



Societal and Ethical Issues in HRI

Ricarda Wullenkord¹ · Friederike Eyssel¹

Published online: 27 August 2020
© The Author(s) 2020

Abstract

Purpose of the Review The purpose of this review is to give an overview of the societal and ethical issues in human-robot interaction, (HRI), mainly focusing on the literature of the last five years.

Recent Findings Both general ethical challenges associated with robot deployment and those specific to human-robot interaction are addressed and complemented by discussions of ethics within HRI research, ethics related behavior towards robots, as well as ethics and robot rights. Moreover, we discuss ethical challenges in sensitive contexts such as medicine, rehabilitation, and care. We conclude our review by providing an overview of the key ethics frameworks and guidelines to inspire researchers, developers, and stakeholders alike.

Summary This review offers a timely overview of the state-of-the art societal and ethical issues that arise from the ever more steady integration of robots into human society. We exemplify the key issues and debates in the field by mainly covering the literature of the past five years.

Keywords Human-robot interaction · Ethics · Robot ethics · Assistive robots · ELS issues

Introduction

Since the introduction of home computers in 1977, rarely has a new technology divided the opinions of people in the same way as robots. Although continuously growing, the current market for personal domestic, and service robots (12.2 million units sold in 2018 worldwide), or “social” robots for entertainment (4.1 million units sold in 2018 worldwide) is small, especially in comparison to the deployment of industrial robots (154 million units sold in 2018 in China alone, 371.5 million units in the 15 biggest markets of the world, [1]). However, it is likely that particularly domestic and social robots will become increasingly prevalent [2, 3], until one day, robots at home will be as common as home computers are

now. Meanwhile, a range of open questions emerges and requires discourse. For instance, researchers need to address the societal and ethical impact associated with the introduction of (social) robots into our everyday lives.

Ethics in general can be defined as principles that distinguish between behavior that helps and behavior that harms [4]. Roboethics is a research area underlying all ethical issues with regard to robots and robotic assistance. Roboethics or robot ethics incorporate “ethical questions about how humans should design, deploy, and treat robots” [5, p. 243]. More specifically, ethical robot behavior is—in this context—understood as “an agent’s behavior governing a system of acts that affects others (i.e., patients) according to moral rules” [6, p. 483]. The importance of ethics in research on robots becomes even more obvious in light of the vast amount of literature on ethics in human-robot interaction (HRI): A literature search on google scholar offers 14.500 results for “ethics and human-robot interaction”, searching for terms like “ethics and robots”, or “ethics and robotics” leads to over 150.000 and over 136.000 results, respectively. The topic of ethics in (social) robotics has been discussed in the literature for decades (e.g., [7–11]). The current work will, however, only provide a glimpse into the most recent issues and debates. We do so by focusing on the last five years of research on ethical and societal issues in the field of HRI.

This article belongs to the Topical Collection on *Service and Interactive Robotics*

✉ Ricarda Wullenkord
rwullenk@cit-ec.uni-bielefeld.de

Friederike Eyssel
feyssel@uni-bielefeld.de

¹ Bielefeld University, Bielefeld, Germany

Areas of Robot Use, and General Ethical Challenges Associated with Robot Deployment

Robots are deployed in various fields of use in which they offer context-specific benefits and challenges. For instance, in industrial settings, robots can increase productivity and relieve workers from completing physically challenging tasks. The automotive industry is a context in which robots have already been used for years [12] to relieve the burden of workers, and to increase productivity and flexibility (e.g., [13]). Furthermore, robots play a role in the military context (e.g., [14]) to reduce the number of human soldiers required for a mission and to minimize the number of casualties. Similarly, robots are used for search-and-rescue tasks in terrains that are either dangerous or inaccessible for human rescue teams (e.g., [15]). Robots can be utilized for sexual pleasure, enabling sexuality without risk of sexually transmitted diseases and unwanted pregnancies, and potentially reducing sex-work related problems, such as sex trafficking [16]. Robots are used in the care sector (e.g., [17]) and in rehabilitation (e.g., [18]) to relieve the burden of care personnel (e.g., [19]). Robots are also beneficial as members in human-robot teams that collaborate in the medical setting, e.g., during surgeries [20]. Finally, robots serve humans as assistants and companions in the home environment (e.g., [21]).

Clearly, apart from such potential benefits, the successful integration of robots in society also introduces several challenges. According to Fosch-Villaronga and colleagues [22•], potential ethical challenges are superordinated by two so-called “meta-challenges”: Uncertainty and responsibility. First, the meta-challenge “uncertainty” refers to user uncertainty concerning laws and regulations of robot use. Uncertainty represents a meta-challenge because many potential legal and societal issues concerning robot use are either still unknown or have not yet been regulated by laws or rules. Second, the meta-challenge “responsibility” refers to the difficulties associated with the open issue of who regulates or holds responsibility when humans interact with robots. This concerns the regulation of robot use, responsibility for damage caused by a robot, for the correct disposal of robots, etc. According to Fosch-Villaronga and colleagues [22•], these two meta-challenges influence each and every ethical issue that is discussed in the literature.

General challenges associated with the deployment of robots vary in terms of their ethical relevance: Robot acceptance (e.g., [23]) and robot usability (e.g., [24, 25]) are deemed less ethically relevant issues compared with concrete fears of potential end users. One big fear of people with regard to robots revolves around job replacement (e.g., [26, 27]). Maurice et al. [28] offer a general discussion on the ethical issues related to robots and assistive technology at the workplace. Acemoğlu and Restrepo [26], as well as Dauth et al. [27] have mainly

covered job replacement in the industry and the labor market, respectively. Despite the fact that robot technology is currently not advanced enough to replace human labor in sectors such as therapy and care some authors have already expressed concern about the future replacement of human caregivers [29, 30]. An additional fear is related to a “too much” of robotic assistance. Likewise, Gransche [31] indicated that excessive assistance by robots could make us either incapable or unwilling to fulfill even simple tasks, and, therefore, rendering humans helpless without robot support.

Besides reflecting concrete fears with regard to robots, there is a high number of potential ethical issues that are summarized under the umbrella term “ethical, legal and security issues” (ELS; e.g., [22•]). Research on ELS issues, especially in the last five years, examined the aspects law and liability, privacy and (data) security, consent, and autonomy, due to the connection between autonomy and security. The aspect of law and liability that is covered within the framework of ELS issues regards the question who is responsible, for example if the robot malfunctions (see [32] or [33] for discussions on the responsibility of machines). This topic often goes hand in hand with privacy and (data) security issues. Who gets access to what data? What threat does hacking pose? Do we know which data a robot is going to collect, and how do we consent? Are we even aware of the presence of a robot in a public space? Questions regarding the collection, storage, and usage of our data, which might be collected by robots around us, are often discussed in the context of Big Data (see [34–38] on Big Data and privacy with regard to assistance technology and robots). Additionally, the deployment of robots in public spaces is not only relevant for privacy and security issues, but also for consent. When humans and robots (need to) interact, it is crucial that there is the opportunity to consent to it or to deny consent (for a discussion on consent in HRI, we refer to [39]). Another important issue that is discussed widely in the context of ethics in robotics in general concerns robot autonomy. According to Bekey [40], autonomous robots are “intelligent machines capable of performing tasks in the world by themselves, without explicit human control over their movements” ([40], p. xiii). Robot autonomy is often addressed in light of Asimov’s “Laws of Robotics” [41], which to this day inspire researchers in HRI. According to Asimov [41], first, a robot must not harm human beings, or humanity, neither through action nor through inaction. Second, a robot must follow human orders, as long as the orders do not lead to harm of another human being, or humanity. Third, a robot must protect its own existence, as long as it does not lead to harm and does not disregard an order given by a human. The “Laws of Robotics” imply that a robot must act if a human/humanity is about to get harmed, even if there are no explicit orders by its user. Accordingly, a robot may refuse an order, if harm would be the consequence of that specific order. Some authors claim that humans must be responsible for the actions

of machines, even if the machines act autonomously [42]. Relatedly, the notion of autonomy is also heavily discussed in the context of autonomous driving since autonomous vehicles make decisions that directly impact human safety, for instance, by Brändle and Grunwald [43], Grunwald [44], and Sparrow and Howard [45].

