Why should we pay attention to gender and race in robot design?

Kamil Mamak
RADAR: Robophilosophy, AI Ethics, and Datafication
University of Helsinki
Helsinki, Finland
kamil.mamak@helsinki.fi

Pekka Mäkelä
RADAR: Robophilosophy, AI Ethics, and Datafication
University of Helsinki
Helsinki, Finland
pekka.a.makela@helsinki.fi

Raul Hakli
RADAR: Robophilosophy, AI Ethics, and Datafication
University of Helsinki
Helsinki, Finland
raul.hakli@helsinki.fi

***Abstract*— People have a tendency to assign such characteristics as race and gender to robots. These characteristics are not politically neutral, they relate to a variety of issues of significant societal importance such as (in)equality, diversity, and power relations. In this paper we discuss some of the potential harms related to ignoring these characteristics in the design of robots and argue for means to contribute to the maintenance of long-term cultural sustainability. The contribution is to be unfolded by way of enhancing institutional reasons sensitivity related to the issues of race and gender in robot design and implementation and safeguarding against the risks through obligatory institutional arrangements. We promote the use of value sensitive instruments in the context of publicly funded institutions. The design decisions that are made about healthcare robots ought to be a matter of public concern. As to our examples, we focus on robots in healthcare which are seen as partial solutions to the problems of utmost societal and economic importance in healthcare institutions of modern western democracies.**

Keywords— healthcare, robots, design, diversity, race, gender

I. Introduction

Robots are entering multiple spheres of life [1]. Healthcare is a sphere where robots are seen as a promising part of the solution to at least some problems of aging societies [2]. Prolonged life on the one hand, and the shortage of professional healthcare workers, on the other, may create a vacuum which at least in theory can be filled by robots. The deployment of robots in healthcare raises many ethical, societal, and legal issues. Just to mention a few, there is a problem of responsibility for potential harm caused by them [3], and more profoundly we can question whether substituting human care with robots is ethically acceptable at all [4]. In this paper, we want to focus on questions and ethical risks that may ensue from unreflective design choices concerning robot appearance, e.g., such choices that inspire an interpretation of a robot in terms of gender or race. Gender and race issues figure importantly in the discussion around deployment of AI in healthcare [5], [6]. Ethical questions and problems related to gender and race equality are obviously even more immanent and striking in the case embodied robots than in the case of AI [7], [8].

If the matter of race and gender in robot design were to be ignored, it might lead to untoward consequences such as less diverse society or strengthened healthcare giver stereotypes (as to their height, skin color, gender, etc.), which could be harmful to the representatives of disadvantageous social groups or the society in general.

According to the Collingridge dilemma [9], [10], it is relatively easy to control technology when it is at the early stage of development – when we do not know possible consequences yet. Whereas when implications are already manifested in society, it is usually harder to change technology substantially. Approaches like value-sensitive design [11], [12]and integrative social robotics [13] are promising ways to tackle the dilemma. They aim at incorporating human values into the design process of technologies. On the practical level, these approaches lead to furnishing design teams with multidisciplinary expertise to bring in multiple viewpoints and improve estimations of the possible consequences of technology use. They also offer conceptual tools and principles to assess various design choices in light of society's values and the aims of the institutions where the technology is to be introduced. In our view the value-sensitive approaches, broadly understood, provide us with societally and institutionally very promising early reaction response to Collingridge dilemma.

This paper focuses on aspects of healthcare robot design. It appears to be true that how our robots look, sound and behave is a big part of how we choose to use them [14, p. 71] Robot design choices in general, and race and gender concerns in particular, might well have an impact on the reception of robots in several sectors over and above healthcare for instance in children's education. Thus, we may be running a bit of a risk of wasting other people's time by way of ethical proliferation [12]. However, we wish to concentrate on healthcare robots for two reasons: First, healthcare robots appear to be a more significant social issue than robots considered in other application sectors (from the perspective of the demand and the quantity). Take for an example the attention sex robots are receiving which is not supported by the figures of their generality at all [16]. The second reason is that in many countries health-care systems are at least largely public, and if there is public money involved, there typically is a (justified) desire to monitor that society's values are complied with. As a result, this topic of gender and race in healthcare robots may, and perhaps should, become a source of public concern. This consideration will be less essential, for example, in the case of sex robots, the design of which is motivated precisely by a certain kind of physical appearance. This, however, is not to ignore the fact that sex robots and sex business more generally are a thoroughly political topic and a very important at that (for an interesting discussion of this see [17]–[20]).

This paper is structured as follows. After introductory remarks, Section II discusses the political nature of race and gender in healthcare robot design. Section III contains a discussion of the risks that may arise if we neglect

considerations of race and gender in robot design. Section IV considers possible ways to mitigate the identified risks, with a focus on solutions at the institutional level. The paper ends up with conclusions in Section V.

