Ethical boundaries for android companion robots: a human perspective

Cindy Friedman[1][0000-0002-4901-9680]

[1] Department of Philosophy, University of Pretoria, Pretoria, South Africa
cindzfriedman@gmail.com

Abstract. Literature on machine ethics tends to position itself on a spectrum where at one end sits notions about whether machines can be *moral patients* and, at the other end, sits notions of whether machines can be *moral agents*. While majority of the literature concerns itself with the latter – for example see [1], [2], [3] – my paper fits into the debate by focusing on the capacity for android companion robots to be moral patients in relation to their capacity to be phenomenally conscious in the sense of *feeling what it is like* [4] to be a moral patient and be morally wronged. It questions whether these robots should be treated morally well by human interactors and ultimately, whether there should be ethical boundaries established for their utilisation in the form of granting robots negative rights. As such, it grapples with both the descriptive and normative aspects of the topic, which seems to often be amiss in current machine ethics literature – an issue that Gunkel [5] has raised.

I start my argument in favour of reflecting on the moral treatment of robots from a *human perspective*, by considering Turner's statement that "protection in law often follows shortly after society has recognised a moral case for protecting something" [6, p. 170] and critically analyse the concepts of there being a moral case to consider, as well as the need for protection, to formulate my argument for why we should establish ethical boundaries, but that we must, first and foremost, consider establishing ethical boundaries from a human perspective. Even though social robots of the kind we are considering here may not actually be conscious, human interactors may anthropomorphise them since they are, as Darling [7] and Scheutz [8] note, designed to elicit the tendency to anthropomorphise in human interactors. This creates the possibility for the formulation of 'real bonds' with these robots, despite robots not having the capacity to genuinely reciprocate human emotion – at least not currently, nor in the near future. The possibility for the formulation of 'real bonds', however, raises the question of whether treating them immorally may lower the moral standards of human interactors [9], [10]. Taking on a human perspective is thus a relational account in terms of the moral consideration of robots [11], [7], [12] since what matters is not whether robots are *actually* phenomenally conscious and can *actually* be moral patients who can feel what it is like to be morally wronged, but whether we view them as possessing the property of phenomenal consciousness due to the way in which we relate to them. Given this, and going back to Turner [6], if we take on a human perspective, there is a moral case for protecting something – this something being the human interactors – and the protection is from themselves. Thus, I argue that we must establish ethical boundaries but that the establishment of ethical boundaries must

firstly be considered from a human perspective so as to protect human interactors from their own potential immoral behaviour.

On the other hand, I suggest that in arguing for ethical boundaries from a *'robot perspective'*, one implies that there is a moral case to consider for the sake of robots, and that robots need protection from being treated immorally by human interactors – that they warrant moral consideration because they *can* be moral patients and can *feel what it is like* to be morally wronged. This is problematic. Firstly, if robots are not actually conscious and cannot then *feel what it is like* to be morally wronged, from what are we protecting them? Secondly, even if they were, or could one day be conscious, we must still approach the topic at hand from a human perspective first and foremost, because even if we cannot disprove the possibility of robot consciousness, or even if robots can mimic consciousness, our misdeeds towards them may still negatively impact our moral standards. Therefore, taking on such a 'robot perspective' would mean taking on a property account towards the moral consideration of robots [11] which pertains to basing moral consideration on whether robots possess particular properties, such as phenomenal consciousness in this instance. This account, however, is highly problematic because, firstly, which property do we take into consideration (consciousness, personhood, sentience)? And, secondly, it is difficult to prove that humans possess such properties, raising the question of how we could even begin to prove such properties in robots?

I then consider granting rights to robots. Given the overlap between morality and legality [13, p. 169] I suggest granting negative rights to robots so as to inhibit the maltreatment of robots by human interactors, thereby preventing negative impact upon the moral fibre and quality of human societies. In literature on robot rights, few authors clarify which kind of rights should, or should not, be granted to robots – something with which Tavani [12] takes issue. Although homing in on the concept of negative rights remains a broad account of the kind of rights that should be granted, it is, nonetheless, a distinction I have not yet come across. I draw the distinction between positive and negative rights from Berlin's [14] account on positive and negative liberty. Negative rights are rights that protect us *from* something. Positive rights, however, are rights we have *to* something. Negative rights, therefore, oblige inaction, since they protect people from being subjected to an immoral action. In this regard, if we considered granting robots negative rights, this would not be granting them for the sake of the robot but rather granting them so as to inhibit maltreatment, thereby possibly lessening negative impacts upon their human interactors. This provides a new perspective regarding the debate surrounding robot rights.

I thus conclude that regardless of the case for treating robots well from their own perspective, the case from the human perspective is enough to argue that robots should be treated morally well. Taking on this humanistic perspective will guide the way as to how we should go about formulating much needed ethical boundaries for our interaction with social robots, as well as what these ethical boundaries should ultimately be.

Keywords: android, companion robot, social robot, robot rights, robot ethics, moral patiency, HRI.

References

1. Anderson, M., Anderson, S. L.: Machine Ethics. Cambridge University Press, Cambridge (2011).
2. Bostrom, N.: Superintelligence: Paths, dangers, strategies. Oxford University Press, New York (2014).
3. Wallach, W., Allen, C.: Moral Machines: Teaching robots right from wrong. Oxford University Press, Oxford (2009).
4. Nagel, T.: What is it like to be a bat?. Philosophical Review 83(October), 435-450 (1974).
5. Gunkel, D. J.: The other question: can and should robots have right?. Ethics and Information Technology (2017).
6. Turner, J.: Why Robot Rights?. In: Robot Rules: Regulating Artificial Intelligence, pp. 145-171. Palgrave Macmillan, Cham (2019).
7. Darling, K.: Extending legal protection to social robots: the effects of anthropomorphism, empathy, and violent behavior towards robotic objects. In: Kerr, I., Froomkin, M., Calo, R. M. (eds.) Robot Law. Edward Elgar, Cheltenham (2016).
8. Scheutz, M.: The Inherent Dangers of Unidirectional Emotional Bonds. In: Lin, P., Abney, K., Bekey, G. A. (eds.) Robot Ethics: The Ethical and Social Implications of Robotics, pp. 205-221. MIT Press, Cambridge (2012).
9. Levy, D.: The Ethical Treatment of Artificially Conscious Robots. International Journal of Social Robotics 1(3), 209-216 (2009).
10. Ramey, C. H.: 'For the sake of others': The 'personal' ethics of human-android interaction. pp. 137-148. Cognitive Science Society, Stresa (2005).
11. Coeckelbergh, M.: Robot rights? Towards a social-relational justification of moral consideration. Ethics and Information Technology 12(3), 209–221 (2012).
12. Tavani, H. T.: Can Social Robots Qualify for Moral Consideration? Reframing the Question about Robot Rights. Information 9(73), (2018).
13. Asaro, P. M.: A Body to Kick, but Still No Soul to Damn: Legal Perspectives. In: Lin, P., Abney, K., & Bekey, G. A., (eds.) Robot Ethics: The Ethical and Social Implications of Robotics, pp. 169-186. MIT Press, Cambridge (2012).
14. Berlin, I.: Two Concepts of Liberty. In: Four Essays on Liberty, pp. 118-172. Oxford: Oxford University Press, Oxford (1969).