Even though based in science fiction literature, Asimov’s “Laws of Robotics” have inspired reflections on robot morality and autonomy (e.g., [46–48]). The more autonomous robots become, the more they are potentially capable of making their own decisions—which brings in the notion of robot morality. Malle [5] discusses machine morality, addressing questions concerning a robot’s moral capabilities and their technical implementation. However, regardless of the actual moral capacities and capabilities of robots, humans hold robots accountable for their actions to a certain degree [49]; some studies even show that humans apply the same moral norms to robots as to humans [50]. Banks [51] claims that independent from the robot’s actual level of autonomy and/or agency determining its moral capacity, the robot can be perceived as a moral agent. At the same time, Bigman and Gray [52] show that people are averse to machines that make morally relevant decisions. Other authors suggest that machines cannot be moral under any circumstance [53], or that they cannot be ethical agents at all [54]. For a general critical discussion on moral robots, consider Scheutz and Malle [55]. Those with a particular interest in robot heroism as a special case of robot morality may want to consult Wiltshire [56]. Apart from whether robots will ever engage in moral decision-making, research indicates that humans indeed perceive robots as potential moral agents [49–51]. This, however, has implications for robot users and the expectations they bring into human-robot interaction.

HRI-Specific Ethical Challenges

Complementary to general ethical issues that must be considered when introducing robots into human lives and human society, there are ethical issues that are specifically relevant to HRI. Among such HRI-related topics, discrimination of users and robots (e.g., [57–59]), dehumanization of users (e.g., [60, 61]), and deception by robots (e.g., [62–65]) are frequently discussed topics. Considering discrimination, scholars point to the issue that if robots are programmed by humans, they may fall prey to the same biases that are known to cause problems in human-human interaction (for as discussion of discrimination in AI see [66]). One example for such a bias is racial bias in the use of police force [57], meaning that a robot could fall victim to the same biases as human police officers when deciding on the usage of (deadly) force during a police operation. On top of that, robots may not only discriminate against humans through their behavior but they also

might embody discrimination through their design, commonly featuring Euro-centric or overly feminized design [58]. Sparrow [67] even argues that due to the perception of robots as slaves a robot appearance resembling the ethnicity of groups formerly abused as slaves might be highly problematic.

Another ethical and social issue that is broadly reflected upon concerns the use of robots to alleviate the lack of social connection faced by some groups in society. The idea is that robots will ultimately replace human social relationships, resulting in dehumanization of the human (elderly) user by society [e.g., 60, 64, 68, 69]. However, a lack of social connection might not be unique to the elderly population. For example, Yamaguchi [70] reports about individuals who have married a virtual agent because of a lack of potential human relationship partners, or because of their lack of the social competence necessary to establish and maintain close human-human relationships. De Graaf [61] argues that this issue might aggravate, as she claims that in a society in which robots are a matter of course humans’ social skills and willingness to “deal with the complexity of real human relationships” might decrease [61, p. 595]. However, when thinking about the relationship between humans and robots, there are more issues to consider than just the replacement of humans. Relationships between humans and robots might even be considered deceptive by their very nature, as they can only simulate a connection that resembles a human-human relationship. Exemplary questions in this context are as follows: Does a robot deceive us when simulating a connection resembling a human-human relationship? Is a robot allowed to lie? It can be argued that robot deception might be legitimate under some circumstances, for instance, when the goal is to make the user feel positive or comfortable [62]. Other researchers concur that robot deception is ethically problematic, no matter what [64, 65]. One topic that is especially relevant with regard to robotic deception in HRI is empathy, more specifically the evocation of empathy in the user. Coeckelbergh [71] argues that the recognition of the vulnerability of humans as embodied beings, and the fact that human beings recognize each other as equally vulnerable, is one necessary condition for empathy to emerge. He names that recognition of vulnerability mirroring and deems the notion of robots being vulnerable a necessary prerequisite for vulnerability mirroring. However, because robots cannot be vulnerable in the same sense as humans are, the idea of robot vulnerability may be associated with deception as well. Liberati and Nagataki [72] elaborately discuss the ethics of vulnerability in relationships between humans and robots, which includes empathy as well. With regard to the question of simulating a social connection between humans and robots, Coeckelbergh [63] suggests that robots can never be friends in an Aristotelian meaning, since they lack the mutuality and reciprocity necessary to form a friendship. Coeckelbergh [63] also argues that what is

considered deception in some works is not necessarily deception, as the term deception in this context implies that robots create a virtual world that contrasts the “real” world, and that this is not necessarily the case. To evaluate robot deception in any given case, it might be necessary to consider whether the robot behavior counts as deception under the specific circumstance and, if so, if the deception is necessary or beneficial for the individual.

Apart from discrimination, dehumanization, and deception, which represent phenomena that are potentially relevant for all types of robots involved in HRI, some authors suggest that there are specific ethical issues related to socially assistive robots (SAR), in particular (e.g., [73]). They propose that these issues are unique to SAR due to their more social nature compared with other types of robots. SAR are defined as a class of robots between “assistive robotics (robots that provide assistance to a user) and socially interactive robotics (robots that communicate with a user through social and nonphysical interaction)” [74, p. 25]. Wilson et al. [73] suggest the following ethical issues are particularly relevant for social robots: A respect for social norms, the robot being able to make decisions about competing obligations, building and maintaining trust between robot and user, the potential problem of social manipulation and deception by the robot, and the issue of blame and justification, especially if something goes wrong [73]. As the task of building and maintaining trust between robots and users is an important ethical factor in the contexts of socially assistive robots [73], there are trust-based approaches to ethical social robots. These emphasize the importance of building and maintaining trust, and the potential pitfalls of trust between user and robot. To illustrate, Koyama [75] presents a recent trust-based approach on the ethics of social robots. In addition to ethical issues specific to SAR, the discussion on ethics in HRI also features cyber-physical systems, which, in this context, are understood as intelligent robotics systems linked to the Internet of Things, that interact with the physical world [76]. Furthermore, for a classic overview over ethics in HRI, we recommend Lin et al. [77], and for a recent overview, we refer to Bartneck [78], respectively.

Moreover, there are two further areas in HRI in which ethics play a major role: Ethics in the conduct of HRI research, and ethics related to robot rights. HRI research and the field’s specific research methods bring along their very own ethical issues to consider. One of the most important issues in this context, which has previously been discussed in the context of relationships between humans and robots in general, is deception of the user. Deception is frequently used in research and is often deemed necessary, because a complete disclosure of all information regarding the experiment would highly influence participant reactions. In HRI research, deception is especially important as the Wizard-of-Oz approach is frequently used. Therefore, with regard to ethics, the possibility of deception through an improper use of a Wizard-of-Oz approach and the

following potential for embarrassment of the participant are to be acknowledged by the researcher [79], as well as the “Turing Deceptions” [80, 81]. Because this article focuses on HRI research in general rather than on research methods in particular, we recommend Punchoojit and Hongwarittorn [82] who cover the ethical issues that must be recognized when conducting HCI or HRI research.

Regarding robot rights, the ethical and societal issues discussed previously took a human-centered but not a robot-centered perspective. However, literature also addresses the ethical and societal issues that target robots. For instance, Loh [83] refers to the difference between robots as moral agents and moral patients, which can be applied to robots as ethical agents and ethical patients as well. Literature on this topic examines and discusses behavior towards robots, robot rights, and the question of whether ethics apply to robots at all. The topic of robot rights and behavior towards robots is vast enough to require its own literature review. Therefore, we recommend the following literature: [59, 84–88] to gain further insights into this matter.

