriate by some scholars (for a philosophical justification, see Sparrow, 2017). Others, like David Levy and Hooman Samani, have set out to suggest (even back in the early 2000s) that love and sex with a robot would be a contemporary reality. We are still not there yet. Döring and Poeschl (2019) analyzed fictional and nonfictional media representations of intimacy between humans and robots. Regarding virtual agents, psychologist Mayu Koike has looked into the role of anthropomorphism in developing social, even romantic relationships with virtual characters (Koike et al., 2022; Koike and Loughnan, 2021). Virtual agents—even life-size versions—are available as companions, communication partners or romantic partners, using the Gatebox device.[4] Despite existing controversy, Bendel (2021) points out contexts in which love dolls and sex robots could eventually be useful while at the same time discussing the ethical issues associated with their use. Despite the growing interest in understanding the underpinnings of positive, close, and even intimate social relationships between humans and novel technologies, it is clear that further research indeed is needed to better understand the psychological underpinnings and consequences of intimate HRI (Borenstein and Arkin, 2019).

[4] See www.gatebox.ai/

12.2.3 Robots in the workplace

A repeatedly expressed worry is that "robots will replace me in the job market." Since the Industrial Revolution, humans have been replacing manual labor with machines, and the recent deployment of robots is no exception. Robots help us to improve our productivity and thereby help to increase our standard of living. Although robots do replace certain jobs, they also create many new jobs, in particular for highly trained professionals. The challenge that society is facing is that the people replaced by robots need to find new jobs, which might require them to undertake additional training or studies. This may be problematic or even impossible for some, for example, due to financial or intellectual constraints.

In many cases, the acceptance of robots in various workplaces will likely depend on their specific roles and how they are integrated into the workforce. Reich-Stiebert and Eyssel (2015) showed that robots are preferred as assistants in the classroom but not as the main teachers. Teachers also voiced concern about the usage and maintenance of the robots, being particularly fearful that the robots would take their resources in terms of time and attention. Interestingly, primary school teachers were particularly reluctant to have robots in schools, maybe because in their view, young students are particularly vulnerable. An analysis of the predictors of such rather negative attitudes and behavioral inclinations toward educational robots revealed that technology commitment was the key predictor of positive attitudes. That is, those teachers who were open to working with novel technologies in general felt more positive about robots and the future use of them in their classrooms. Another field in which people are concerned about the application of robots is assistive robots designed for use in the home (Reich-Stiebert and Eyssel, 2015, 2013). Again, technology commitment was found to predict people's reluctance to accept robots in their lives.

12.2.4 Ambivalent attitudes toward robots

Haegele (2016) claims that more and more robots will be sold on the market in upcoming years. Their acceptance into society, however, will remain a challenge, and further research on technology-related attitudes and how to change them is necessary to increase society's acceptance of robots. This is particularly relevant in light of the current reconceptualization of attitudes toward robots. That is, research by Stapels and Eyssel (2022, 2021) has shown that attitudes toward robots are not—as suggested by a meta-analysis by Naneva et al. (2020)—neutral or even mildly positive. Indeed, whereas the notion of ambivalent attitudes has been widely studied in social psychology, it has not been widely applied to social robots yet (Stapels and Eyssel, 2022, 2021). However, this is highly relevant because it is plausible that allegedly neutral attitudes toward robots are actually ambivalent. What do we mean by *ambivalence*? This refers to the simultaneous evaluation of the same attitude object in both positive and negative terms. From this, a person might

experience inner conflict, which, too, comes with distinct social and cognitive consequences (see van Harreveld et al., 2015). Research by Stapels and Eyssel (2022, 2021) was the first to demonstrate robot-related ambivalence, and further data are needed that use proper attitude measurements that include ambivalence so that the state of people's true attitudes toward robots can be explored. People's ambivalence toward robots may also shift to more positive or negative perceptions based on the context of the robot's use, so more testing in specific task and use contexts is important for understanding people's preferences about the deployment of robots in their everyday environments.

12.2.5 A more diverse and inclusive HRI

A number of researchers have joined forces to emphasize the multifaceted notion of diversity and its value for HRI researchers, their work, and the community at large. Diversity can be looked at from various angles, taking into account researcher characteristics (e.g., age, gender, geographic distribution), demographics or other features of research participants (i.e., belonging to a social minority, being part of a vulnerable group, socioeconomic status, etc.) under study, and how the design of the robot might affect diverse stakeholders or embody particular social and cultural stereotypes. Research that takes a human-centered perspective will take into account the first two aspects, whereas robot developers also need to be mindful of how they frame and design their robots, their appearance, and other robot characteristics. This, too, is relevant because none of the people involved in a robot development cycle are free from bias, and implicit as well as explicit biases may have an impact on design choices.

