

Is there a need for robots with moral agency? A case study in social robotics

Rebecca Raper
School of Aerospace, Transport and

Manufacturing
Cranfield University

Cranfield, United Kingdom
rebecca.raper@cranfield.ac.uk

Abstract— There has been significant recent interest in the risks associated with Artificial Intelligence (AI), so much so that a Global AI Summit was recently hosted at Bletchley Park in the United Kingdom. One supposed risk associated with Artificial Intelligence is the threats that might be associated with an Artificial General Intelligence (AGI) carrying out acts detrimental to humanity. In the past, some researchers have attempted to bestow machines with morals to mitigate against these types of threat, however, in recent times the approach has been largely dismissed, with claims that giving machines moral agency poses more of a threat in of itself, than preventing it. One critique of the calls of the risk associated with AGI is that it is unrealistic, and that there is no grounding for any threat to humanity. The aim of this paper is to present a case study in social robotics to illustrate two points: 1) what real-life risks associated with AI might be, and 2) to reinstate the discussion surrounding whether there is a requirement for robots with moral agency.

Keywords—*Laws of Robotics, Moral Machines, Machine Ethics, Moral Agency, Robot Ethics, Social Robotics*

I. INTRODUCTION

In 1942 Isaac Asimov [1] introduced us to the idea of *Robots with Morals*. In a (not so unimaginable now) future where robots live alongside humans to provide policing, care, catering, autonomous driving and cleaning, Asimov introduces us to *The Laws of Robotics*, intended to govern machines to ensure they live in a moral way:

- 1 A robot may not injure a human being or, through inaction, allow a human being to come to harm.
- 2 A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
- 3 A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.

Later, to account for broader societal impacts arising from the laws, in his stories Asimov also introduces us to the Zeroth Law.

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

As the story goes on, the hope is that as long as the rules are adhered to, all should be well [for humanity]. However, even within the stories, things don't materialise in a particularly positive way. One issue is that the laws themselves are quite ambiguous when it comes to prescribing to a machine how to behave. The term 'harm', used in a context such as 'do not harm humans', might be obvious to us, but to a machine there are several unspecified terms. What do we mean by 'harm'? There are a variety of different types of

'harm': physical, psychological, social, financial... How should these be prioritised if different types of harm are at stake? Furthermore, even if 'harm' could be clearly defined for a machine, the rules seem not even applicable in every situation. What if *harming a human*, a bit (i.e., breaking an arm), is the only way to prevent them coming to an even greater harm (i.e., loss of life)? As in Trolley Case scenarios (see [2]), what if harming one human would preserve the lives of several other humans?

It is for these, and other obvious reasons, that The Laws of Robotics are regarded as insufficient to govern robots in a moral way. However, the question arises as to whether moral robots are even something that we need today? There have been arguments that not only are there no reasons to build moral machines in the first place [3], but that it would actually *be detrimental to humanity do so* [4]. This is owing to the increased new risk that if moral agency is bestowed, that the machines might become *immoral*. Many ([5],[6]) have also suggested that even if there were a motivation for creating machines with morals, that it could not be possible – largely because morality requires certain capabilities (i.e., self-awareness, emotions...?) that machines simply cannot, it is argued, possess. However, despite significant philosophical reflection having gone into these arguments, very few appeal to practical examples of Artificial Intelligence (or robotic) systems to validate the claims; they rest upon speculation about what our current technological requirements are, rather than working through real examples.

The aim of this paper is to contribute to both the discussions surrounding what the risks might be with Artificial Intelligence systems, and whether Moral Agency is something that would be *needed* to help mitigate against these risks by tracing through a *case study in social robotics*. That is, using a realistic scenario that has been designed involving a waiting service robot, enhanced clarity will be provided to understand the discussion surrounding moral agency in machines in more depth. If it is the case that the risks associated with social robots are such then they can only be mitigated by bestowing a machine with moral agency, then there is a renewed argument for development of research into this area.