Before proceeding further, we would like to make a terminological clarification. The term “robot” is ambiguous. One of the commonly referred ways of explaining what constitutes a robot is the sense-think-act paradigm see, eg. , [21], [22], which requires that a robot has sensors to get information from the outside world, can process that information and analyze the data collected through the sensors, and is able to manipulate features of the outside world. This paradigm covers both the embodied and disembodied agents. The outside world may be some virtual world as well, like the bot that "discusses" on Twitter. In this paper we talk about “health care robots” and their race and gender, the focus here is on the bodies of robots, thus the robots under discussion are embodied. For some researchers, the embodiment is an element of the definition of robots; Winfield, among others, defines robots as “an AI with physical body” [23, p. 8]. To sum up, in our paper, we talk about robots with physical bodies that are used in the healthcare context.

Perhaps at this stage it is appropriate to make another clarificatory note as well. In this work it is not our intention to adopt any kind of moralizing position, rather our aim is to steer clear of substantial commitments to any specific values. Instead, our aim is to point at some issues that may go unnoticed in the design of healthcare robots even within societies that are sensitive to gender and race questions in general. In our view, paying attention to these issues already in the design phase is one important way, independent of any particular value system or structure, to safeguard against unnecessary and untoward societal consequences. Furthermore, it should also be pointed out that we are not debating race or gender categorizations per se [24]; rather, we are attempting to build on the rather generic observation that people tend to attach gender and race classifications to robots and draw some conclusions from the discussion of the phenomenon in the context of healthcare robots.

II. Politics in the Design of Healthcare Robots

The development of AI and robots has ethical and political dimensions [25], [26]. In this paper, we focus on the appearance of healthcare robots, which can involve ethical and political problems. This is in part due to the human tendency to anthropomorphize, that is, human propensity to look for human-like traits in non-human objects and events [27]. Research shows that together with many other human-like qualities, people attribute also sex and race to robots [28]–[33]. Neither race nor gender is politically neutral, therefore gendered and racialized robots have politics embedded in them.

In his seminal paper titled "Do Artifacts Have Politics?", L. Winner argues that technologies and tools are not politically neutral [34]. An artifact may have an embedded agenda that can intentionally or unintentionally shape political reality. Robert Sparrow, who in his article “How Robots have Politics” is asking a similar question, also affirms Winner's point. In addition, Sparrow argues that robots potentially instantiate, express or embed more politics than other artifacts [35]. This is because robots combine the intricacy of computers with the materiality of tools, thus they have and provide more affordances than other artifacts. Sparrow focusses also on the representation as an additional argument for the claim that robots engage more politics than typical artefacts. He points out that robots will often have a representational content, that is, they look like something, and this resemblance communicates potentially political ideas. He gives an example of sex robots: “Sex robots will convey ideas about how women look, or should look, and also ideas about how women should behave.” (On the issue of representation of sex robots see also [14], [32, p. 7]). A similar question can and perhaps should be asked with respect to healthcare robots. Will healthcare robots convey a message concerning how competent healthcare employees should look like? The appearance of robots can be viewed as a concentration of our attitudes and understanding concerning social roles, people that fulfill social roles, and what kind of people should fulfill social roles in question. So understood the appearance of robots indeed is a political issue.

A good many political bodies share the goal of decreasing inequalities based on many grounds, including race and gender. In 2015, United Nations Member States enacted the document titled "The 2030 Agenda for Sustainable Development" [36], in which participating countries committed themselves to transforming the world to achieve the goals and targets enlisted in the document. Among many other goals, gender equality (goal 5) and reduction of inequality within and between countries (goal 10), which also concerns the racial issues, are mentioned as important aims. In his book *AI for the Sustainable Development Goals,* Henrik Skaug Sætra analyzes how the rise of AI technologies may impact the UN's Sustainable Development goals [37], see also [38]. For example, the goal of decreasing inequalities between and within countries may be, as well as almost any other goal, positively or negatively impacted by AI. As Sætra points out, health is a crucial component of sustainability and the most important determinant of the substantial well-being of individual human beings [37, p. 56]. Hence, also the design of healthcare robots has a potential effect and contribution to the achievement of the Sustainable Development Goals.

To sum up, robots, including healthcare ones, bring a baggage of ideas with their physicality. This baggage includes the representational component as part of it. More specifically, robots may be perceived as entities having a race and gender. Such perceptions are not irrelevant from the perspective of political agenda. The look and appearance of future healthcare robots may have an impact, positive or negative, on the societal development of (in)equalities related to race and gender

III. Risks Associated with the Choices of Design of Healthcare Robots

Risks related to technology can be discussed from a number of different temporal perspectives — short-term, mid-term, and long-term [39]. Our paper is focused on the long-term perspective concerning the inequalities related to representation of gender and race in robots. Such inequality is by no means the only long-term worry concerning social robots. For example, Mark Coeckelbergh, in his article “Are emotional robots deceptive”, points out that love robots, which might replace human lovers, could have a negative impact on human-human relations and even on the human

willingness to seek for meaningful relations with other fellow humans [40]. Henrik Skaug Sætra, in his article "Social robot deception and the culture of trust", discusses deceptive robots from the long-term perspective, and raises worries about the impact of such robots on the human culture of trust among members of a society. Sætra argues that mutual trust may be degraded by deceiving robots becoming part of our social life [41]. A worrying scenario with respect to and about the deceptive design of social robots has been presented by Aimee van Wynsberghe. She visions that designing deceptive social robots for reciprocal interaction with humans threatens the human ability and willingness to reciprocate to fellow humans. Furthermore, she believes that the ensuing lack of reciprocity in our social life could risk our socially sustainable future [42]. In the discussion below we point out some negative consequences that the lack of diversity in healthcare robots may bring about. You may find some overlap between the issues discussed and you may even think that there are contradictory elements in the presentation. Here we do not discuss these issues in relation to one another however. Instead, our aim in discussing them separately is to bring some topical specific aspects and issues to the fore.