Several recent overviews of HRI research suggest that the field needs to become more inclusive and diverse in relation to the participants who are asked to evaluate robots, the researchers who develop robots, and the contexts in which robots are envisioned as being deployed. A systematic analysis of the HRI literature showed that HRI, like many other scientific fields, relies on studies from "Western, educated, industrial, rich, and democratic" (WEIRD) populations and that there is insufficient consideration of key axes of diversity—sex and gender, race and ethnicity, age, sexuality and family configuration, disability, body type, ideology, and domain expertise—in the HRI literature (Seaborn et al., 2023). Furthermore, a meta-review of studies from HRI conferences in the 2006–2021 period found that men were overrepresented among research participants and that the field generally treats gender as a binary, in contradiction to best-practice guidelines (Winkle et al., 2023a). In an overview of studies relating to sexbots as an HRI application domain, only one study included nonmale users of these robots (González-González et al., 2020). Finally, among robot developers, people from WEIRD countries are also overrepresented; we have few people from developing nations contributing to the design of robots, and few developers focus on creating solutions that can be affordable and usable in more resource-constrained environments, including rural or lower socioeconomic areas (Johnson et al.,

2017). This lack of diversity in the process and aims of robotics research and development can exacerbate bias in robot design.

To consider the interplay of bias and stereotyping in robot design, think of what happens when we meet people: in order to initiate an impression-formation process, we use central social categories—namely, age, ethnicity, and gender—as reference categories to derive judgments about individuals and their characteristics. Because we often do not have the time and motivation to process information systematically and deeply, this happens relatively quickly and automatically. Would this translate to our impressions about robots as well? Researchers have explored the role of various social categories (e.g., gender, ethnicity) for robot perception by seeing if manipulating specific visual cues or merely the name of the robot to suggest such categories will result in a change in people's perceptions (Eyssel and Loughnan, 2013; Eyssel and Hegel, 2012; Bernotat et al., 2017; Bartneck et al., 2018; Perugia et al., 2023). Studies have shown that even robots designed to be gender-neutral can activate harmful biases in people's perceptions of them because people bring their previous experiences and assumptions to their understanding of robots (Guidi et al., 2023).

When discussing bias, social psychologists like to refer to in-groups versus out-groups, thereby differentiating between the group to which one belongs and that is generally perceived more positively, and "the others." This is called *in-group bias* or *in-group favoritism* (Scheepers et al., 2006) and represents a form of discrimination. In North America, for instance, the intergroup context of African Americans versus Whites has been studied extensively. However, what does this have to do with robots—and with diversity? One online study investigated whether White American people also discriminate between in-group (i.e., robots that look White) and out-group robots (i.e., robots that look Black). At first glance, this experiment produced results that gave some hope: the prediction that people would evaluate the out-group robot as having less "mind" was not supported. However, Eyssel and Loughnan (2013) were able to show that White American participants devalued the robot from the out-group, especially if these people showed a high degree of modern racism. People with racist anti-Black attitudes were also among those who ascribed less mind to the out-group robot in terms of agency and experience. However, it is important to note that individual attitudes did indeed play a role—the prototypical devaluation of an out-group could only be demonstrated when individual prejudices were taken into account. Earlier work (Eyssel and Kuchenbrandt, 2012) found that it was not even necessary to manipulate visual cues for group membership. German participants presented with a picture of the same robot with different names and country-of-production cues (Eyssel and Kuchenbrandt, 2012) manifested in-group bias. They preferred the in-group product over the alleged out-group platform, even at the level of design evaluation.

Moreover, research by Correll et al. (2002) has documented that people discriminate in a way that—not only in their laboratory experiments—can have fatal consequences. In the classic shooter bias paradigm, photos of people

with and without a weapon are shown. The task is to react as quickly as possible to press the button for "shoot" in the event of danger and "do not shoot" when unarmed persons are depicted. The skin color of the people in the pictures had a clear influence on the reaction time. If an African American–looking man held a gun in his hand, he was shot faster than when participants were confronted with a White armed man. If the dark-skinned person carried a harmless cell phone, it took participants longer to refrain from shooting. Bartneck et al. (2018) have replicated the paradigm of the shooting bias experiments with White versus dark-skinned in-group and out-group robots that appeared armed versus unarmed and found analogous results, suggesting that similar implicit racial biases can be at play in human–robot interactions as well.

