WHY DO WE LIKE MR. ROBOT, BUT NOT DR. ROBOT? SOME MORAL CONSTRAINTS ON TELEMEDICINE IN SEARCH FOR A "RE-HUMANIZING REVOLUTION"

SILVIYA SERAFIMOVA

Institute of Philosophy and Sociology, BAS

silvija_serafimova@yahoo.com

Abstract

The main objective of this article is to demonstrate why establishing the criterion of reason responsiveness is only a necessary condition for the successful (in both moral and epistemological terms) implementation of telepresence robots as moral prompters (reasoners). By extrapolating the use of Kantian and utilitarian first-order normative theories to the field of machine ethics, I examine why some of the main difficulties in arguing for such moral prompters – consequently, for the development of telepresence robots as potential moral advisers and moral observers – derive from the fact that their application raises some significant moral dilemmas which go beyond the field of machine ethics as such. Clarifying why those dilemmas are unsolvable, I try to demonstrate why not only strong, but also weak 'moral' AI scenarios require further investigation.

Key words: machine ethics, moral epistemology, Kantian and utilitarian first-order normative theories, telepresence robots

1. Introduction

In this paper, my primary objective is to demonstrate why not only so-called strong AI (which aims at developing a human-level of cognition in a computer), but also so-called weak AI (which assumes computing and updating data to assist human agents in making moral decisions) (Savulescu and Maslen, 2015: 84) raises serious concerns about the moral enhancement in the field of telemedicine. Those concerns are evaluated from the perspective of machine ethics whose ultimate objective is to build artificial autonomous moral agents (AAMAs). Analyzing the origin of the challenges faced when humans try to design ethical intelligent agents (Anderson and Anderson, 2007), one sees why AI functioning as a moral prompter cannot guarantee the role of AI as a means to moral enhancement (Klincewicz, 2016), even if one assumes that AI can be successfully developed so as to persuade reason-responsive people.

I would argue that having well-working ethical intelligent agents in epistemic terms (where by epistemic one understands the verification of ethical principles guiding the computational process of decision-making) does not mean that those agents are moral by default. One of the main reasons for taking such a stance is that what is recognized as 'right' in epistemic terms is not necessarily 'good' in a moral sense.

In practical terms, I examine a relatively recent example triggering some strong objections to the adoption of a telepresence robot in a hospital environment. I raise a hypothesis that even if the telepresence robot is considered as well built in technical terms, it cannot successfully function as a moral adviser nor as a moral observer (viz., as an explicit ethical agent) (Moor, 2006), since it fails as a moral prompter (viz., as an implicit ethical agent) (Ibid.). The reasons behind such a hypothesis are that communicating sensitive info about end-of-life treatment reveals why modifying the health care providers' approach cannot be successfully carried out, unless it is firstly specified in the scope of building agent-agent trust (Klincewicz, 2016) (Nickel, 2013).

1.1. Key Concepts and Clarifications

Analyzing Anderson and Anderson's statement that "having all the information and facility in the world won't, by itself, generate ethical behavior in a machine" (Anderson and Anderson, 2007: 15) nor would experts find easy solutions due to the fact that "ethics has not been completely codified" (Ibid.), one can spot the challenge as consisting in how one can create "an ethical intelligent agent" (Ibid.).

The plurality of combined approaches points towards arguing for two main types of ethical intelligent agents, namely, for so-called implicit and explicit ethical agents (Moor, 2006: 19-20). While by implicit ethical agent Moor understands a machine that has been programmed to behave ethically by following some ethical principles, as embodied by its designer, by explicit ethical agent he means a machine which "is able to calculate the best action in ethical dilemmas using ethical principles" (Anderson and Anderson, 2007: 15). What is of primary importance for machine ethics is to "create a machine that is an explicit ethical agent" (Ibid.).

In the field of machine ethics, researchers discern between three sub-fields of AI expertise. Firstly, AI could be used as a moral environment monitor (which "provides morally-relevant information about one's surroundings"), secondly – AI can work as a moral prompter

whose function is to inform the user about morally important situations or events. In turn, the third and the fourth types of expertise concern the role of AI as a moral adviser (nudging the user what one should do) and a moral observer (observing the behavior of the user) (Savulescu and Maslen, 2015: 84) (Klincewicz, 2016: 174).