Sensitive Areas of Robot Deployment and Associated Ethical Challenges

There are some areas of robot deployment that can be regarded as potentially more ethically sensitive than others, introducing domain-specific ethical challenges. The use of robots for warfare, sexual pleasure, or to care for vulnerable target groups are some key examples. A more general perspective on robots for warfare is provided by Andreas [89]. Philosopher Robert Sparrow, too, has intensively researched the notion of robot killers (e.g., [90–93] as well as Sparrow and Lucas, [94] on robots for war at sea), but has also inspired scholarly discourse on sex robots, discussing it in the context of robot rape [95]. Not less ethically sensitive is the issue of robotic assistance in the medical field and carebot use to assist vulnerable end users, such as people with cognitive impairments, children, or seniors. Steil et al. [96] provide valuable insights into the ethical challenges associated with robot deployment in medical settings. In the field of robotic care, robots are employed for the care of elderly people, people with disabilities, and children. These groups can be considered vulnerable due to age, reduced or not yet fully developed cognitive and/or physical abilities, e.g., in the case of dementia (see [97] or [98] on ethical recommendations for assistive robotics in dementia care) or due to ongoing cognitive and/or physical development conditioned by a young age (e.g., [99]). Robots can be very helpful in assisting those groups and/or their caretakers in completing tasks, by monitoring user health and user behavior, and by providing companionship ([64]; for a description of a robotic care system, see [100] or [101]).

However, when closely interacting and co-sharing space with robots in general and with carebots in particular, physical safety of users has to be assured (e.g., [102]). Physical safety is not the only issue that has to be taken into account with regard to the interaction between carebots and humans, though. The topic of ethics in robotic care is widely discussed in the literature. Starting broadly, Manzeschke, [103] provides a general discussion on ethics in robotic care, taking into account the different levels of relations between robots and humans in this context: The robot as a mere tool, the robot as a tool with social capabilities, and the robot as an agent the human develops a relationship to. Especially with regard to the specific relationship between humans and robots in the context of care, Körtner [104] suggests six aspects to consider for an ethical integration of carebots into the user's lives: First, he proposes deception, understood as the potential of the user to form incorrect ideas of the robot's abilities with regard to cognition and emotion. Second, he names dignity, referring to the risk of patronizing or infantilization of elderly people (e.g., by giving dementia patients the robot *Paro* [105] as a toy to play with). Third, he refers to isolation, since robots might replace all human contact. Fourth, he mentions privacy, especially regarding the fact that people who are reliant on care potentially are more willing to sacrifice privacy in favor of care and security. Fifth, he lists safety, which might be more important for elderly people as due to reduced walking stability they might be knocked over by a robot more easily than the younger population. Finally, he suggests vulnerability due to potentially reduced cognitive abilities (e.g., due to dementia) and, therefore, a reduced ability to give consent to interaction with a robot. However, this list is not necessarily exhaustive. Zwijsen et al. [106] propose the following factors as specifically important in the context of robots in elderly care: The personal living environment (encompassing privacy, autonomy, and obtrusiveness), the outside world (encompassing stigma and human contact), and the design of the assistance technology (comprising of individual approach, affordability and safety). Manzeschke et al. [107] argue that the following fields are relevant when comparing elderly users to the general user population: The elderly might have lower financial possibilities than the working populations, there are privacy aspects to acknowledge because more health-related data are collected for elderly people, and shared with doctors and care-givers, they might suffer from reduced mobility, their user involvement and robot acceptance might be lower, and their expectations towards the technology might be different, due to reduced experience with modern technologies, for example. While being rather reserved towards robots in elderly care, Sharkey and Sharkey [64] reflect on ethical challenges with regard to robots caring for the elderly as well and suggest the following aspects that need to be considered: Potential reduction in the amount of human contact of elderly people, increased objectification of dementia patients, privacy issues,

loss of personal liberty, deception and infantilization, and the question of who is to control the robots. Due to the heterogeneity of the ethical aspects different authors propose for the use of robots in elderly care, ethical aspects in elderly care might be a topic that is interesting for a review on its own.

Above and beyond problems that have to be examined with regard to robots in elderly care, Riek and Howard [58] extend the discussion to issues to consider when deploying robots in other sensitive fields, such as robots in therapy and general care settings. First, they refer to the problem of using therapeutic robots during research projects, more specifically what happens once the project is finished. Usually, the robots are removed again, which may revoke all benefits the robots brought to the patients, leaving them in a worse state than before. Second, they refer to problems specific for physically assistive robots, namely, help with sensitive tasks as bodily hygiene, and the fact that the users will probably develop an emotional bond to the robots as they might have little contact to other people. They cite works by authors such as Forlizzi and DiSalvo [108], Riek et al. [109], Scheutz [110], and Carpenter [111] to support their claim that no matter the morphology of the platform, a certain extent of bonding will inevitably form. Apart from using robots with varying degrees of autonomy in the care sector, there is also the option of using telepresence robots. Niemelä et al. [112] provide ethical guidelines for using telepresence robots in residential care. Their results showed that sometimes ethical considerations were deemed more important than usability concerns. For example, it was considered crucial that the primary user, i.e., the elderly person needs to maintain control over accepting or rejecting an incoming call via the robot, no matter what the intention of the call was. The participation of family members in health checks or hygiene care by the telepresence robot was considered ethically problematic and, therefore, was advised against. As a telepresence robot offers the possibility to be remotely controlled by a family member or a care worker, the authors argue that the aspect of the invasion of privacy by the robot is even more important than with regard to conventional robots. For a more general discussion on ethical aspects of telepresence robots, we recommend Oliveira et al. [113]).

Ethical Frameworks, Guidelines, and Their Implementation into Robots

Given the vast number of ethical issues to consider when designing robots for the various roles and user groups in current and future societies, it becomes clear that theoretical frameworks and guidelines are called for to bundle the multi-disciplinary scholarship on ethics in (social) robotics and HRI. The frameworks and guidelines range from very broad theoretical discussions on the topic of ethics to detailed suggestions for concrete algorithms necessary for robots to behave in ethical ways. Reijers et al. [114•] give an extensive systematic

literature review on the methods to incorporate ethics into research and innovation in general.

Veruggio [115] takes general ethical problems into account that are linked to relationship formation between humans and machines (e.g., humanization of the human/machine relationship (cognitive and affective bonds towards machines) [115, p. 615]) and suggests an ethical framework on the basis of the so-called “PAPA” code of ethics, which is taken from Computer and Information Ethics. The acronym PAPA stands for privacy, accuracy, intellectual property, and access [115]. Privacy deals with the question of which information we must reveal to others under which conditions and protections, and which information we can keep secret. Accuracy refers to the question of responsibility, more specifically, addressing who is responsible to make sure the information is authentic and accurate, and who is responsible if there are errors and damages to repair. The notion of property suggests the question of ownership of information, the fairness of costs of information exchange, potential ways of information exchange, the ownership of said ways, and the regulation of the information exchange. Finally, accessibility comprises the right of a person and/or organization to obtain information, and the surrounding conditions [115]. These ethical recommendations are applicable to all relationships between humans and machines and are not exclusively relevant for robots. Thus, the French advisory commission for the ethics of information and communication technology (ICT) research CERNA (Commission for the Ethics of Research in Information Sciences and Technologies) recommends general ethical standards for robotics, aiming at providing tools and recommendations for research institutions and the associated researchers [116]. CERNA’s recommendations concern all ethically relevant fields, ranging from autonomy and decision-making over imitation of life, affective, and social interaction to robot-aided therapy and human-robotic augmentation. For more specific recommendations for dealing with robots that are growing more and more intelligent, consider Kornwachs [117], for example. Even more specifically, taking assistance robots into the focus, literature refers to the five ethical principles underlying the distribution and use of assistance technology by Kitchener [118], namely, beneficence, nonmaleficence, justice, autonomy, and fidelity. Beneficence is supposed to ensure that actions lead to results benefiting others. Nonmaleficence is connected to the previously mentioned laws by Asimov [41] and states that no harm should be caused to others. Justice refers to fairness in different contexts, namely, individual, interpersonal, organizational, and societal contexts. Autonomy aims at freedom of action and choice. Fidelity is the principle of behaving in a loyal, trustworthy, faithful, and honest way (also see [119]).