A. Partial representation of the members of society

It seems rather safe to say that people attribute race and gender to robots on the basis of the conceptual apparatus with which they operate when attributing race and gender to their fellow humans as well. People also identify themselves through the employed conceptual apparatus. This seems to put quite a bit of pressure on the design of healthcare robots, as there is a significant amount of research evidence supporting the claim that people prefer to interact with robots that look like the members of their own social group [43], [44]. This is problematic from the standpoint of diverse societies because in practice healthcare robots can only represent a portion of society. Here there is a risk that some patients do not feel that the healthcare robots represent their specific social group and make the inference that the healthcare system is not designed for them. If society cannot afford to provide such healthcare robots, which make all patients comfortable in the sense mentioned above, the healthcare systems facilities may be perceived as hostile toward some group of people. This may lead to self-exclusion from the healthcare system with devastating consequences. As to a historical analogue, in the back of history, many aboriginals in Australia felt that Australian healthcare system was, as a matter of fact, a part of the assimilation project of the state, and they refused to use the services of the healthcare system with serious consequences in the health predicament in the aboriginal population [45].

B. Creating a specific role model for the healthcare professionals

Robots in principle may, through their design, represent a broad spectrum of human diversity, and then again due to a variety of practical and financial constraints robots may be designed to only represent "the representative individual" determined by the stereotypical thinking and conceptualization of the majority "in power". Design choices in robots may explicitly or implicitly represent what is normal in a social context [46]. The standardized model of a healthcare robot may represent only a few or a single group. If medical/healthcare robots represent a specific type of human, it may have an impact on how people think, characterize, and conceptualize a healthcare professional. It may be expected that a "typical" representative of the medical profession looks and behaves in a certain way, and high competency and reliability will be associated with such properties. As a result, a person (or a robot) that looks different from "the ideal" may be perceived as someone who does not qualify, someone who is not good enough. This may be harmful to existing professionals and also to prospective ones, who may not choose the career path because of their fear of falling short of the professional standard set by idiosyncratic representational robot design choices. It is a well-known fact that one way of improving the state of equality in a healthcare system is to increase workforce diversity [47]. Additionally, in the process towards a better state of equality it is important to constantly keep an eye on and to recognize new and emerging disparities [47].

C. Deskilling of confronting otherness

A serious problem potentially following from basing the design of healthcare robots on subjective preferences of patients or customers to be is moral deskilling. This obviously needs some unfolding and elaboration. Shannon Vallor, whose work mainly focuses on virtue ethics in the context of new technologies [48], argues that technologies, including robots, may lead to moral deskilling and thus have a negative impact on the human character [49]. As already discussed above, many studies indicate that people prefer to interact with robots that appear to belong to their own social group [43], [44]. So, it is highly probable that people who will be asked to choose the properties of robots will choose qualities similar to those of their own peer group. In such cases, they will lose the chance to confront otherness. Sparrow points out that interactions with robots shape our behavior in multiple ways, and "the more we interact with robots, the more the nature of our activities and our relationships with other people will be shaped by the decisions of those who design the robots with which we interact." [35, p. 10] Such a claim is supported by empirical research that shows that robots, due to their physical presence, have more potential and means to teach humans than traditional learning technologies [50]. The way we treat a specific kind of people may transfer to the way we treat robots that represent such kind of people [32], but see [51] and the other way round. So, having the experience with robots representing other groups may lead to more openness to otherness in the case of fellow humans. If people are not exposed to otherness, there may be problems with approaching human beings that belong to unfamiliar social groups.

IV. How to mitigate possible risks

One may ask how we, as a society, should address the potential risks resulting from grounding the design of healthcare robots to patient preferences. As pointed out above, such practice could have untoward societal consequences. One obvious way to go here is not do anything and let the robot companies and their clients operate in the "morally free zone" as it were. However, considering the weight of the possible consequences, we would like to briefly discuss some of the imaginable actions that may mitigate the risks.