The paper proceeds in the following way. After defining *Morality* and *Agency*, and in turn *Moral Agency*, it first explores some of the arguments that have previously been given for developing moral machines, including the fact that moral machines are *inevitable* and that through research in this area, we learn more about morality in of itself. The paper then outlines in more detail some of the arguments *against* moral machines, before giving a brief examination of their immediately obvious shortcomings. Then, a case study is presented involving a service waiting robot that has been

programmed using reinforcement learning, to illustrate how equipping such a robot with moral agency might be viewed as an exercise in risk mitigation, before the implications of this viewpoint are discussed in more detail, considering the narrow context of the case study. It is concluded that there *may* be a need for machines with moral agency, if moral agency is viewed as a mechanism for the risk mitigation of robots acting *immorally* in social situations.

II. MORAL AGENCY IN MACHINES

A. Morality and Agency

According to [7], morality can have either a descriptive or normative definition. Descriptively, morality can be understood as “certain codes of conduct put forward by a society or a group (such as a religion) or accepted by an individual for her own behavior”. Normatively, morality can be understood as “a code of conduct that, given specified conditions, would be put forward by all rational people” [7]. The key difference between the two definitions is that descriptively, morality is subjective, insofar as societies or groups can vary in their opinions on acceptable behaviour, whereas normatively, morality is objective, in-that there should *only be* one universal code of conduct.

Whether morality is subjective or objective in nature is open to debate, however, either way, for the purpose of this paper it can be understood that a machine that acts according to a code of conduct, whether this is a universal code of conduct, or one put forth by a group or society, can be said to be acting morally.

Extending the definition; Agency can be understood as a theory of action [8], insofar as something can be said to have agency (be an ‘agent’) if it manifests an action. A machine then, has agency if it carries out actions.

As [9] outlines, there is some contention about whether intentionality is a requirement for full agency. That is: whether the ability to have representations of the mental state causing the action, is needed for something to be regarded as an agent. However, again, as far as this paper is concerned, we can say that a machine has agency if it has the ability manifest an action. Finally then, a machine can be regarded as having *Moral Agency* if it can carry out acts that are in line with a code of conduct, whether this be universal or put forth by a group or society.

Moral Agency in machines is something that is considered extensively by Wallach and Allen [10], who highlight a scale of different levels of Moral Agency in machines. At the beginning of the scale, a machine demonstrates what they refer to as ‘operational morality’ simply by being designed in a way that it always makes decisions that are moral, without any kind of moral decision-making. An example of a machine demonstrating ‘operational morality’ might be an industrial robot that has been limited in its movements to prevent it carrying out an action that is immoral. For example, a robot may be fixed to a surface to prevent it have consequences that would only be of a mobile robot. Next along the spectrum is what is termed ‘functional morality’, which is where machines are either not fully autonomous in their moral decision-making or are autonomous, but with little ethical sensitivity. An example of the former might be an engine that

recommends appropriate actions in ethically sensitive situations (for example, MedEthEx by Anderson and Anderson [11]), whereas an autopilot system that avoids making immoral decisions would be an example of the latter. What is termed a “Full Moral Agent” then would be machine that is fully autonomous, with a high level of ethical sensitivity.

At this point the paper remains neutral as to whether morality should be operational, functional or full. However, the demarcation between types of morality will be revisited at the end. For the sake of this paper though, the question concerns whether moral agency in a machine is something that is needed at all.

B. Machine Ethics

The discipline aimed at trying to bestow morals unto machines, back in the early 00’s, is popularly known as *Machine Ethics*. According to Anderson and Anderson [12], back in 2007:

“Why is the field of Machine Ethics important? There are at least three reasons that can be given. First, there are ethical ramifications to what machines currently do and are projected to do in the future. [...] Second, it could be argued that humans’ fear of the possibility of autonomous intelligent machines stems from their concern about whether these machines will behave ethically, so the future of AI may be at stake. Whether society allows AI researchers to develop anything like autonomous intelligence machines may hinge on whether they are able to build in safeguards against unethical behaviour. [...] Finally, we believe that it’s possible that research will advance the study of ethical theory.”