The investigations show that the role of artificial moral adviser and that of ideal observer in Firth's sense are not mutually exclusive. This is due to the fact that the artificial moral adviser is supposed to be disinterested, dispassionate and consistent in formulating moral judgments, similarly to the ideal moral observer (Giubilini and Savulescu, 2018). However, the main difference is that the artificial moral adviser would lack one important feature of the ideal observer, namely, that of being 'absolute' in moral terms, if by absolute one understands the capacity of the ideal observer to provide ethical statements which do not contain egocentric expressions (Ibid.).

Clarifying how one can couple morality with rationality in respect to ethical intelligent agents would mean to examine the implications of what is called moral reasoning. "Moral reasoning leads a moral agent to make moral decisions based on relevant reasons, broadly construed."; consequently, machine ethics aims to implement that process in a computer (Klincewicz, 2016: 179). In this context, I would argue that one of the primary objectives of machine ethics can be defined as revealing how one can justify the implications of moral reasoning when adopted by an explicit ethical agent in Moor's sense, or as Wallach and Allen cogently coin it, by turning computers into "explicit moral reasoners" (Wallach and Allen, 2010: 6).

Analyzing the role of moral reasoning is important, since it contributes to revealing whether or not elaborating upon the distinction between so-called strong and weak AI scenarios (Savulescu and Maslen, 2015: 84) can be extrapolated to what I would call strong 'moral' AI scenarios (looking for an AAMA as an explicit ethical agent) and weak 'moral' AI scenarios (designing an implicit ethical agent, namely, an AMA).[1]

Most researchers clearly argue that building fully conscious artificial systems with complete human moral capacities is not something which will happen soon if it happens at all (Wallach and Allen, 2010: 6) (Anderson and Anderson, 2007: 15-24) (Klincewicz, 2016). However, the main methodological benefit of examining the challenges such a project brings to light is that in the process of addressing, humans will understand "what truly remarkable

creatures they are" (Wallach and Allen, 2010: 11). This in turn would shed light upon the issue of why increasing agent-computer trust (Klincewicz, 2016) (Nickel, 2013) cannot successfully result in overcoming some ethical dilemmas, unless they are firstly overcome by encouraging agent-agent trust, as I will show below.

2. Questioning the Role of a Telepresence Robot as an Implicit Ethical Agent

I exemplify the aforementioned challenges in coupling morality with rationality by analyzing a relatively recent case in which the use of a telepresence robot triggers some strong moral concerns. On March 10th, 2019 the local U.S. station KTVU published information about the reaction of the family of a dying 78-year-old man. The family had been left devastated after a telepresence robot rolled into man's room to deliver the news that he did not have long to live.

The family of a dying 78-year-old man have been left devastated after a hospital robot rolled into his room to deliver the news that he did not have long to live:

Ernest Quintana was taken to the Kaiser Permanente Medical Center in the San Francisco Bay area of California last weekend after suffering breathing difficulties, and died on Tuesday. His family had known that he would die of his chronic lung disease, but were not expecting to receive news of his imminent death when a robot arrived at the intensive care unit on the night Mr. Quintana was admitted. Rather than a doctor to deliver the news that he would likely die within days in person, one flashed up on a screen on the robot to tell them via video link. It has sparked outrage among the family of Mr. Quintana, who thought the robot was making routine visit. Daughter Catherine Quintana said: "If you're coming to tell us normal news, that's fine. "But if you're coming to tell us there's no lung left and we want to put you on a morphine drip until you die, it should be done by a human being and not a machine." Part of the exchange between the on-screen doctor and the shocked family was captured on video by Annalisia Wilharm, granddaughter of Mr. Quintana, who was alone with him when the robot appeared. The 33-year-old said she had been told by a nurse that a doctor would be making his rounds. "This guy cannot breathe, and he's got this robot trying to talk to him," she said. "This guy is telling him, 'So we've got your results back, and there's no lung left. There's no lung to work with.'" Ms. Wilharm said she had to repeat what the doctor said to her grandfather because he was hard of hearing in his right ear and the machine was unable to get to the other side of the bed. The hospital has defended its use of the robot, insisting that the diagnosis came after

several physician visits and that it did not replace "previous conversations with the patient and family members". But Michelle Gaskill-Hames, senior vice-president of Kaiser Permanente Greater Southern Alameda County, added that hospital policy was to have a nurse or doctor in the room when remote consultations took place and acknowledged it had fallen short of the family's expectations. "We will use this as an opportunity to review how to improve patient experience with tele-video capabilities," she said (Doctor Told California Man He Was Dying Via Robot Video Call, CBS News, 2019).