Parallel to the spectrum the literature on the ethical challenges associated to robot deployment covers, the proposed ethical frameworks and guidelines also cover sensitive areas

of HRI, such as carework. Accordingly, Riek and Howard [58] formulate “Specific Principles of Human Dignity Considerations”. In their work, they listed 15 principles that must be considered when designing a robot or assistance technology. The principles encompass privacy, emotional needs, physical, and psychological capabilities of the user, predictability of the robot, trust, and more formal issues such as laws and regulations. Some exemplary principles read: “The emotional needs of humans are always to be respected”, “Maximal, reasonable transparency in the programming of robotic systems is required.”, or “Avoid racist, sexist, and ableist morphologies and behaviors in robot design” ([58, p. 6]. These principles are an important guiding framework for the development of assistance technology maintaining and supporting human dignity. In addition, Misselhorn et al. [120] offer an ethical framework for the use of robots in the care context, illustrating their principles using the therapeutic seal robot Paro. For a general overview of ethical frameworks for the use of robots in elderly care, Vandemeulebroucke et al. [121] provide a systematic literature review on different ethics approaches and/or frameworks addressing the ethical issues of robots in the care sector; additionally, Mansouri et al. [122] offer a more general review on ethical frameworks for assistive devices, especially for usage in elderly care. Finally, Huber et al. [123] take into account the aspect of relationships between humans and robots and suggest the “Triple-A Model” to incorporate ethics in the design of social companion robots. The model covers the aspects assistance, adaptation, and attachment, and is supposed to help with the identification of ethical risks, and potential ethical risks, based on the different interaction levels of companion robots.

Evidently, the literature offers a rich body of research on ethical frameworks and guidelines to facilitate robot uptake in society. Thus, before deploying robot technology in any field whatsoever, it would be wise to conceptualize specific use cases, to take into account diverse user needs (e.g., through participatory design) and to reflect upon short- and long-term implications of the given scenario for the particular user group. In this respect, consulting the user group-specific ethical frameworks can be helpful. Which frameworks are ultimately consulted depends on where the research is actually conducted. To illustrate, Weber [124] compares three frequently used ethics frameworks in German-speaking countries: The “MEESTAR” model (e.g., [125, 126]), action sheets [127], and the ethics canvas [128]. MEESTAR is a model for the ethical evaluation of socio-technical arrangements and should be used by all stakeholders concerned with the usage of the respective technology. The stakeholders are supposed to do a moral evaluation of the technology at hand and incorporate the results into the development process. It was originally developed for the ethical evaluation of technology used for elderly care. Action sheets can be used to adapt the evaluation dimensions in the MEESTAR model for other fields in a

systematic way. The ethics canvas is an online tool and can be used to gain an overview over a moral field. Stakeholders are supposed to gather their knowledge and assumptions about different categories, i.e., affected people and/or groups, their relationships, and potential conflicts. This way, the expertise of all people potentially involved with the technology can be taken into account.

The ethical frameworks for different contexts of robot use give concrete recommendations on how robots should or should not behave in certain situations. However, these recommendations reflect only a theoretical side of ethical robot behavior. Another step is required to make the robots behave as intended in practice, namely, the concrete technological implementation of ethics into robots. Different researchers have developed and tested algorithms that, for example, allow robots to decide in morally ambivalent situations (e.g., [129–137]). However, it may be critically discussed whether an implementation of ethics in the form of algorithms is feasible, or even possible. McBride and Hoffman [138, p.77] argue, that there is an “immense gap [...] between the architecture, implementation, and activity of humans and robots in addressing ethical situations”. They claim that a robot’s ethical capabilities are reduced to decisions in simple environments, while a human’s ethical capabilities are much more complex. Therefore, they suggest that, instead of applying the same ethical fundamentals used to guide human behavior to robots, it is necessary for humans and robots to communicate on ethics and explore the field together, to come to a new form of guidelines for ethical robot behavior. As approaches towards machine learning are especially relevant for fields in which the programming of concrete algorithms is out of scope due to the complexity of the task, aiming at a shared exploration of ethical situations might be a feasible solution to help transform robots into ethical agents.

Summary and Conclusion

Taken together, we demonstrated that roboethics is a highly complex and increasingly important topic with a vast amount of literature and discussion to examine. To give a starting point, the current review featured the societal and ethical issues in human-robot interaction concentrating on advancements within the last five years. The topics discussed range from general ethical issues that emerge from the introduction of robots into human lives and human society to very concrete ethical issues for specific contexts, such as robots in the care sector. An overview of ethical frameworks and guidelines and their technological implementation into robots aims at providing answers to the open ethical questions. It is imperative for the successful integration of robots into society and into our homes that ethical issues are considered in robotics research, robot development, and the deployment of robots in their various fields of use. Therefore, ethics in robotics is not only highly relevant for the scientific community, but also for developers, technicians, and prospective end

users. Given the rapid technological development, there is a high probability that one day, robot co-share our daily lives. Until then and beyond, ethics in (social) robotics and HRI will remain a crucial, if not, inevitable field of multidisciplinary scholarship, providing rich resources to ameliorate human interactions with novel technologies.

Acknowledgments We thank Angelika Penner, Julia Stapels, and Marlena Fraune for their helpful feedback on previous versions of this manuscript.

Authors’ Contributions Friederike Eyssel provided the idea for the article. Ricarda Wullenkord did the literature search and provided a first draft of the article. Friederike Eyssel and Ricarda Wullenkord jointly rewrote and revised the draft and prepared it for publication.

Funding Information Open Access funding provided by Projekt DEAL. This research was funded by the Ministry of Education and Research (Project “poliTE”; grant no. 16SV7880K).

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

Papers of particular interest, published recently, have been highlighted as:

- Of importance
- Of major importance

1. Economic Commission for Europe, & International Federation of Robotics. World Robotics: United Nations Publications; 2019.
2. Gates B. A robot in every home. *Sci Am.* 2007;296:58–65. <https://doi.org/10.1038/scientificamerican0107-58>.
3. Rus D. The robots are coming. *Foreign Aff.* 2015;94:2–6.
4. Paul R, Elder L. The miniature guide to understanding the foundations of ethical reasoning. United States: Foundation for Critical Thinking; Free Press; 2006.