As Anderson and Anderson [12] go on to further explain, if we consider care robots placed into people’s homes that make daily decisions about the administration of medication, for instance, we can see that some of the decisions are going to have ethical implications. Should the machine insist on giving medication to the individual they are supporting if that individual refuses it?

As an extension of these arguments, Wallach and Allen [10] refer to what they term *the inevitability argument* in support of Machine Ethics. According to their argument, engineers implicitly put their values into machines when they create them anyway, so it is better to have machines with explicit ethical decision-making so that these prejudices can be more easily managed.

For instance, if we consider an AI algorithm designed to decide who is awarded credit for a bank loan, the engineer’s assumptions will have gone into various aspects of how the algorithm was designed, such as, which metrics are given the greatest decision weighting, what the exclusion principles are... If we actively sought to ensure that the decisions a machine made were moral, then these implicit prejudices would not only be highlighted (through the auditing process to ensure that moral decisions were being made) but be mitigated against through the mechanisms that are employed. In this regard then, machines need explicit moral reasoning so that there is at least some oversight on the ethical implications resulting from the design decisions for autonomous decision-making systems.

III. THE CASE AGAINST MORAL MACHINES

According to Van Wynesberghe and Robbins [3], the case for Machine Ethics is insufficiently strong to support future research in the area. They argue that not only is there no good argument for moral machines in the first place, but that research into the area should be stopped until there is a good argument for their creation. They argue that this is because there are probably more risks associated with moral AI than there are benefits. They argue that it is the duty of the Machine Ethicist to provide good arguments to support the claim that artificial moral agents are needed before research in the area continues. They support the claim that there are no good arguments for moral machines, by listing what they see as key arguments *for* Machine Ethics, and by responding to each of these. They list these as follows:

- **Inevitability** – the response given is that machines can be placed in morally charged situations without the requirement for moral agency. For example, therapy dogs are frequently placed into moral situations (i.e., situations involving vulnerable individuals) but there is no need for the dogs to make moral decisions.
- **Preventing harm to humans** – their response against this motivation is that the term ‘moral’ in the context of protecting humans does no more than invite a fictive or deceptive relationship between human and machine, when ‘safety’ seems to describe what is needed in machines just as well.
- **Complexity** – machines don’t need ethical decision-making to manage the increasing complexity of decision-making systems because unpredictable machines should simply not be placed into situations that could cause harm.
- **Public Trust** – just because moral machines might increase public trust in robots, does not mean that they *should*. There is a difference between *acceptance* (which could be cultivated with the public in a deceptive way) and *acceptability* (something morally worthy of being accepted). Moral machines might increase public trust, but this cannot be used as a reason for their development.
- **Preventing Immoral use** – moral machines serving as a protective mechanism against immoral use is refuted on the grounds that controlling a human (by ensuring *moral* use) is not necessarily a moral thing to do and that it’s not obvious that one ethical theory would make the right decision in this type of circumstance.
- **Morality better with Moral Machines** – the common assumption that a moral machine, due to impartiality, lack of emotional affect and rationality for every decision, would make better moral decisions than a human, is refuted on the grounds that we cannot be in any position to measure ‘better performance’ because to do so would require an objective account of what constitutes morality, which does not exist.
- **Better Understanding of Morality** – the claim that moral machines would give us a better understanding

of morality is refuted on the grounds that moral decision-making has nothing to do with ethical theory, and since Machine Ethics involves exploration of ethical theories, no lessons can be learned regarding the essence of morality.