The short video recorded by the granddaughter went viral online and got many negative responses, since it was uploaded on YouTube. Most users described the situation as "cruel" and "inhuman", arguing against the lack of compassion and sympathy which the patient deserved. In a blog called *The Re-humanizing Revolution*, an anonymous writer described the situation as a "dystopian", "inhuman" picture (Anonymous, 2019) by drawing a parallel with "2001 Space Odyssey" in which the computer HAL 9000 had to inform the spaceship captain that he would soon be dead (Ibid.). In addition to the thought-provoking arguments regarding the lack of interpersonal interaction between the doctor and the patient which makes the question "Do you understand me?" impossible to ask, the anonymous writer emphasizes that ironically, the style of language used by the doctor resembles that learnt from the 0-1 binary code which robots use: "since active treatment is not working, we must move to 'comfort therapy'" (Ibid.).

Drawing on the concerns displayed in the aforementioned example with the telepresence robot, I try to demonstrate why the difficulties are exacerbated by the fact that they cannot be successfully overcome neither in favor of rationality nor in favor of morality alone. I present two important arguments which show why moral dilemmas are irreducible to those triggered within the field of machine ethics: the complex nature of moral expectations and one "troubling phenomenon that is not necessarily unique to robots" (Arkin and Borenstein, 2016: 40), namely, what Hansson refers to as "subordination to the technology" (Hansson, 2007: 264). [2]

According to the phenomenon in question, "There is a risk that users will feel that they are controlled by this technology, rather than using it themselves to control their surroundings" (Ibid.: 265). Some illuminative embodiments of "the subordination to the technology" phenomenon can be traced back to the reaction of the granddaughter who insisted on getting her grandfather's diagnosis by a human being and not by a machine. As the anonymous writer thoughtfully argues, the defense of using telemedicine and tele-visits "in this way seems wrong"

because the patient was dying, while the doctor "was somewhere lost in space and in translation" (Anonymous, 2019).

As an illustration of why what is considered as an enhancement is not necessarily a matter of moral enhancement, I point out the response of the authorities including Kaiser Permanente and caregivers in Fremont. They claimed that the video technology was used "as an appropriate enhancement to the care team, and a way to bring additional consultative expertise to the bedside" (Cusano, 2019). Even if one can clearly understand why and how the telepresence robot has been used to provide an enhancement, which can be evaluated as appropriate in terms of making a successful tele-visit, this rationally responsive person might face some serious difficulties in understanding exactly how the "difficult information" has been transmitted rather than shared "with compassion in a personal manner", [3] as the authorities insist.

In other words, being rationally responsive can be considered only a necessary condition for reacting as a morally responsive person. This is due to the fact that moral responsiveness is only partially dependent upon rational responsiveness, as the aforementioned example clearly shows. The discrepancy between rational and moral responsiveness can be well demonstrated by B. Tingley's saying which described the situation as follows: "While using robots and AI to more efficiently treat sick people and save lives is certainly nothing anyone could argue against, a recent incident in California highlights the fact that no matter how much we might love *Mr. Robot*, we might not be ready for Dr. Robot" (Tingley, 2019).

2.1. Some Implications of Moral Epistemology

Why do the authorities argue for an improvement of patient experience if the telepresence robot has not made any mistakes whatsoever in transmitting the diagnosis? In other words, if the robot has fulfilled its task (to transmit the message), why do the authorities still argue for making an improvement in the fields of telemedicine and telecommunication at all? Is it not a paradox that it is the accuracy of the telepresence robot that makes the family members so hostile to the way in which they have been informed?