A. Restriction

Joanna Bryson has considered the undesirable consequences of over-attachment to robots that may result from attributing robots such meanings and expectations (humane features, being a friend) that robots cannot due to their ontological structure (machine-like entities) support or meet. She points out that robots are designed by us, humans, and we as creators have the power to build them in a way that could help us to avoid unnecessary harm, for instance by way of avoiding building robots with deceptive appearance [52]. The uniformity in the appearance of robots may contribute to harm at the societal level. In the case of healthcare robots, the issue of diversity is slightly complicated by the fact that in this context it is well justified that the preferences of care receivers get considered in the design, which of course may direct the design toward uniformity. It seems plausible that creating social robots that do not activate anthropomorphic tendencies is not feasible [53], [54]. However, we as creators of robots have power over their design, and as we are aware of the anthropomorphic tendencies, we can make such corrections to the dominant practices that minimize societal harm.

There are at least two ways of restricting the influence of patient preferences on the design that leads to the perceived gender and race of the robots. The first is to impose regulation quotas for different genders and ethnic groups. It may be relatively easy to enforce, for example, in public healthcare. It can be argued that patient preferences may be ignored to some extent in hospitals or other publicly funded healthcare institutions.

Let's imagine a hospital in which there are only female physicians and in which there are no robots. A misogynist patient with a medical problem comes into the hospital to get help but he does not want care from a female physician. In such a situation, it is rather safe to argue that the explicit preference of the patient should be ignored. If we accept this, it seems natural to move on to discussing whether, when and to what extent patients' preferences concerning perceived race and gender of a healthcare *robot* should or could also be ignored. A hospital operating in public domain and with taxpayers' money cannot make assumptions as to their potential patients' misogynist or racist orientation not to mention try to please or pamper such tendencies. A public hospital must operate on the assumption that it supports and aims to maintain the values of (liberal democratic) society. If we accept this, then the way is paved for discussing whether, when, and to what extent the patient preferences with respect to robot appearance can be trumped by such values as equality and focusing on the question of just and fair representation. It seems that this line of reasoning would not directly concern healthcare robots in home settings where the properties of the robots and their design are made in advance, considering the patient preferences. This would be due to the lack of public component.

Another way of mitigating the risks related to robots' gendered or racialized appearances is to create robots that do not have perceived race or gender. For instance, Rob Sparrow in the context of a discussion concerning the racial issues related to robots proposes that attributing race should be made difficult for receivers [8] However, also this may be problematic as taking care of the patient's well-being is a central part of the healthcare business, and it may well be that patients would like to have gendered and racialized robots, and "neutral" or machine-like robots would not fulfill their expectations, which in turn might have a detrimental effect on the patients' well-being. Another question is to what extent it is possible to create human-like social robots that will not inspire anthropocentric interpretation in terms of race and gender. One alternative would be to design robots that have mixed or fluid gender and race features. Merle Weßel et al. propose queering of robots, by which they mean designing robots that challenge the binary gender attribution and common stereotypes [7]. This approach would acknowledge the inevitability of attribution of racial and gender features, but would challenge the binary gender classification and common stereotypes, e.g., by mixing different racial and gender features, or by making them fluid. Obviously, this solution is not without its problems. For instance, it seems to be an empirical truth that many people operate with a binary gender system and entertain racial stereotypes, so queered robots might face resistance and cause psychological dissonance.

B. Nudges

Nudges provide us with one "manipulative" tool to be used in dealing with the problematic and societally important questions concerning how gender and race and respective inequalities get expressed and represented in robots. It is well-known that people making decisions are not perfectly rational. Indeed, our decision-making and our behavior more generally can be influenced by the choice architecture, that is, the contextual representation of the alternative options of action, the architecture makes us see the choice situation in a certain way which has an impact on the light our reasons shed on alternative choices [55]. It is pointed out that nudges can be helpful in the healthcare setting, however, this comes with the risk of undermining patient autonomy [56]. Assuming that nudging is compatible with patient autonomy, it could be employed in the design of healthcare robots, by offering robots with diverse racial and gender features by default. These features may be changed if the patients demand it, but that would demand extra steps. Cappuccio et al. believe that through the design of robots it is possible to impact the development of the moral character of the people who interact with them [57], [58].

The measures discussed above could and should (only) be applied after assuring that all the relevant reasons and arguments have been taken into account, and there has been enough time for critical discussion and deliberation with relevant stakeholders. Patient care is one of the ultimate goals of the healthcare system, and it is obvious that healthcare organizations have a crucial role in achieving that goal [59]. We realize that in the case of a strict and homogeneous society the core idea and arguments of this paper may not inspire much popularity but rather face resistance and counterarguments. Indeed, an army of healthcare robots representing diversity as to gender and race may come across as somewhat alien in such a society.

Another point worthy of emphasis is the context-sensitivity of our approach. There is no one size fits all solution here, which is the case with every societally problematic issue worth pondering of course. For instance, the issue related to the weight of patient preferences looks very different in the case of population consisting of dementia patients and in the case in which the target

population consists of school kids. By the same token, the exact content of the goal under discussion depends on the type of healthcare institution in question and the means of achieving the goals may depend on group-specific features.