Relatedly, research by Eyssel and Hegel (2012) and Bernotat et al. (2017) investigated the role of gender in the perception of social robots and showed that widely known stereotypes about men and women in society are upheld even in the context of robots. Currently, in a second wave of interest, various researchers have focused on social categories, including gender, to explore the potential detrimental effects of categorizing not only humans but also robots and to demonstrate the importance of taking such features—on the part of user, researcher, or robot—into account (Perugia and Lisy, 2022; Perugia et al., 2022; Roesler et al., 2022; Winkle et al., 2023a). Notably, most research in the realm of gendering of robots has explored the notion of robot or participant gender in a dichotomous fashion—that is, contrasting "male" versus "female." Contemporary approaches, however, would refrain from such a dichotomous conceptualization and integrate a more diverse, gender-fluid range of gender categories. If one aims to research the impact of the traditional male-versus-female gender categories on social judgments, though, it seems fair to study exactly that. At the same time, research on other forms of gender and effects associated with them is still scarce. Thus, this area holds a plethora of open research questions to be investigated.

Likewise, the richness of potential robot user groups, representing persons with cognitive, physical, or other forms of diversity (e.g., neurodiversity), is yet to be adequately mirrored in HRI studies; the experiences of individuals who are less frequently studied need greater inclusion. Some researchers would argue that when doing so, it is valuable and relevant to give room to voices from these target groups, even as part of the research process. Indeed, this would be a truly human-centered approach.

In addressing the various sources of bias in robotics, such as those mentioned previously, Howard and Borenstein (2018) call on the robotics community not only to identify issues but also to create solutions to problems of bias and racism in robotics by developing a more inclusive moral imagination and proactive stance to address ethical issues and bias before technology is deployed and creates negative societal effects. Howard and Kennedy III (2020), in turn, call on the robotics community to explicitly consider ethical use and equity in performance when designing and deploying robots, and they discuss the formation of the Black in Robotics (BiR) community to start

addressing some of these issues. Winkle et al. (2023b) provide a feminist framing of work in HRI to suggest that we need to examine and challenge, as needed, power relationships in HRI research and development. This may involve being mindful of and at times subverting the power relationships and hierarchies between researchers and participants, such as through participatory design, which also provides opportunities for robot design to incorporate more diverse voices. It can also take into account the differential effects of robotic technologies on people who decide to purchase and deploy them (e.g., corporate managers) and the people who end up having to use them (e.g., factory floor workers). This perspective suggests it is important to empower potential users of robots to participate more substantively in decision-making regarding their appropriate use and deployment and for HRI researchers to actively question the assumptions and power dynamics involved in the research.

12.3 Conclusion

It is important to realize that robots, humans, and their interactions are part of broader societies that encompass different kinds of people, technologies, institutions, and practices. In these different social and cultural contexts, people may hold different initial attitudes and beliefs about robots based on their prior exposure to fictional narratives and popular media. Potential users of robots will also hold different social and cultural values and norms. Both these cultural narratives and values will affect how people perceive and respond to robots and how the use of robots might affect existing social structures and practices. HRI researchers should be conscious of and sensitive to prevailing cultural narratives and values when they design and deploy robots in society, and they should also consider whether they want robots to reproduce or challenge existing practices and norms. HRI research, although already quite interdisciplinary, should open up more space to participants from diverse sociocultural and application-oriented backgrounds to better include the varied experiences and perspectives of those who will be affected by the future adoption and use of robots.

Questions for you to think about:

- What was the last movie or series you watched or book you read that depicted robots?
- List the characteristics of the robot protagonists you have recently seen in a film or TV series. What were their capabilities? Did they appear humanlike? Did they pose a threat to humanity, or did they save the world?
- How will the availability of new forms of media, such as YouTube, change people's expectations toward robots?

- Think of professions that have been replaced by machines. Which ones come to mind? What are the potential positive and negative implications of this replacement?
- Is there an activity that you are happy to have a machine do? What is an activity that you would not want to be replaced by a machine? How do you think others might feel about your choices—who might disagree?
- Discuss whether it is ethical to use a social robot as comfort for lonely elderly people. Describe relevant issues, and explain your opinion.
- In a future where highly intelligent robots are available, would it be ethical to develop robot nannies or robot teachers? Describe the potential issues.
- Some HRI studies are provocative or thought-provoking, for example, Bartneck et al.'s (2018) study on the presence of racism in HRI. Is it ethical to run controversial HRI studies? Are there particular themes, such as religion, where HRI should not tread?

12.4 Exercises

The answers to these questions are available in the Appendix.

*** Exercise 12.1 Sci-Fi media** What was the last movie or series you watched or book you read that depicted robots?

**** Exercise 12.2 *Bicentennial Man*** What is the fictional robot Andrew Martin prepared to do to be fully recognized as a human? Select one or more options from the following list:

1. It agrees not to let any other robot become human.
2. It becomes mortal.
3. It accepts becoming unaware of its own robotic nature.
4. It gives up all friendships.
5. It enters a legal trial.