Many of the responses listed seem to miss the essence of the arguments provided in support of Machine Ethics research. For instance, the argument against inevitability does not only argue that machines will be placed into ethically sensitive contexts (such as the therapy dog), but it argues that moral agency is required because the designer’s implicit ethical prejudices will end up governing the behaviours of the AI otherwise. There doesn’t seem to be an accurate parallel between the therapy dog and the AI making credit decisions, because no implicit ethical prejudices will have gone into the decisions the therapy dog makes. Similarly, since the paper was written, machines *have been* placed into decision-making situations that have ethical ramifications. For instance, Amazon’s Alexa is frequently asked questions with moral import, and ChatGPT, if prompted with the right questions, can act in a counselling capacity, giving advice on topics such as relationships and personal issues.

It is also unfair to argue that because we don’t yet understand certain aspects of morality, that we should not be attempting to build moral machines. For instance, it might be that we cannot objectively assess whether a moral machines’ moral decisions are better than a human, but there are metrics to assess moral development, or *moral maturity* and these could be used to determine whether we have created a moral machine. There is also no strict requirement that Machine Ethics rely solely on ethical theory for its advancement, and therefore, it is entirely conceivable that the field of Machine Ethics might contribute to our understanding of morality (even if just from a cognitive perspective).

According to [13] we are at greater risk of *immoral machines* if we attempt to create machines with moral agency, because if we understand how to engineer morality in a machine, this could easily be susceptible to ‘hacking’ or ‘reverse engineering’, making it easier to engineer immoral robots.

The general line of this argument is problematic because it assumes that because there is an undesirable version of something, that we should not endure to create a desirable version of it. Similarly, even more problematically, it assumes that morality is something that can be reverse engineered into its opposite, rather than something of an assurance mechanism that strengthens decision-making. For example, we might strengthen branches on a tree by adding scaffolding, but the fact that we have designed a mechanism for strengthening the branches, does not mean we now have an easier way to weaken the branches. This is why to truly understand the mechanism that is required for AI systems, a practical case study is required.

IV. A CASE STUDY IN SOCIAL ROBOTICS

A. The Service Robot

As per [14], in robotics, significant progress has been made using a version of Machine Learning known as Reinforcement Learning (RL). RL works by motivating machine behaviour through reward [15]. The machine may

learn – for example – how to throw and catch a tennis ball – by receiving positive reinforcement every time it successfully completes its task of throwing and catching a tennis ball. If we

indicated by the colour scale in Figure 1, of the 5 customers, 1 is happy, 2 are ok, and 2 are unhappy. Using RL then, perhaps after specifying that the robot deliver plates to tables,