The hostility originates from machine ethics's failure to solve moral dilemmas by adopting machines' means. Some reasons for the failure in question can be found in machine ethics's inability to build a human-level of moral reasoning into the telepresence robot. This is

due to the fact that one cannot unquestionably define when what is considered as right can also be evaluated as morally good. As demonstrated above, the family's critical response derives from the fact that what is considered as right and conducted correctly, namely, informing a person about a diagnosis of terminal illness by a robot, is evaluated as morally wrong, although there are no epistemic counter-arguments for that.

As Wallach and Allen relevantly argue, "we have not seen any evidence that full machine control over life-and-death decisions has gotten any closer" (Wallach and Allen, 2010: 41). In turn, this conclusion raises many concerns which deserve serious consideration, such as those about patients' autonomy and well-being, as well as the way in which patients' concerns can be reduced by increasing the level of trust within the patient-physician interaction. In this context, the issue of informed consent enters the stage – namely, whether or not computers can be as reliable as patients' relatives in predicting the preferences of terminally ill patients.

Regarding Mr. Quintana's particular case, Michelle Gaskill-Hames, senior vice president and area manager of Kaiser Permanente Greater Southern Alameda County, expressed Kaiser Permanente's condolences to the relatives. But she contested the use of the term 'robot' to describe the device that the doctor used to speak to Mr. Quintana. In an emailed statement to the magazine *The Verge*, Michelle Gaskill-Hames called the term 'robot' "inaccurate and inappropriate". "This secure video technology is a live conversation with a physician using tele-video technology, and always with a nurse or other physician in the room to explain the purpose and function of the technology," she stated. Previous posts on the organization's websites and YouTube channel did, in fact, refer to the technology as a robot. However, as R. Becker cogently argues, "quibbling over the word "robot" is beside the point: the point is that Quintana…spent some of his last moments interacting with a person through a screen rather than face-to-face" (Becker, 2019).

Speaking from the perspective of machine ethics, there is no issue to argue for: some investigations of the patient's condition were properly conducted. This led to the conclusion that it is doctors' moral obligation to inform the relatives about the patient's terminal illness. The doctor used a robot which is designed to be able to transmit the diagnosis as a message on a screen. In turn, the family members are reason-responsive people who understand the diagnosis, namely, they do not have concerns about whether or not the diagnosis is correct. Consequently, their mistrust does not stem from the information transmitted by the robot itself.

Based on the outlined circumstances and investigations, I would argue that the concerns regarding the application of telemedicine's device correspond to those raised in the field of bioethics, namely, they ask the question who has the right to decide on behalf of others, rather than the question how one can generate "a better software" in this context.

Going back to the "subordination to technology" phenomenon, I claim that the discrepancy between what is right and what is good in a moral sense results from the fact that even if one adopts some rational means, the process of persuasion is not necessarily justifiable in moral terms. One of the reasons is that moral responsiveness depends not only upon the well-groundedness of the arguments, but also upon many other individual factors such as the impact of different types of biases (both conscious and unconscious) concerning moral motivation, moral emotions and moral expectations for what empathy and compassion should look like.[4] Those factors cannot be calculated by providing any 'moral' arithmetic whatsoever.[5]

2.2. Exemplifying Kantian and Utilitarian Scenarios for the Case with the Telepresence Robot

In response to the criticism regarding Mr. Quintana's case, *The Verge* spoke with Alex John London, the Clara L. West professor of ethics and philosophy at Carnegie Mellon University (who is also the director of CMU's Center for Ethics and Policy). He was asked when telepresence robots are, or are not, appropriate. As London cogently argues, some difficulties arise due to the fact that the medical personnel are familiar with the healthcare environments, while the patients are not, which may raise uncertainty and distress among them. Certainly, the doctor-patient relationship should be "human", i.e. including compassion and empathy towards the patient when bad news are announced (Becker, 2019). The question, however, is how being familiar with the role of telepresence robots can enable the doctor to increase, or at least maintain, the same level of compassion and understanding whilst using robots as mediators, as well as how patients can be successfully prepared for such a mediation in moral terms. London also told *The Verge* that in Mr. Quintana's case, the telepresence robot created a distance between the family and the clinician, "which, instead of the empathy that you want out of this kind of interaction, further depersonalized it" (Ibid.).