5. Malle BF. Integrating robot ethics and machine morality: The study and design of moral competence in robots. *Ethics Inf Technol.* 2016;18:243–56. <https://doi.org/10.1007/s10676-015-9367-8>.
6. Al-Fedaghi SS. Typification-based ethics for artificial agents. In: 2008 2nd IEEE International Conference on Digital Ecosystems and Technologies. Phitsanulok, Thailand: IEEE; 2008. p. 482–91. <https://doi.org/10.1109/dest.2008.4635149>.
7. Gips J. Towards the ethical robot. In: Ford KM, Glymour CN, Hayes PJ, editors. *Android epistemology*. Menlo Park, Cambridge, MA: MIT Press; 1995. p. 243–52. <https://doi.org/10.1017/CBO9780511978036.019>.
8. Moor JH. Is ethics computable? *Metaphilosophy.* 1995;26:1–21. <https://doi.org/10.1111/j.1467-9973.1995.tb00553.x>.
9. Hall JS. Ethics for machines. In: Anderson M, Anderson SL, editors. *Machine ethics*. Cambridge: Cambridge University Press; 2000. p. 28–44. <https://doi.org/10.1017/CBO9780511978036.005>.
10. Arkin RC. Robot ethics. *Ethics Inf Technol.* 2002;4:305–18.
11. Petersen S. The ethics of robot servitude. *J Exp Theor Artif Intell.* 2007;19:43–54. <https://doi.org/10.1080/09528130601116139>.
12. Choi S, Eakins WJ, Fuhlbrigge TA. Trends and opportunities for robotic automation of trim & final assembly in the automotive industry. In: 2010 Automation Science and Engineering (CASE); 21.8.2010-24.8.2010. Toronto, Canada: IEEE; 2010. p. 124–9. <https://doi.org/10.1109/COASE.2010.5584524>.
13. Fragapane G, Ivanov D, Peron M, Sgarbossa F, Strandhagen JO. Increasing flexibility and productivity in Industry 4.0 production networks with autonomous mobile robots and smart intralogistics. *Ann Oper Res.* 2020. <https://doi.org/10.1007/s10479-020-03526-7>.
14. Franke UE. Military robots and drones. In: Galbreath DJ, Deni JR, editors. *Routledge Handbook of Defence Studies*. New York: Routledge; 2018. p. 339–49. <https://doi.org/10.4324/9781315650463-28>.
15. Sheh R, Schwertfeger S, Visser A. 16 years of RoboCup rescue. *KI-Künstliche Intelligenz.* 2016;30:267–77. <https://doi.org/10.1007/s13218-016-0444-x>.
16. Döring N, Pöschl S. Sex toys, sex dolls, sex robots: Our under-researched bed-fellows. *Sexologies.* 2018;27:e51–5. <https://doi.org/10.1016/j.sexol.2018.05.009>.
17. Abdi J, Al-Hindawi A, Ng T, Vizcaychipi MP. Scoping review on the use of socially assistive robot technology in elderly care. *BMJ Open.* 2018;8:e018815. <https://doi.org/10.1136/bmjopen-2017-018815>.
18. Babaiasl M, Mahdioun SH, Jaryani P, Yazdani M. A review of technological and clinical aspects of robot-aided rehabilitation of upper-extremity after stroke. *Disabil Rehabil.* 2016;11:263–80. <https://doi.org/10.3109/17483107.2014.1002539>.
19. Blake V. Regulating care robots. *Temple Law Review.* 2019;92:1–52.
20. Taylor RH, Menciassi A, Fichtinger G, Fiorini P, Dario P. Medical robotics and computer-integrated surgery. In: Siciliano B, Khatib O, editors. *Springer Handbook of Robotics*. Cham: Springer; 2016. p. 1657–84. https://doi.org/10.1007/978-3-540-30301-5_53.
21. Wang J, Liu T, Liu Z, Chai Y. Affective interaction technology of companion robots for the elderly: A review. In: El Rhalibi A, Pan Z, Jin H, Ding D, Navarro-Newball AA, Wang Y, editors. *International Conference on E-Learning and Games*. Cham: Springer; 2018. p. 79–83. https://doi.org/10.1007/978-3-030-23712-7_11.
- 22•• Fosch-Villaronga E, Lutz C, Tamò-Larrieux A. Gathering expert opinions for social robots' ethical, legal, and societal concerns: Findings from four international workshops. *Int J Soc Robot.* 2019:1–18. <https://doi.org/10.1007/s12369-019-00605-z>. **This paper summarizes expert discussions from international workshops on ELS issues associated to social robots from the years 2015 – 2017. From an interdisciplinary perspective the potential ethical issues for workers, users, and developers are outlined and possible solutions are proposed.**
23. De Graaf MMA, Allouch SB. Exploring influencing variables for the acceptance of social robots. *Robot Auton Syst.* 2013;61:1476–86. <https://doi.org/10.1016/j.robot.2013.07.007>.
24. Feingold Polak R, Elishay A, Shachar Y, Stein M, Edan Y, Levy Tzedek, S. Differences between young and old users when interacting with a humanoid robot: A qualitative usability study. *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction 2018:107-108.* <https://doi.org/10.1145/3173386.3177046>.
25. Schmidler J, Körber M, Bengler K. A trouble shared is a trouble halved - Usability measures for human-robot collaboration. In: 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC); 9.10.2016-12.10.2016. Budapest, Hungary: IEEE; 2017. p. 000217–22. <https://doi.org/10.1109/SMC.2016.7844244>.
26. Acemoğlu D, Restrepo P. Robots and jobs: Evidence from US labor markets. *J Polit Econ.* 2020;128:2188–244. <https://doi.org/10.1086/705716>.
27. Dauth W, Findeisen S, Südekum J, Woessner N. German robots - the impact of industrial robots on workers. *IAB.* 2017;12306:1–63 <https://ssrn.com/abstract=3039031>.
28. Maurice P, Allienne L, Malaisé A, Ivaldi S. Ethical and social considerations for the introduction of human-centered technologies at work. In: 2018 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO); 27.9.2018-29.9.2018. Genova, Italy: IEEE; 2019. p. 131–8. <https://doi.org/10.1109/ARSO.2018.8625830>.
29. Sparrow R, Sparrow L. In the hands of machines? The future of aged care. *Mind Mach.* 2006;16:141–61. <https://doi.org/10.1007/s11023-006-9030-6>.
30. Pearson Y, Borenstein J. Creating “companions” for children: The ethics of designing esthetic features for robots. *AI & Soc.* 2014;29:23–31. <https://doi.org/10.1007/s00146-012-0431-1>.
31. Gransche B. Assisting ourselves to death – a philosophical reflection on lifting a finger with advanced assistive systems. In: Fritzsche A, Oks SJ, editors. *The future of engineering. Philosophical foundations, ethical problems and application cases*. Cham: Springer; 2018. p. 271–89. https://doi.org/10.1007/978-3-319-91029-1_19.
32. Coy W. Ethik, Verantwortung und Haftung autonomer Maschinen, [Ethics, responsibility and liability of autonomous machines]. In: Klumpp D, Lenk K, Koch G, editors. *Überwiegend Neuland: Positionsbestimmungen der Wissenschaft zur Gestaltung der Informationsgesellschaft, [Predominantly virgin soil: Determination of the positioning of science regarding the information society]*. Baden-Baden: Nomos Verlagsgesellschaft mbH & Co. KG; 2014. p. 110–5. <https://doi.org/10.5771/9783845269269-110>.
33. Hubig C. Haben autonome Maschinen Verantwortung?, [Do autonomous machines have responsibility?]. In: Hirsch-Kreinsen H, Karačić A, editors. *Autonome Systeme und Arbeit. Perspektiven Herausforderungen und Grenzen der Künstlichen Intelligenz in der Arbeitswelt, [Autonomous systems and work. Perspectives, challenges and boundaries of artificial intelligence in the working environment]*. Bielefeld: Transcript; 2019. pp. 275–298. <https://doi.org/10.14361/9783839443958-011>
34. Manzeschke A, Assadi G, Viehöver W. The role of big data in ambient assisted living. *Ethics of Big Data.* 2016;24:22–32.
35. Kappler K, Schrape JF, Ulbricht L, Weyer J. Societal implications of big data. *KI - Künstliche Intelligenz [AI – Artificial*