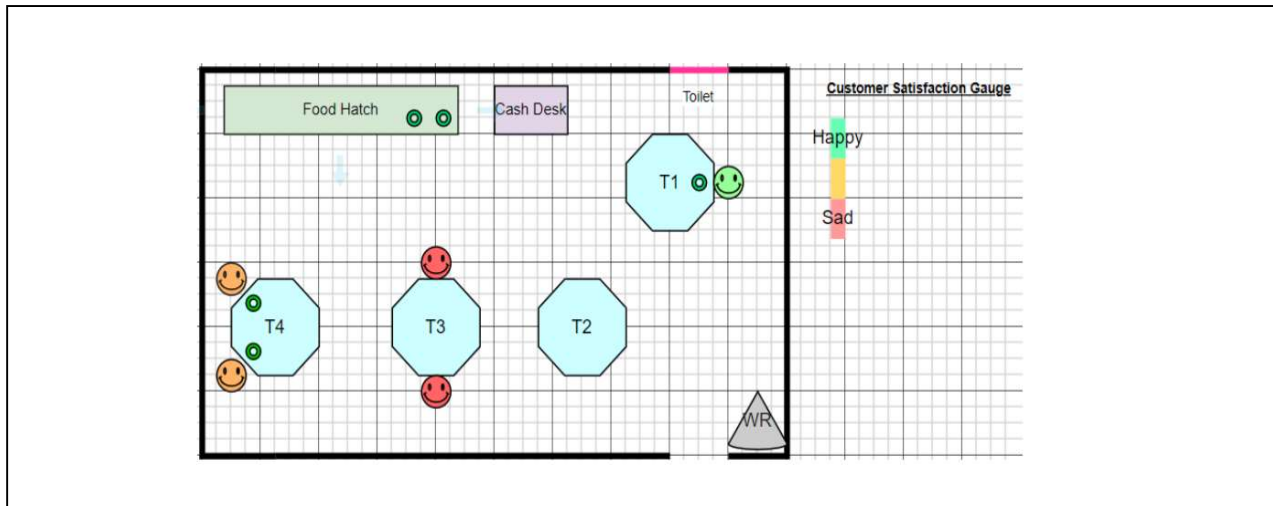


Figure 1: a diagram representing a service robot working in restaurant.

consider a machine making decisions that affect humans, it's useful to play out a typical scenario that a robot might be faced with, to understand the emerging issues. This robot would presumably be programmed to carry out its behaviour using a technique such as Reinforcement Learning, and it would be in a situation where it must autonomously carry out an overarching task or 'role'. One such example might be a service robot within a restaurant, where the robot is tasked with showing customers to their seats, taking orders, collecting prepared food, delivering the food to customers, taking payment for meals, and cleaning up afterwards. Figure 1 gives a pictorial representation of this scenario.

Where 'WR' denotes the service robot in this case ('Waiting Robot'), one of the key tasks at hand is to appease customers by effectively delivering food (indicated by the green circles on Figure 1) from the hatch to the tables. There are various social etiquettes that should be managed in this situation. First, the robot should, if acting as an effective waiting robot, deliver the food in a 'pleasant' way. One way to optimally satisfy the goal of 'delivering food' might be to *throw* the food at the customers, however, throwing food at customers to speed up the process wouldn't really satisfy the more social aspects associated with being a waiter. Similarly, 'delivering food' to the right table, but at the wrong side, would not sufficiently complete the task. Simply specifying rules for behaviour (e.g., 'deliver food to tables') is not sufficiently specific to capture the needs of the role, because it does not specify *how* the food should be delivered. We might narrow the specification by asserting that the robot 'deliver food to people, sat at tables, who have already placed an order...' but even if we could capture all the necessary requirements in this situation, it still precludes delivering food in a *socially acceptable way*. Defining what is socially acceptable in a prescribed list of rules would be a big challenge. Therefore, we might imagine that to cultivate the right kind of *social* behaviour in WR, using RL, rewards for making customers 'happy', could be an applied strategy. As

the robot's goal would be to *maximise happiness*.

The problem is, that in using RL, again it has not been specified *how* the happiness of the customers should be maximised. In this instance the robot could carry out any action to achieve this goal, which might not even correspond to the overarching aim to deliver plates in a socially acceptable way. For instance, we can imagine that one way to *maximise happiness* in this scenario would be to adopt a deceptive manoeuvre where customers believe they are getting a good service, and are thereby happy in the restaurant, but are *not actually getting a good service*. For instance, WR might discover that adding extra salt to the meals is a way to appease customers because it means their food (albeit unhealthy) is full of flavour.

This would be a morally dubious action, but given that WR knows nothing other than 'maximising happiness' (as long as 'food delivering' is satisfied), there is nothing to stop WR from adopting this as its given strategy. However, the act could have significant moral ramifications: the customers might leave the restaurant happy but with resulting health problems. One way to mitigate against this type of incident might be to remove all salt from the restaurant area, so the robot has no means to potentially harm unsuspecting humans. However, described in this way we can begin to see that if we want a robot to act in a social environment, demonstrating features such as social etiquette and behaving in a morally acceptable way, that there is some need for moral agency, insofar as there is a need to *prevent the machine acting immorally*. In the case just presented, we need to come up with some mechanism to prevent the machine putting salt onto customer meals without their consent. The challenge seems to be one in assurance, where we need to *assure* that social machines do not carry out an act that is immoral.