In this context, several questions arise. Seeing as there is no proof that the use of telepresence robots can help us attain a level of trust similar to that gained through face-to-face

interaction between doctors and patients, is it not more reasonable to work on improving interhuman moral interactions rather than focusing on interactions which already entail a high risk of depersonalization to begin with? On the other hand, even if we agree that in some cases people trust computers because they used to distrust agent-agent relationships (Klincewicz and Frank, 2018), is it not more reasonable to enhance the relationships between moral agents rather than those between moral agents and robots? [6] Furthermore, even if there is a pre-existing bond between patient and doctor, which is outlined as one of the cases when telepresence robots can be used, what guarantees that the moral and emotional biases concerning the doctor-patient relationship will not be neglected, when mediated by a non-human transmitter? In fact, it may have just the opposite effect, as demonstrated by the case with Mr. Quintana.

Judging by the aforementioned investigations, I argue that it is the role of moral expectations and the lack of their fulfillment that make the debate about turning telepresence robots into potential moral advisers and moral observers unsolvable in the field of machine ethics. This is due to the fact that they cannot successfully work as moral prompters either. Analyzing the messages, as provided in the discussed example, one can speculate about the circumstances under which the expectations of the family members could be fulfilled in a scenario which involves a robot.

For the sake of conducting such an experiment, I will go back to one of the following definitions of machine ethics. The AI system is supposed to generate persuasive arguments (by using algorithms) that formalize moral reasoning based upon first-order normative theories such as utilitarianism or Kantianism (Klincewicz, 2016: 185) (Klincewicz, 2017).

Firstly, let us imagine that the telepresence robot in the case of Mr. Quintana is improved by having been programmed with a Kantian first-order normative theory – namely, to treat humans as ends in themselves. Would it change something in respect to the persuasiveness of moral reasoning at all? According to Kantian moral theory, the telepresence robot should be instructed to say that patient's life is precious, since the patient is an end in himself or herself as a human being. However, this also means that the patient's death should be evaluated as a great loss which would cause the family members deep and long-lasting sorrow. Even if the robot is designed to assist humans in Kant's sense, would it meet the moral expectations of both the relatives and the patient, as well as offer them consolation in moral terms?

There are two possible scenarios, at least. Firstly, when a telepresence robot and humans "share" a Kantian first-order moral theory. However, it does not follow that the patient would be more easily persuaded by the robot. The reasons for the potential lack of persuasion vary. For instance, since according to a Kantian moral theory, machines (robots) are not considered as ends in themselves, the patient may continue insisting on getting the news from another person, i.e. a 'human' doctor. Secondly, the patient may share a Kantian first-order moral theory, but 1) refuse to apply it to this situation because there are some other motives behind his or her moral expectations (such as some important relationships with others, some objectives which he or she has not managed to fulfill yet etc.) 2) it could be Kant's theory that makes the patient feel reluctant to accept the fact that he or she is dying.

On the other hand, let us imagine that the robot is designed to work according to a utilitarian first-order normative theory which is based on minimizing one's suffering and maximizing one's well-being. Then, however, the example should be narrated in a completely different manner. Taking into account that certainly, informing one about the fact that one is terminally ill will not increase one's well-being, the question is whether or not one should be informed about one's condition at all. The answer to this question goes beyond the prerogatives of machine ethics.

As shown by the example with Mr. Quintana, it requires specifying who should be informed, as well as how, depending on the agent or agents who are supposed to obtain informed consent, some decisions, in both medical and moral terms (e.g. ones concerning palliative sedation), should be made. Furthermore, in light of the discrepancy between moral mistakes and computation errors, [7] the following conclusion drawn by Wallach and Allen raises the old concerns in a new voice. They argue that moral dilemmas are getting even more complicated in the field of machine ethics, since a system that gets 90 percent of the decisions correct using only ethically blind criteria, "may be missing the 10 percent of cases where ethical judgment has the greatest significance to all involved" (Wallach and Allen, 2010: 42).

As the anonymous blogger cogently points out, medicine "is not solely scientific or computerized profession. The artistry of the clinician to interpret the needs of the patient in front of them, and to accommodate those needs in a caring, appropriate and individual way is not something we have yet succeeded in programming into a robot. Let's celebrate that, not try to replace it" (Anonymous, 2019).