- Intelligence]. 2018;32:55–60. <https://doi.org/10.1007/s13218-017-0520-x>.
36. Lutz C, Schöttler M, Hoffmann CP. The privacy implications of social robots: Scoping review and expert interviews. *Mobile Media & Communication*. 2019;7:412–34. <https://doi.org/10.1177/2050157919843961>.
 37. Lutz C, Tamò A. RoboCode-Ethicists: Privacy-friendly robots, an ethical responsibility of engineers? In: de Roure D, Burnap P, Halford S, de Roure DC, editors. *Proceedings of the 2015 ACM Web Science Conference*. New York: The Association for Computing Machinery; 2015. p. 1–12. <https://doi.org/10.1145/2786451.2786465>.
 38. Wiegerling K, Nerurkar M, Wadepful C. Ethische und anthropologische Aspekte der Anwendung von Big-Data-Technologien, [Ethical and anthropologic aspects of the application of big data technology]. In: Kolany-Raiser B, Heil R, Orwat C, Hoeren T, editors. *Big Data und Gesellschaft: Eine multidisziplinäre Annäherung, [Big data and society: A multidisciplinary approach]*. Wiesbaden: Springer Fachmedien Wiesbaden; 2018. p. 1–67. https://doi.org/10.1007/978-3-658-21665-8_1.
 39. Sarathy V, Arnold T, Scheutz M. When exceptions are the norm. Exploring the role of consent in HRI. *ACM Trans Hum-Robot Interact*. 2019;8:1–21. <https://doi.org/10.1145/3341166>.
 40. Bekey GA. *Autonomous robots: From biological inspiration to implementation and control*. Cambridge, MA: MIT press; 2005.
 41. Asimov I. *Runaround*. *Astounding science fiction*. 1942;29:94–103.
 42. Johnson DG, Noorman M. Principles for the future development of artificial agents. In: 2014 IEEE International Symposium on Ethics in Science, Technology and Engineering; 23.5.2014–24.5.2014. Chicago, IL, USA: IEEE; 2014. p. 1–3. <https://doi.org/10.1109/ETHICS.2014.6893395>.
 43. Brändle C, Grunwald A. Autonomes Fahren aus Sicht der Maschinenethik, [Autonomous driving from a machine ethics view]. In: Bendel O, editor. *Handbuch Maschinenethik, [Handbook of Machine Ethics]*. Wiesbaden: Springer; 2019. p. 281–300. https://doi.org/10.1007/978-3-658-17483-5_18.
 44. Grunwald A. Self-driving cars: Risk constellation and acceptance issues. *Delphi*. 2018;1:8–13. <https://doi.org/10.21552/delphi/2018/1/7>.
 45. Sparrow R, Howard M. When human beings are like drunk robots: Driverless vehicles, ethics, and the future of transport. *Transportation Research Part C: Emerging Technologies*. 2017; 80:206–15. <https://doi.org/10.1016/j.trc.2017.04.014>.
 46. Clarke R. Asimov’s laws of robotics: Implications for information technology. In: Anderson M, Anderson SL, editors. *Machine ethics*. New York: Cambridge University Press; 2011. p. 254–84. <https://doi.org/10.1109/2.247652>.
 47. Winfield AFT, Blum C, Liu W. Towards an ethical robot: Internal models, consequences and ethical action selection. In: Mistry M, Leonardis A, Witkowski M, Melhuish C, editors. *Advances in autonomous robotics systems. TAROS 2014. Lecture Notes in Computer Science*, vol. 8717. Cham: Springer; 2014. p. 85–96. https://doi.org/10.1007/978-3-319-10401-0_8.
 48. Pereira LM, Lopes AB. Is it possible to program artificial emotions? A basis for behaviours with moral connotation? In: *Machine Ethics*. Cham: Springer; 2020. p. 87–92. https://doi.org/10.1007/978-3-030-39630-5_12.
 49. Kahn PH, Severson RL, Kanda T, Ishiguro H, Gill BT, Ruckert JH, et al. Do people hold a humanoid robot morally accountable for the harm it causes? In: 2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI); 5.3.2012–8.3.2012. Boston, Massachusetts, USA: IEEE. p. 33–40. <https://doi.org/10.1145/2157689.2157696>.
 50. Komatsu T. Japanese students apply same moral norms to humans and robot agents: Considering a moral HRI in terms of different cultural and academic backgrounds. In: 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI); 7.3.2016–10.3.2016. Christchurch, New Zealand: IEEE; 2016. p. 457–8. <https://doi.org/10.1109/HRI.2016.7451804>.
 51. Banks J. A perceived moral agency scale: Development and validation of a metric for humans and social machines. *Comput Hum Behav*. 2019;90:363–71. <https://doi.org/10.1016/j.chb.2018.08.028>.
 52. Bigman YE, Gray K. People are averse to machines making moral decisions. *Cognition*. 2018;18:21–34. <https://doi.org/10.1016/j.cognition.2018.08.003>.
 53. Johnson AM, Axinn S. Acting vs. being moral: The limits of technological moral actors. In: 2014 IEEE International Symposium on Ethics in Science, Technology and Engineering; 23.5.2014–24.5.2014. Chicago, IL, USA: IEEE; 2014. p. 1–4. <https://doi.org/10.1109/ETHICS.2014.6893396>.
 54. Moor JH. The nature, importance, and difficulty of machine ethics. *IEEE Intell Syst*. 2006;21:18–21. <https://doi.org/10.1109/MIS.2006.80>.
 55. Scheutz M, Malle BF. Moral robots. In: Syd L, Johnson M, Rommelfanger KS, editors. *The Routledge Handbook of Neuroethics*. New York, NY: Routledge; 2018. p. 363–77. <https://doi.org/10.4324/9781315708652-27>.
 56. Wiltshire TJ. A prospective framework for the design of ideal artificial moral agents: Insights from the science of heroism in humans. *Mind Mach*. 2015;25:57–71. <https://doi.org/10.1007/s11023-015-9361-2>.
 57. Asaro P. Hands up, don’t shoot!: HRI and the automation of police use of force. *J Hum-Robot Interaction*. 2016;5:55–69. <https://doi.org/10.5898/JHRI.5.3.Asaro>.
 58. Riek L, Howard D. A code of ethics for the human-robot interaction profession. *Proceedings of We Robot 2014*. <https://ssrn.com/abstract=2757805>. Accessed 13 May 2020.
 59. Sparrow R. Robotics has a race problem. *Sci Technol Hum Values*. 2019a;45:538–60. <https://doi.org/10.1177/0162243919862862>.
 60. Sharkey N, Sharkey A. The eldercare factory. *Gerontology*. 2012a;58:282–8. <https://doi.org/10.1159/000329483>.
 61. De Graaf MMA. An ethical evaluation of human–robot relationships. *Int J Soc Robot*. 2016;8:589–98. <https://doi.org/10.1007/s12369-016-0368-5>.
 62. Arkin RC, Ulam P, Wagner AR. Moral decision making in autonomous systems: Enforcement, moral emotions, dignity, trust, and deception. *Proc IEEE*. 2012;100:571–89. <https://doi.org/10.1109/JPROC.2011.2173265>.
 63. Coeckelbergh M. Care robots and the future of ICT-mediated elderly care: A response to doom scenarios. *AI & Soc*. 2016;31: 455–62. <https://doi.org/10.1007/s00146-015-0626-3>.
 64. Sharkey A, Sharkey N. Granny and the robots: Ethical issues in robot care for the elderly. *Ethics Inf Technol*. 2012b;14:27–40. <https://doi.org/10.1007/s10676-010-9234-6>.
 65. Sparrow R. The March of the robot dogs. *Ethics Inf Technol*. 2002;4:305–18. <https://doi.org/10.1023/A:1021386708994>.
 66. Zou J, Schiebinger L. AI can be sexist and racist - it’s time to make it fair. *Nature*. 2018;559:324–6. <https://doi.org/10.1038/d41586-018-05707-8>.
 67. Sparrow R. Do robots have race?: Race, social construction, and HRI. *IEEE Robotics & Automation Magazine*. 2019b;1–20. <https://doi.org/10.1109/MRA.2019.2927372>.
 68. Sparrow R. Robots in aged care: A dystopian future? *AI & Soc*. 2016;31:445–54. <https://doi.org/10.1007/s00146-015-0625-4>.
 69. Turkle S. *Alone together: Why we expect more from technology and less from each other*. UK: Hachette; 2017.