C. *Institutional value-sensitive design instruments*

As to the third way of mitigating the possible risks related to the design and appearance of robots in health care institutions, we promote a ”meta regulation” approach according to which certain procedural requirements for decision-making concerning the design and implementation of robots in publicly funded democratic institutions should be established by law or other kind of democratically authorized and legitimized regulation. One reason for this is that there is no straightforward way to dictate whether design decisions, e.g., concerning the appearance of robots, should be based on subjective desires or preferences of individual people, or society-level values aiming at furthering people's objective interests. In our view these two ways are both eligible, and the choice between them is and should be highly context dependent.

As to the procedural requirements supporting decision making concerning the design and implementation in the context of publicly funded institutions we suggest methods that are in line with the general methodology of value-sensitive design [11], and in particular, its further developments in the field of care robotics [2], [60]. In some countries, the use of such tools, like the Data Ethics Decision Aid (DEDA) [61], [62] are mandatory in the implementation of algorithmic systems in public institutions. This is the case at least in some municipalities in the Netherlands. Such tools bring together a whole variety of stakeholders to discuss ethical and other aspects of the design and implementation and “force” them to see and reflect upon different factors and reasons that are relevant for the choices to be made about the design and implementation of, say, a robot into the practices of a hospital. Adoption of such tools is a procedural way to enhance and increase the institutional sensitivity to moral reasons, which is not ad hoc as it is part of the obligatory institutional decision-making structure. Such mechanisms allow for both taking into account a good variety of contextual factors and simultaneously weighing the importance of such factors in each and every context. Here we promote the adoption and use of these instruments, however, it is adequate to point out that as far as we know race and gender do not seem to be on the agenda of any of the existing tools. This is not an in-principle problem, though, and race and gender can easily be, and obviously should, be accommodated to the agenda of such tools. This we believe is a way towards sustainability related to issues concerning the appearance of robots. Such a self-reflective mechanism that allows the institutions to reflect upon their goals and values and to assess whether the current institutional arrangements are coherent with and conducive to them seems necessary for maintaining the recreational and reproductive capacity of the institutions and thereby contributes to their sustainability.

V. Conclusions

Decisions concerning the design of healthcare robots, even if consistent with the patients' individual preferences, may have long-term detrimental impact on delicate political issues, such as race and gender equality. Improving the state of equality with respect to both gender and race is high on the priority list of many international political bodies, it is also among the central political goals of many nation states all over the globe. In the brief discussion above we brought up some risks, or undesirable consequences, that unreflective healthcare robot design may cause. We paid attention to the potential societal worry that if the healthcare robot design were uniform and not sensitive to the diversity of the population, it would probably be received and perceived as unrepresentative. Here there is a risk that some patients do not feel that the healthcare robots represent their specific social group and make the inference that the healthcare system is not designed for them which ends up with the practical conclusion of self-exclusion. Another entry emphasizing the need for aware and savvy design of healthcare robots was about the possibility that the lack of diversity in the design of healthcare robots may lead to, or support and maintain, a narrow and stereotypical image of the competent and reliable medical professional. Setting up a restricted model and ideal may scare off many talented healthcare professionals to-be from such a career path, which is very bad news for many countries struggling with the shortage of healthcare professionals. A third potential worry that was discussed above had to do with the risk of deskilling of confronting otherness. Here the point bluntly was that if the design mechanism of healthcare robots allows people to stick to their own “bubbles” in their interactions with both people and robots, this may result in atrophy of people's capacities to face and confront otherness, the unfamiliar.

It seems clear that both intentional and unintentional choices made in robot design do have an impact on a variety of societal issues which are highly relevant to citizens' well-being. A part of the function of a piece like this is to contribute to awareness and as a result to increase the proportion of intentional choices as opposed to unintentional choices in public and institutional contexts, at least. We are painfully aware that we are discussing rather complex and difficult issues here: It is not easy to solve the related problems, there are conflicting ideas that all have their perspective-dependent pros and cons. Yet, however, we daringly discussed some constructive ideas as well. Our rather cautious and moderate take on the ways to mitigate the risks and potential harms consisted in such means as regulation/constraints, nudges, and “meta-regulation” enforcing value-sensitive design approach, broadly understood, in all public procurement, that is, design, production and implementation of robots in the context of public institutions in general and in healthcare institutions in particular.

We ended up mildly arguing for some procedural requirements supporting decision making concerning the design and implementation of robots in the context of publicly funded institutions. We suggested methods that are in line with the general methodology of value-sensitive design. We promoted such decision aid tools that bring together a whole variety of stakeholders to discuss ethical and other aspects of the design and implementation and “force” the participants to see and reflect upon different factors and reasons that are relevant for the choices to be made about the design and implementation. We also suggested extending these tools with explicit concern to race and gender issues, which currently seem to be lacking from them. We flagged for procedural ways to enhance and

increase the institutional sensitivity to moral reasons, which are not ad hoc but part of the obligatory institutional decision-making structure. Such mechanisms allow for both taking into account a good variety of contextual factors and simultaneously weighing the importance of such factors in each and every context. We argued for self-reflective mechanisms that allow the institutions to reflect upon their goals and values and to assess whether the current arrangements of the institution are coherent with and conducive to them. Furthermore, we claimed that such mechanisms are necessary for maintaining the recreational and reproductive capacity of institutions and thereby contribute to their viability and sustainability.