**** Exercise 12.3 Robotic revolution** The robot uprising is a common theme in the media. Why do the robots typically rebel? Select one or more options from the following list:

1. They mirror humanity's poor behavior during colonization.
2. They compete for resources with humanity and only see the option to kill or enslave humanity.
3. They want to protect life on Earth by removing the people that pollute it.
4. Humans programmed them to do so.
5. They are annoyed by having to take orders from less intelligent beings.

**** Exercise 12.4 Relationship** A robot companion, may it be for elderly care, social companionship, or training for people on the autism spectrum,

might raise ethical issues. Which of the following statements are true? Select one or more options from the following list:

1. Robots are smarter and stronger than humans.
2. Robots have no legal status.
3. Robots will want to deceive humans
4. The imitation of reciprocal affect can never be as meaningful as authentic affect.
5. Robots could set unrealistic expectations for human-to-human relationships.

*** **Exercise 12.5 Trust in robots** Watch this video, and then answer the question that follows.

Ayanna Howard, "Should We Trust Robots and Should They Trust Us?" https://youtu.be/P86kv-v7XJU

1. Ayanna Howard discusses how the general public perceives and interacts with robots. She explains that people often trust, perhaps even over-trust, robots. She mentions the emotional connection to robots and people's preconceptions of robots based on their ideas about robots as advanced technology as some of the reasons for this trust. Explain how these two factors can lead to positive as well as negative outcomes—what are those outcomes, and how do they stem from our relationships and expectations of robots? Furthermore, how can we address these potential problems in our design of robots?

*** **Exercise 12.6 Ethical issues in HRI** Watch this video, and then answer the question that follows.

Kate Darling, "Ethical Issues in Human-Robot Interaction," https://youtu.be/m3gp4LFgPX0?si=ztu7xUShqNYSTTT3

1. Kate Darling describes the new paradigm of social robots that engage with people in diverse contexts, similarly to what we have been discussing so far, and then points out several ethical issues that emerge from the design of and people's interactions with such robots. Based on her talk, explain why social robots may be different from other robots in terms of their ethical implications. Also describe which of the ethical implications Kate Darling describes you found the most surprising or important. How does this implication affect the way you think about designing social robots?

Future reading:

- Jonze, dir., Spike. *Her*. Warner Bros., Burbank, CA, 2013. URL www.imdb.com/title/tt1798709/?ref_=fn_al_tt_1
- Isaac Asimov's Robot series is a collection of short stories and novels published between 1950 and 1986 that were never formally published

as a series but as separate works. https://en.wikipedia.org/wiki/Robot_series

- Dick, Philip K. *Blade Runner: Do Androids Dream of Electric Sheep?* Ballantine Books, New York, 25th-Anniversary edition, 2007. ISBN 9780345350473. URL http://worldcat.org/oclc/776604212
- Schreier, dir., Jake. *Robot and Frank*. Sony Pictures Home Entertainment, Culver City, CA, 2013. URL www.imdb.com/title/tt1990314/
- Sharkey, Amanda J. C. Should we welcome robot teachers? *Ethics and Information Technology*, 18(4):283–297, 2016. doi: 10.1007/s10676-016-9387-z. URL https://doi.org/10.1007/s10676-016-9387-z
- Singer, Peter W. *Wired for War: The Robotics Revolution and Conflict in the Twenty-First Century*. Penguin, New York, 2009. ISBN 9781594201981. URL http://worldcat.org/oclc/857636246
- Veruggio, Gianmarco, Operto, Fiorella, and Bekey, George. Roboethics: Social and ethical implications. In Siciliano, Bruno, and Khatib, Oussama, editors, *Springer Handbook of Robotics*, pages 2135–2160. Springer, New York, 2016. ISBN 978-3-319-32550-7. doi: 10.1007/978-3-319-32552-1. URL https://doi.org/10.1007/978-3-319-32552-1
- Awad, Edmond, Dsouza, Sohan, Kim, Richard, Schulz, Jonathan, Henrich, Joseph, Shariff, Azim, Bonnefon, Jean-François, and Rahwan, Iyad. The moral machine experiment. *Nature*, 563:59–63, 2018. ISSN 1476-4687. doi: 10.1038/s41586-018-0637-6. URL https://doi.org/10.1038/s41586-018-0637-6
- Sparrow, Robert. Robots, rape, and representation. *International Journal of Social Robotics*, 9(4):465–477, 2017. ISSN 1875-4805. doi: 10.1007/s12369-017-0413-z. URL https://doi.org/10.1007/s12369-017-0413-z
- Lin, Patrick, Abney, Keith, and Bekey, George A. *Robot Ethics: The Ethical and Social Implications of Robotics*. Intelligent Robotics and Autonomous Agents. MIT Press, Cambridge, MA, 2012. ISBN 9780262016667. URL http://worldcat.org/oclc/1004334474