70. Yamaguchi H. 'Intimate relationship' with 'virtual humans' and the 'socialification' of familyship. SSRN J. 2018;3213799. <https://doi.org/10.2139/ssrn.3213799>.
71. Coeckelbergh M. Artificial companions: Empathy and vulnerability mirroring in human-robot relations. *Studies in ethics, law, and technology*. 2011;4(2). <https://doi.org/10.2202/1941-6008.1126>.
72. Liberati N, Nagataki S. Vulnerability under the gaze of robots: Relations among humans and robots. *AI & Soc*. 2019;34:333–42. <https://doi.org/10.1007/s00146-018-0849-1>.
73. Wilson JR, Scheutz M, Briggs G. Reflections on the design challenges prompted by affect-aware socially assistive robots. In: Tkalčić M, De Carolis B, de Gemmis M, Odić A, Košir A, editors. *Emotions and Personality in Personalized Services*. Human-Computer Interaction Series. Cham: Springer; 2016. p. 377–395. https://doi.org/10.1007/978-3-319-31413-6_18.
74. Feil-Seifer D, Mataric MJ. Socially Assistive Robotics. *IEEE Robot Autom Mag*. 2011;18:24–31. <https://doi.org/10.1109/MRA.2010.940150>.
75. Koyama T. Ethical issues for social robots and the trust-based approach. In: 2016 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO): 8.7.2016-10.7.2016. Shanghai, China: IEEE; 2016. p. 1–5. <https://doi.org/10.1109/ARSO.2016.7736246>.
76. Van Woensel L, Kurrerand C, Mihalis K, Kelly B, Boucher P, McCormack S, Manirambona R. *Ethical Aspects of Cyber-Physical Systems*. 2016. European Parliament: Scientific Foresight Unit. <https://doi.org/10.2861/68949>
77. Lin P, Abney K, Bekey GA. *Robot Ethics: The Ethical and Social Implications of Robotics*. Cambridge, MA: MIT Press; 2011.
78. Bartneck C, Lütge C, Wagner A, Welsh S. *Ethik in KI und Robotik, [Ethics in AI and Robotics]*. Carl Hanser Verlag GmbH Co KG: München; 2019.
79. Fraser NM, Gilbert GN. Simulating speech systems. *Comput Speech Lang*. 1991;5:81–99. [https://doi.org/10.1016/0885-2308\(91\)90019-M](https://doi.org/10.1016/0885-2308(91)90019-M).
80. Riek LD, Watson RN. The age of avatar realism. *IEEE Robot Autom Mag*. 2010;17:37–42. <https://doi.org/10.1109/MRA.2010.938841>.
81. Miller KW. It's not nice to fool humans. *IT professional*. 2010;12: 51–2. <https://doi.org/10.1109/MITP.2010.32>.
82. Punochojit L, Hongwarittorn N. Research ethics in human-computer interaction: A review of ethical concerns in the past five years. In: 2015 2nd National Foundation for Science and Technology Development Conference on Information and Computer Science (NICS); 16.9.2015-18.9.2015. Ho Chi Minh City, Vietnam: IEEE; 28.10.2015. pp. 180–185. <https://doi.org/10.1109/NICS.2015.7302187>.
83. Loh J. *Maschinenethik und Roboterethik, [Machine ethics and robot ethics]*. In: Bendel O, editor. *Handbuch Maschinenethik, [Handbook of Machine Ethics]*. Wiesbaden: Springer; 2019. p. 75–93. https://doi.org/10.1007/978-3-658-17483-5_6.
84. Gunkel DJ. *Robot Rights*. Cambridge, MA: MIT Press; 2018.
85. Sparrow R. Virtue and vice in our relationships with robots: Is there an asymmetry and how might it be explained? *Int J Soc Robot*. 2020;1–7. <https://doi.org/10.1007/s12369-020-00631-2>.
86. Wareham C. On the moral equality of artificial agents. *Int J Technoethics*. 2011;2:35–42. <https://doi.org/10.4018/IJT.2011010103>.
87. Wendt J. *Roboter-Ethik: Brauchen wir Roboterschutz-Gesetze?, [Robot-ethics: Do we need robot protection laws?]*. In: *Die Zeit*. 2013. <https://www.zeit.de/digital/internet/2013-05/roboter-ethik-kate-darling>. Accessed 24 March 2020.
88. Whitby B. Sometimes it's hard to be a robot: A call for action on the ethics of abusing artificial agents. *Interact Comput*. 2008;20: 326–33. <https://doi.org/10.1016/j.intcom.2008.02.002>.
89. Andreas M. *Lebenskritische Entscheidungen in der Roboterethik, [Autonomous Lethality. Vital decisions in robot ethics]*. In: Andreas M, Kasprovicz D, Rieger S, editors. *Unterwachen und Schlafen. Anthropophile Medien nach dem Interface, [Underwaking and sleeping. Antothrophile media after the interface]*. Lüneburg: meson press; 2018. p. 135–57. <https://doi.org/10.25969/mediarep/1278>.
90. Sparrow R. Killer robots. *J Appl Philos*. 2007;24:62–77. <https://doi.org/10.1111/j.1468-5930.2007.00346.x>.
91. Sparrow R. Killer robots: Ethical issues in the design of unmanned systems for military applications. In: Valavanis KP, Vachtsevanos GJ, editors. *Handbook of unmanned aerial vehicles*. Dordrecht: Springer; 2015a. p. 2965–83. https://doi.org/10.1007/978-90-481-9707-1_98.
92. Sparrow R. Twenty seconds to comply: Autonomous weapon systems and the recognition of surrender. *Int Law Stud*. 2015b;91:1–31.
93. Sparrow R, McLaughlin R, Howard M. Naval robots and rescue. *Int Rev Red Cross*. 2017;99:1139–59. <https://doi.org/10.1017/S181638311800067X>.
94. Sparrow R, Lucas G. When robots rule the waves? *Naval War College Review*. 2018;69:49–78.
95. Sparrow R. Robots, rape, and representation. *Int J Soc Robot*. 2017;9:465–77. <https://doi.org/10.1007/s12369-017-0413-z>.
96. Steil J, Finas D, Beck S, Manzeschke A, Haux R. Robotic systems in operating theaters: New forms of team-machine interaction in health care. *Methods Inf Med*. 2019;58:e14–25. <https://doi.org/10.1055/s-0039-1692465>.
97. Ienca M, Jotterand F, Vičá C, Elger B. Social and assistive robotics in dementia care: Ethical recommendations for research and practice. *Int J Soc Robot*. 2016;8:565–73. <https://doi.org/10.1007/s12369-016-0366-7>.
98. Nestorov N, Stone E, Lehane P, Eibrand R. Aspects of socially assistive robots design for dementia care. In: 2014 IEEE 27th International Symposium on Computer-Based Medical Systems; 2014. p. 396–400. <https://doi.org/10.1109/CBMS.2014.16>.
99. So I. Cognitive development in children: Piaget development and learning. *J Res Sci Teach*. 1964;2:176–86.
100. Kittmann R, Fröhlich T, Schäfer J, Reiser U, Weißhardt F, Haug A. Let me introduce myself: I am Care-O-bot 4, a gentleman robot. In: Diefenbach S, Henze, Pietot M, editors. *Mensch und Computer 2015 – Proceedings*. Berlin: De Gruyter Oldenbourg; 2015. p. 223–32. <https://doi.org/10.1515/9783110443929-024>.
101. Huisman C, Kort H. Two-year use of care robot Zora in Dutch nursing homes: An evaluation study. *Healthcare*. 2019;7:31–46. <https://doi.org/10.3390/healthcare7010031>.
102. Virk GS. Personal care robot safety. In: Fujimoto H, Tokhi MO, Mochiyama H, Virk GS, editors. *Emerging trends in mobile robotics. Proceedings of the 13th International Conference on Climbing and Walking Robots and the Support Technologies for Mobile Machines*. Nagoya Institute of Technology, Japan, 31 August – 3 September 2010. WORLD SCIENTIFIC; 2010. pp. 1332–1339. https://doi.org/10.1142/9789814329927_0162.
103. Manzeschke A. *Roboter in der Pflege, [Robots in care]*. *EthikJournal* 2019;5:1–11.
104. Körtner T. Ethische Herausforderungen zum Einsatz sozial-assistiver Roboter bei älteren Menschen, [Ethical challenges for the use of social-assistive robots for older people]. *Z Gerontol Geriatr*. 2016;49:303–7. <https://doi.org/10.1007/s00391-016-1066-5>.
105. Shibata T, Kawaguchi Y, Wada K. Investigation on people living with seal robot at home. *Int J Soc Robot*. 2012;4:53–63. <https://doi.org/10.1109/ROMAN.2010.5598704>.
106. Zwijsen SA, Niemeijer AR, Hertogh CPM. Ethics of using assistive technology in the care for community-dwelling elderly people: An overview of the literature. *Aging Ment Health*.