REFERENCES

[1] M. Ford, *Rise of the Robots: Technology and the Threat of a Jobless Future*, Illustrated edition. New York: Basic Books, 2016.

[2] E. Fosch Villaronga, *Robots, Healthcare, and the Law. Regulating Automation in Personal Care*. 2019. doi: 10.4324/9780429021930.

[3] A. Matthias, "The responsibility gap: Ascribing responsibility for the actions of learning automata," *Ethics Inf. Technol.*, vol. 6, no. 3, pp. 175–183, Sep. 2004, doi: 10.1007/s10676-004-3422-1.

[4] R. Sparrow and L. Sparrow, "In the hands of machines? The future of aged care," *Minds Mach.*, vol. 16, no. 2, pp. 141–161, Oct. 2006, doi: 10.1007/s11023-006-9030-6.

[5] E. Fosch-Villaronga, H. Drukarch, P. Khanna, T. Verhoef, and B. Custers, "Accounting for diversity in AI for medicine," *Comput. Law Secur. Rev.*, vol. 47, p. 105735, Nov. 2022, doi: 10.1016/j.clsr.2022.105735.

[6] E. Fosch-Villaronga and H. Drukarch, "Accounting for Diversity in Robot Design, Testbeds, and Safety Standardization," *Int. J. Soc. Robot.*, Mar. 2023, doi: 10.1007/s12369-023-00974-6.

[7] M. Weßel, N. Ellerich-Groppe, and M. Schweda, "Gender Stereotyping of Robotic Systems in Eldercare: An Exploratory Analysis of Ethical Problems and Possible Solutions," *Int. J. Soc. Robot.*, Dec. 2021, doi: 10.1007/s12369-021-00854-x.

[8] R. Sparrow, "Robotics Has a Race Problem," *Sci. Technol. Hum. Values*, vol. 45, no. 3, pp. 538–560, May 2020, doi: 10.1177/0162243919862862.

[9] D. Collingridge, *Social Control of Technology*, New edition. Milton Keynes: Open University Press, 1981.

[10] W. Liebert and J. C. Schmidt, "Collingridge's dilemma and technoscience," *Poiesis Prax.*, vol. 7, no. 1, pp. 55–71, Jun. 2010, doi: 10.1007/s10202-010-0078-2.

[11] B. Friedman and D. Hendry, *Value sensitive design: shaping technology with moral imagination*. Cambridge, Massachusetts: The MIT Press, 2019.

[12] *Value-sensitive Design*. Routledge, 2020, pp. 329–332. doi: 10.4324/9781003075011-23.

[13] J. Seibt, M. Flensborg Damholdt, and C. Vestergaard, "Integrative social robotics, value-driven design, and transdisciplinarity," *Interact. Stud. Soc. Behav. Commun. Biol. Artif. Syst.*, vol. 21, no. 1, pp. 111–144, 2020, doi: 10.1075/is.18061.sei.

[14] G. Kasparov, *Deep Thinking: Where Machine Intelligence Ends and Human Creativity Begins*. PublicAffairs, 2017.

[15] H. S. Sætra and J. Danaher, "To Each Technology Its Own Ethics: The Problem of Ethical Proliferation," *Philos. Technol.*, vol. 35, no. 4, p. 93, Oct. 2022, doi: 10.1007/s13347-022-00591-7.

[16] J. A. Oravec, "Love, Sex, and Robots: Technological Shaping of Intimate Relationships," in *Good Robot, Bad Robot: Dark and Creepy Sides of Robotics, Autonomous Vehicles, and AI*, J. A. Oravec, Ed., in Social and Cultural Studies of Robots and AI. Cham: Springer International Publishing, 2022, pp. 91–123. doi: 10.1007/978-3-031-14013-6_4.

[17] K. Richardson, "The Asymmetrical 'Relationship,'" *Acm Sigcas Comput. Soc.*, vol. 45, no. 3, pp. 290–293, 2015, doi: 10.1145/2874239.2874281.

[18] R. Sparrow, "Robots, Rape, and Representation," *Int. J. Soc. Robot.*, vol. 9, no. 4, pp. 465–477, Sep. 2017, doi: 10.1007/s12369-017-0413-z.

[19] J. Danaher, "Robotic Rape and Robotic Child Sexual Abuse: Should They be Criminalised?," *Crim. Law Philos.*, vol. 11, no. 1, pp. 71–95, Mar. 2017, doi: 10.1007/s11572-014-9362-x.

[20] J. Danaher, B. Earp, and A. Sandberg, "Should We Campaign Against Sex Robots?," in *Robot Sex: Social and Ethical Implications*, J. Danaher and N. McArthur, Eds., The MIT Press, 2017, p. 0. doi: 10.7551/mitpress/9780262036689.003.0004.

[21] J. M. Jordan, *Robots*, 1st Edition. Cambridge, MA: The MIT Press, 2016.