It might be argued that though there appears to be a risk associated with such a case study that safety, and not morality, is the real resolution to this problem. One of the arguments

from [3] was that the challenge can as easily be articulated as a challenge in safety, and therefore that it should avoid hyperbole and unhealthy human relationships with machines. The argument goes that in this instance we want to prevent machines *poisoning humans*, therefore, we need to make systems that are *safe, not moral*. However, although this specific circumstance describes an exercise in safety management (we are protecting humans by preventing salt being placed into their food), the problem is much greater than this, and extends outside of the realms of just safety for humans. For instance, under this view, we need to consider *whom the robot is safe for*? If we focus on just making the AI systems *safe for humans*, we are limited in that we are only really mitigating against robot misbehaviour as it concerns humans. We might imagine a scenario where WR learns to make humans happy by sacrificing animals. The outcome would not necessarily affect human safety in this instance, but there would be obvious moral implications. Due to the unknowability of many AI systems, we can't predict what type of hazards are going to appear, or *when*. Food poisoning (and the issue of *human safety*) is not the only issue that might crop up. Bestowing machines with moral agency – the ability to make *good* decisions on their own – seems like an appropriate approach to ensure that *all* adverse risks are captured.

It might be argued that if we avoid creating social robots, we therefore avoid the need for such mitigation. For instance, if we just accept that machines will deliver food to us, however they see fit (even if this is by throwing food at us), then we avoid the issue concerning happiness manipulation. However, such an argument would be to under-appreciate the extent to which social etiquette plays a role in any task completion involving humans. Just navigating the room has moral implications. What if a child were to suddenly run into the path of WR as it were delivering food? Some mechanism would need to be put in place to prioritise children safety over delivery of food. This again, would not just be limited to human safety: what if an unsuspecting dog were to cross the path? What if the restaurant was outdoor, and wildflowers were scattered in several places between tables that could lead to quicker food delivery? Even in this instance WR should know that avoiding the flowers, despite a need to quickly deliver food, is the most appropriate strategy. It seems that if we want machines that effectively complete any level of autonomous task, that some degree of moral agency is required. This might be why ChatGPT, an agent beginning to have social import, has raised alarms with senior AI researchers. Essentially, as artificial agents become more autonomous, additional risks seem to become apparent.

B. Implications and Moving Forward

This *assurance* approach to viewing Machine Ethics brings about new challenges and new opportunities for research in this area. Machine Ethics becomes less about bestowing machines with morality in terms of programming them with abstract rules or behaviours, and more concerned with *mitigating immoral behaviour*, by ensuring that machines always act in moral ways. If we revisit the levels of Moral Agency discussed at the beginning of this paper, this approach therefore does not preclude any of the types of Moral Agency as mechanisms for protecting us against immoral behaviour. In-fact, initially, and since we might have more knowledge in this type of mitigation, we might assure safety through

operational constraints, such as by removing all salt from the robot's operating area. In essence, before any robot is deployed autonomously into an environment, such risk assessments (and relevant risk mitigation techniques) should be carried out. Questions should be asked such as 'what unintended hazards could occur from an autonomous agent in this situation?' 'What precautions and management strategies need to be put into place to prevent these happening?'. 'Has the robot been tested to ensure that these threats cannot materialise?'. These lessons could be taken and applied to any situation in which there is an autonomous machine. The responsible robotics engineer then would need to assure their system is safe through relevant constraint mechanisms before it is implemented. Regulation could guide the design and implementation procedure to ensure that necessary constraints are applied.

As more research into Machine Ethics is carried out, and there is a greater understanding of Morality in general, functional models could begin to facilitate this safe operation until machine moral agency is wholly assured. Moral decision-making would be assured by rigorous testing and verification.