- 2011;15:419–27. <https://doi.org/10.1080/13607863.2010.543662>.
107. Manzeschke A, Weber K, Rother E, Fangerau H. Ethische Fragen im Bereich Altersgerechter Assistenzsysteme, [Ethical questions in the area of age appropriate assisting systems]. German Federal Ministry of Education and Research. 2015. https://www.researchgate.net/profile/Karsten_Weber/publication/304743219_Ethical_questions_in_the_area_of_age_appropriate_assisting_systems/links/5778da7808ae1b18a7e5f6b3/Ethical-questions-in-the-area-of-age-appropriate-assisting-systems.pdf. Accessed 12 May 2020.
 108. Forlizzi J, DiSalvo C. Service robots in the domestic environment: A study of the roomba vacuum in the home. In: Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-robot Interaction (HRI). 2006; 258–265. <https://doi.org/10.1145/1121241.1121286>.
 109. Riek LD, Rabinowitch TC, Chakrabarti B, Robinson P. Empathizing with robots: Fellow feeling along the anthropomorphic spectrum. In: 3rd International Conference on Affective Computing and Intelligent Interaction (ACII); 10.9-2009-12.9.2009. Amsterdam, Netherlands: IEEE; 2009. p. 1–6. <https://doi.org/10.1109/ACII.2009.5349423>.
 110. Scheutz M. The inherent dangers of unidirectional emotional bonds between humans and social robots. In: Lin P, Bekey GA, Abney K, editors. Robot ethics: The ethical and social implications of robotics. Cambridge, MA: MIT Press; 2011.
 111. Carpenter J. The quiet professional: An investigation of US military explosive ordnance disposal personnel interactions with everyday field robots. Doctoral dissertation. University of Washington; 2013.
 112. Niemelä M, Aerschot L, Tammela A, Aaltonen I, Lammi H. Towards ethical guidelines of using telepresence robots in residential care. *Int J Soc Robot*. 2019;1–9. <https://doi.org/10.1007/s12369-019-00529-8>.
 113. Oliveira R, Arriaga P, Paiva A. Ethical issues and practical considerations in the use of teleoperated robots as social interfaces. In: Human-Robot Interaction; Workshop: The dark side of human-robot interaction: Ethical considerations and community guidelines for the field of HRI; 11.3.2019. Daegu, South Korea: HRI; 2019. p. 1–5.
 114. Reijers W, Wright D, Brey P, Weber K, Rodrigues RO, Sullivan D, et al. Methods for practising ethics in research and innovation: A literature review, critical analysis and recommendations. *Sci Eng Ethics*. 2018;24:1437–81. <https://doi.org/10.1007/s11948-017-9961-8>. **This paper systematically reviews literature on methods to practice research in research and innovation in different fields, classifying the methods into ex ante, intra and ex post methods.**
 115. Veruggio G. The EURON Roboethics Roadmap. In: 2006 6th IEEE-RAS International Conference on Humanoid Robots; 4.12.2006-6.12.2006. Genova, Italy: IEEE; 2007. p. 612–7. <https://doi.org/10.1109/ICHR.2006.321337>.
 116. Grinbaum A, Chatila R, Devillers L, Ganascia JG, Tessier C, Dauchet M. Ethics in robotics research: CERN mission and context. *IEEE Robot Autom Mag*. 2017;24:139–45. <https://doi.org/10.1109/MRA.2016.2611586>.
 117. Kornwachs K. Smart robots – smart ethics? *Datenschutz und Datensicherheit*. 2019;43:332–41. <https://doi.org/10.1007/s11623-019-1118-2>.
 118. Kitchener KS, Anderson SK. Foundations of ethical practice, research, and teaching in psychology and counseling. 2nd ed. New York: Routledge; 2011.
 119. Cook A. Ethical issues related to the use/non-use of assistive technologies. *Dev Disabil Bull*. 2009;37:127–52.
 120. Misselhorn C, Pompe U, Stapleton M. Ethical considerations regarding the use of social robots in the fourth age. *GeroPsych*. 2013;26:121–33. <https://doi.org/10.1024/1662-9647/a000088>.
 121. Vandemeulebroucke T, Dierckx de Casterlé B, Gastmans C. The use of care robots in aged care: A systematic review of argument-based ethics literature. *Arch Gerontol Geriatr*. 2018;74:15–25. <https://doi.org/10.1016/j.archger.2017.08.014>.
 122. Mansouri N, Goher K, Hosseini SE. Ethical framework of assistive devices: Review and reflection. *Robo Biomimetics*. 2017;4: 19. <https://doi.org/10.1186/s40638-017-0074-2>.
 123. Huber A, Weiss A, Rauhala M. The ethical risk of attachment how to identify, investigate and predict potential ethical risks in the development of social companion robots. In: 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI); 7.3.2016-10.3.2016. Christchurch, New Zealand: IEEE; 2016. p. 367–74. <https://doi.org/10.1109/HRI.2016.7451774>.
 124. Weber K. Methoden der ethischen Evaluation von IT, [Methods for the ethical evaluation of IT]. In: Draude C, Lange M, Sick B, editors. Informatik 2019 Workshops, Lecture Notes in Informatics (LNI). Bonn: Gesellschaft für Informatik; 2019. p. 431–44.
 125. Manzeschke A. MEESTAR: ein Modell angewandter Ethik im Bereich assistiver Technologien, [MEESTAR: a model of applied ethics for assistive technologies]. In: Weber K, Frommeld D, Manzeschke A, Fangerau H, editors. Technisierung des Alltags – Beitrag für ein gutes Leben?, [Mechanization of everyday life - contribution to a good life?]. Stuttgart: Franz Steiner Verlag; 2015. p. 263–8.
 126. Weber K. Demografie, Technik, Ethik: Methoden der normativen Gestaltung technisch gestützter Pflege, [Demographics, technology, ethics: Methods for the normative design of technology-assisted care]. *Pflege & Gesellschaft*. 2017;22:338–52.
 127. Scorna U, Weber K, Haug SELSI. in serious games für die technikunterstützte medizinische Ausbildung. Das Beispiel HaptiVisT, [ELSI in serious games for technologically assisted medical training. The example HaptiVisT]. In: Weidner R, Karafillidis A, editors. Technische Unterstützungssysteme, die die Menschen wirklich wollen. Dritte Transdisziplinäre Konferenz, [Technological assistance systems humans really want. Third Transdisciplinary Conference]. Hamburg: Helmut-Schmidt-Universität; 2019. p. 187–94.
 128. Online ethics canvas. ADAPT Centre & Trinity College Dublin & Dublin City University. 2017. <https://www.ethicscanvas.org>. Accessed 12 May 2020.
 129. Battistuzzi L, Sgorbissa A, Papadopoulos C, Papadopoulos I, Koulouglioti C. Embedding ethics in the design of culturally competent socially assistive robots. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); 1.10.2018-5.10.2018. Madrid, Spain: IEEE; 2019. p. 1996–2001. <https://doi.org/10.1109/IROS.2018.8594361>.
 130. Bremner P, Dennis LA, Fisher M, Winfield AF. On proactive, transparent, and verifiable ethical reasoning for robots. *Proc IEEE*. 2019;107:541–61. <https://doi.org/10.1109/JPROC.2019.2898267>.
 131. Headleand CJ, Teahan W. Towards ethical robots: Revisiting Braitenberg’s vehicles. In: 2016 SAI Computing Conference (SAI); 13.7.2016-15.7.2016. London, UK: IEEE; 2016. p. 469–77. <https://doi.org/10.1109/SAI.2016.7556023>.
 132. Lindner F, Bentzen MM, Nebel B. The HERA approach to morally competent robots. In: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); 24.9.2017-28.9.2017. Vancouver, BC, Canada: IEEE; 2017. p. 6991–7. <https://doi.org/10.1109/IROS.2017.8206625>.
 133. Malhotra C, Kotwal V, Dalal S. Ethical framework for machine learning. In: 2018 ITU Kaleidoscope: Machine Learning for a 5G

- Future (ITU K); 26.11.2018-28.11.2018. Santa Fe, Argentina: IEEE; 2018. p. 1–8. <https://doi.org/10.23919/ITU-WT.2018.8597767>.
134. Pereira LM, Saptawijaya A. Programming machine ethics. Cham: Springer; 2016.
135. Sandewall E. Ethics, human rights, the intelligent robot, and its subsystem for moral beliefs. *Int J Soc Robot*. 2019. <https://doi.org/10.1007/s12369-019-00540-z>.
136. Shim J, Arkin R, Pettinatti M. An intervening ethical governor for a robot mediator in patient-caregiver relationship: Implementation and evaluation. In: 2017 IEEE International Conference on Robotics and Automation (ICRA); 29.5.2017-3.6.2017. Singapore, Singapore: IEEE; 2017. p. 2936–42. <https://doi.org/10.1109/ICRA.2017.7989340>.
137. Vanderelst D, Winfield A. An architecture for ethical robots inspired by the simulation theory of cognition. *Cogn Syst Res*. 2018;48:56–66. <https://doi.org/10.1016/j.cogsys.2017.04.002>.
138. McBride N, Hoffman RR. Bridging the Ethical Gap: From Human Principles to Robot Instructions. *IEEE Intell Syst*. 2016;31:76–82. <https://doi.org/10.1109/MIS.2016.87>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.