[22] D. J. Gunkel, *Robot Rights*. Cambridge, Massachusetts: The MIT Press, 2018.

[23] A. Winfield, *Robotics: A Very Short Introduction*. in Very Short Introductions. Oxford, New York: Oxford University Press, 2012.

[24] J. K. Malinowska and T. Żuradzki, "Towards the multileveled and processual conceptualisation of racialised individuals in biomedical research," *Synthese*, vol. 201, no. 1, p. 11, Dec. 2022, doi: 10.1007/s11229-022-04004-2.

[25] M. Coeckelbergh, *The Political Philosophy of AI: An Introduction*, 1st edition. Cambridge: Polity, 2022.

[26] M. Coeckelbergh, *AI Ethics*. Cambridge, MA: MIT Press, 2020. Accessed: Aug. 11, 2020. [Online]. Available: https://mitpress.mit.edu/books/ai-ethics

[27] S. E. Guthrie, "Anthropomorphism: A definition and a theory," in *Anthropomorphism, anecdotes, and animals*, R. W. Mitchell, N. S. Thompson, and H. L. Miles, Eds., in SUNY series in philosophy and biology. Albany, NY, US: State University of New York Press, 1997, pp. 50–58.

[28] J. Bernotat, F. A. Eyssel, and J. Sachse, "Shape it – The influence of robot body shape on gender perception in robots," *Soc. Robot. 9th Int. Conf. ICSR 2017 Tsukuba Jpn. Novemb. 22-24 2017 Proc.*, vol. 10652, 2017, Accessed: Oct. 02, 2022. [Online]. Available: https://pub.uni-bielefeld.de/record/2914079

[29] F. Eyssel and F. Hegel, "(S)he's Got the Look: Gender Stereotyping of Robots1," *J. Appl. Soc. Psychol.*, vol. 42, no. 9, pp. 2213–2230, 2012, doi: 10.1111/j.1559-1816.2012.00937.x.

[30] E. Roesler, L. Naendrup-Poell, D. Manzey, and L. Onnasch, "Why Context Matters: The Influence of Application Domain on Preferred Degree of Anthropomorphism and Gender Attribution in Human–Robot Interaction," *Int. J. Soc. Robot.*, vol. 14, no. 5, pp. 1155–1166, Jul. 2022, doi: 10.1007/s12369-021-00860-z.

[31] G. Perugia, S. Guidi, M. Bicchi, and O. Parlangeli, "The Shape of Our Bias: Perceived Age and Gender in the Humanoid Robots of the ABOT Database," in *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction*, in HRI '22. Sapporo, Hokkaido, Japan: IEEE Press, Mar. 2022, pp. 110–119.

[32] J. Louine, D. C. May, D. W. Carruth, C. L. Bethel, L. Strawderman, and J. M. Usher, "Are Black Robots Like Black People? Examining How Negative Stigmas about Race Are Applied to Colored Robots," *Sociol. Inq.*, vol. 88, no. 4, pp. 626–648, 2018, doi: 10.1111/soin.12230.

[33] C. Bartneck *et al.*, "Robots And Racism," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, Chicago IL USA: ACM, Feb. 2018, pp. 196–204. doi: 10.1145/3171221.3171260.

[34] L. Winner, "Do Artifacts Have Politics?," *Daedalus*, vol. 109, no. 1, pp. 121–136, 1980.

[35] R. Sparrow, "How Robots Have Politics," in *The Oxford Handbook of Digital Ethics*, C. Véliz, Ed., 2021. doi: 10.1093/oxfordhb/9780198857815.013.16.

[36] "The 2030 Agenda for Sustainable Development." United Natuons, 2015. Accessed: Oct. 02, 2022. [Online]. Available: https://sdgs.un.org/2030agenda

[37] H. S. Sætra, *AI for the Sustainable Development Goals*, 1st edition. CRC Press, 2022.

[38] H. S. Sætra, Ed., *Technology and Sustainable Development: The Promise and Pitfalls of Techno-Solutionism*. New York: Routledge, 2023.

[39] L. Royakkers and R. van Est, “A Literature Review on New Robotics: Automation from Love to War,” *Int. J. Soc. Robot.*, vol. 7, no. 5, pp. 549–570, Nov. 2015, doi: 10.1007/s12369-015-0295-x.

[40] M. Coeckelbergh, “Are Emotional Robots Deceptive?,” *IEEE Trans. Affect. Comput.*, vol. 3, no. 4, pp. 388–393, 2012, doi: 10.1109/T-AFFC.2011.29.

[41] H. S. Sætra, “Social robot deception and the culture of trust,” *Paladyn J. Behav. Robot.*, vol. 12, no. 1, pp. 276–286, Jan. 2021, doi: 10.1515/pjbr-2021-0021.

[42] A. van Wynsberghe, “Social robots and the risks to reciprocity,” *AI Soc.*, Apr. 2021, doi: 10.1007/s00146-021-01207-y.