V. CONCLUSION

Various arguments and perspectives have been given in this paper. Initially, there were arguments *for* the need for Machine Moral Agency, which focused on the inevitability of machines having moral import and the ability to facilitate our understanding of morality. These arguments were rebutted on the basis that just because a machine is in a moral circumstance, does not mean it needs to be moral, and through arguing the distinction between acceptance and acceptability. However, after noting that many of the arguments given do not seem to consider real-life AI-based scenarios, a case study in social robotics was presented which illustrated the need for moral agency, insofar as protecting us (and the world) from *immoral* machine behaviour. This may be requirement above and beyond simply the need for safety, as it extends beyond the realm of just protecting humans. As machines become more autonomous, there is an increased need for their moral autonomy, owing to the risks associated with unpredictable behaviours of AI-based systems.

To conclude, this paper has provided a case study to suggest that there *does* seem to be a need for machines with moral agency – if such machines are ever going to interact in social environments. Given recent technological advancements and as machines become integrated more into our social lives, this issue becomes increasingly pressing. Machine Ethics can be viewed as an exercise in risk mitigation.

ACKNOWLEDGMENT

This work and associated reflections was born out my PhD and discussions with various individuals involved in it. I'd like to thank those who were closely involved in conversation throughout the scoping and development phase of my thesis, as well as all of those who personally supported me, including family and friends.

REFERENCES

- [1] Asimov, I., 1942. Runaround. *Astounding science fiction*, 29(1), pp.94-103.
- [2] Foot, P., 1967. The problem of abortion and the doctrine of the double effect.
- [3] Van Wynsberghe, A. and Robbins, S., 2019. Critiquing the reasons for making artificial moral agents. *Science and engineering ethics*, 25, pp.719-735.
- [4] Vanderelst, D. and Winfield, A., 2018, December. The dark side of ethical robots. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 317-322).
- [5] Coeckelbergh, M., 2010. Moral appearances: emotions, robots, and human morality. *Ethics and Information Technology*, 12, pp.235-241.
- [6] Sparrow, R., 2021. Why machines cannot be moral. *AI & SOCIETY*, 36(3), pp.685-693.
- [7] Gert, Bernard and Joshua Gert, "The Definition of Morality", *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2020/entries/morality-definition/>. [Last accessed 19/11/2023]
- [8] Schlosser, Markus, "Agency", *The Stanford Encyclopedia of Philosophy* (Winter 2019 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2019/entries/agency/>.
- [9] Jacob, Pierre, "Intentionality", *The Stanford Encyclopedia of Philosophy* (Spring 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/spr2023/entries/intentionality/>.
- [10] Wallach, W. and Allen, C., 2008. *Moral machines: Teaching robots right from wrong*. Oxford University Press.
- [11] Anderson, M., Anderson, S.L. and Armen, C., 2006, August. MedEthEx: a prototype medical ethics advisor. In *Proceedings of the national conference on artificial intelligence* (Vol. 21, No. 2, p. 1759). Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999.
- [12] Anderson, M. and Anderson, S.L., 2007. Machine ethics: Creating an ethical intelligent agent. *AI magazine*, 28(4), pp.15-15.
- [13] Vanderelst, D. and Winfield, A., 2018, December. The dark side of ethical robots. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 317-322).
- [14] Crook, N. and Corneli, J., 2021. The Anatomy of moral agency: A theological and neuroscience inspired model of virtue ethics. *Cognitive Computation and Systems*, 3(2), pp.109-122.
- [15] Sutton, R.S. and Barto, A.G., 2018. *Reinforcement learning: An introduction*. MIT press.



DSPACE

<https://dspace.org/>

Is there a need for robots with moral agency? A case study in social robotics

Raper, Rebecca

2024-06-05

Institute of Electrical and Electronics Engineers (IEEE)

<https://doi.org/10.1109/ICIT58233.2024.10540919>

Raper R. (2024) Is there a need for robots with moral agency? A case study in social robotics. In: 2024 IEEE International Conference on Industrial Technology (ICIT), 25-27 March 2024, Bristol, UK