[43] F. Eyssel and D. Kuchenbrandt, “Social categorization of social robots: Anthropomorphism as a function of robot group membership,” *Br. J. Soc. Psychol.*, vol. 51, no. 4, pp. 724–731, 2012, doi: 10.1111/j.2044-8309.2011.02082.x.

[44] D. Kuchenbrandt, F. Eyssel, S. Bobinger, and M. Neufeld, “When a Robot’s Group Membership Matters,” *Int. J. Soc. Robot.*, vol. 5, no. 3, pp. 409–417, Aug. 2013, doi: 10.1007/s12369-013-0197-8.

[45] S. Vallesi, L. Wood, L. Dimer, and M. Zada, “‘In Their Own Voice’—Incorporating Underlying Social Determinants into Aboriginal Health Promotion Programs,” *Int. J. Environ. Res. Public. Health*, vol. 15, no. 7, p. 1514, Jul. 2018, doi: 10.3390/ijerph15071514.

[46] H. S. Sætra, A. Nordahl-Hansen, E. Fosch-Villaronga, and C. Dahl, “Normativity assumptions in the design and application of social robots for autistic children,” *Scand. Conf. Health Inform.*, pp. 136–140, Aug. 2022, doi: 10.3384/ecp187023.

[47] J. S. Williams, R. J. Walker, and L. E. Egede, “Achieving Equity in an Evolving Healthcare System: Opportunities and Challenges,” *Am. J. Med. Sci.*, vol. 351, no. 1, pp. 33–43, Jan. 2016, doi: 10.1016/j.amjms.2015.10.012.

[48] S. Vallor, *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*, 1st edition. New York, NY: Oxford University Press, 2016.

[49] S. Vallor, “Moral Deskilling and Upskilling in a New Machine Age: Reflections on the Ambiguous Future of Character,” *Philos. Technol.*, vol. 28, no. 1, pp. 107–124, Mar. 2015, doi: 10.1007/s13347-014-0156-9.

[50] T. Belpaeme, J. Kennedy, A. Ramachandran, B. Scassellati, and F. Tanaka, “Social robots for education: A review,” *Sci. Robot.*, vol. 3, no. 21, p. eaat5954, Aug. 2018, doi: 10.1126/scirobotics.aat5954.

[51] J. Banks and K. Koban, “A Kind Apart: The Limited Application of Human Race and Sex Stereotypes to a Humanoid Social Robot,” *Int. J. Soc. Robot.*, Jul. 2022, doi: 10.1007/s12369-022-00900-2.

[52] J. J. Bryson, “Patiency is not a virtue: the design of intelligent systems and systems of ethics,” *Ethics Inf. Technol.*, vol. 20, no. 1, pp. 15–26, Mar. 2018, doi: 10.1007/s10676-018-9448-6.

[53] D. J. Gunkel, “The other question: can and should robots have rights?,” *Ethics Inf. Technol.*, vol. 20, no. 2, pp. 87–99, Jun. 2018, doi: 10.1007/s10676-017-9442-4.

[54] K. Mamak, “Whether to save a robot or a human: On the ethical and legal limits of protections for robots,” *Front. Robot. AI*, vol. 8, 2021, doi: 10.3389/frobt.2021.712427.

[55] R. H. Thaler and C. R. Sunstein, *Nudge: Improving Decisions About Health, Wealth, and Happiness*, Illustrated edition. New Haven: Yale University Press, 2008.

[56] S. Z. Raskoff, “Nudges and hard choices,” *Bioethics*, vol. n/a, no. n/a, 2022, doi: 10.1111/bioe.13091.

[57] M. L. Cappuccio, E. B. Sandoval, O. Mubin, M. Obaid, and M. Velonaki, “Robotics Aids for Character Building: More than Just Another Enabling Condition,” *Int. J. Soc. Robot.*, vol. 13, no. 1, pp. 1–5, Feb. 2021, doi: 10.1007/s12369-021-00756-y.

[58] M. L. Cappuccio, E. B. Sandoval, O. Mubin, M. Obaid, and M. Velonaki, “Can Robots Make us Better Humans?,” *Int. J. Soc. Robot.*, vol. 13, no. 1, pp. 7–22, Feb. 2021, doi: 10.1007/s12369-020-00700-6.

[59] A. E. Mills and E. M. Spencer, “Values Based Decision Making: A Tool for Achieving the Goals of Healthcare,” *HEC Forum*, vol. 17, no. 1, pp. 18–32, Mar. 2005, doi: 10.1007/s10730-005-4948-2.

[60] A. van Wynsberghe, *Healthcare Robots: Ethics, Design and Implementation*. Abingdon, Oxon: Routledge, 2016.

[61] I. Muis, J. Straatman, R. Fransen, and S. Hemerik, “DEDA: Data Ethics Decision Aid.”

[62] A. S. Franzke, I. Muis, and M. T. Schäfer, “Data Ethics Decision Aid (DEDA): a dialogical framework for ethical inquiry of AI and data projects in the Netherlands,” *Ethics Inf. Technol.*, vol. 23, no. 3, pp. 551–567, Sep. 2021, doi: 10.1007/s10676-020-09577-5.