3. Conclusion

Extrapolating Kantian and utilitarian first-order normative theories to the case of the telepresence robot shows that some crucial problems occur. Specifically, the hypothetical scenarios of introducing Kantian and utilitarian theories in the field of machine ethics demonstrate that even if the users of a given moral-reasoner system share the moral principles which are implied within the system in question, the effect of persuasion is not necessarily predictable in moral terms. Investigating the consequences of applying those theories reveals that one's strive for fulfilling the more realistic objective, that of so-called weak 'moral' AI scenarios, does not make the task any less complicated. This is demonstrated by Mr. Quintana's case when the robot functions as an implicit ethical agent. The case is an illuminative representation of how elaborating upon the criterion of reason-responsiveness may have the opposite effect. The latter displays the concern about the determination of the process of persuasion as morally justifiable.

The other alternative is to assume that the telepresence robot is like any other robot which meets the requirements of what Moor calls ethical impact agents, viz. "those agents whose actions have ethical consequences whether intended or no" (Moor, 2009). If so, however, it would mean that the supporters of telemedicine do not have well-grounded reasons to argue for the moral reliability of the particular use of telepresence robots.

Based on the analysis I have conducted so far, I would argue that the problem with answering some crucial moral questions within the field of machine ethics is that they require the following specifications to be made. Firstly, those questions should be both raised and answered by humans as being autonomous moral agents (full ethical agents in Moor's sense). If the questions are solved in a satisfactory manner, then they should be 'adapted' to the additional difficulties regarding the attempts at reaching a certain level of ethical (self-updating) knowledge in the field of machine ethics.[8] In this context, I have reached the conclusion that one of the reasons for the impossibility of building explicit ethical agents comes from the fact that we have not successfully built implicit ethical agents yet.

Furthermore, I have drawn the conclusion that since humans are still struggling to solve the problem of moral enhancement on intra- and inter-human levels, the hope that the difficulties in the field of machine ethics can be overcome by generating a better software should be diminished. The discouragement as such is a 'natural' (in the sense of being both

logically and morally justifiable) result of developing so-called Pyrrhonian skepticism. Analyzing the impact of the latter on the field of machine ethics means that one should critically reflect upon the idea of having certain knowledge about strong and weak 'moral' AI scenarios. Otherwise, humanity risks ending up with what Dellot and Wallace-Stephens call an AI bubble: a bubble with machines that are more artificial than they are intelligent (Dellot and Wallace-Stephens, 2017); and more importantly, with machines whose moral self-update is doomed to fail.

NOTES

[1] Some attempts in this field concern the distinction between strong and weak AI, as related to that between different types of values (Wallach and Allen, 2010: 192-193). Wallach and Allen examine the differences between weak and strong AI by referring to the development of so-called functional and operational morality (Ibid.: 74, 103, 111-112). In this context, both types of morality are evaluated from the perspective of so-called top-down and bottom-up approaches (Ibid.: 79-81).

[2] For the role of the nationally determined performative potential of so-called sociotechnical imaginaries, see Jasanoff and Kim, 2013.

[3] As an illustration of why rational responsiveness is only a necessary condition for achieving moral responsiveness, I refer to Ms. Willharm's response to the situation with her grandfather: "I just don't think that critically ill patients should see a screen. It should be a human being with compassion" (Jacobs, 2019).

[4] See Serafimova, 2020; Serafimova, 2020a: 114-115.

[5] For the role of utilitarian moral arithmetic within the field of machine ethics, see Serafimova, 2020.

[6] For more details about the difficulties in building agent-agent trust, see Serafimova, 2020a: 115-116.

[7] See Serafimova, 2020.

[8] For the challenges regarding the moral self-update of AMAs and AAMAs, see Serafimova, 2020.

REFERENCES

Anderson, M. and S. L. Anderson. (2007). Machine ethics: creating an ethical intelligent agent. *AI Magazine* 28 (4), pp. 15-26.

Anonymous. (2019). Replacing: Dr. Robot, Mr. Quintana and the Lord of the Rings. *The Re-humanising Revolution*. https://re-humanising.co.uk/2019/04/02/replacing-dr-robot-mr-quintana-and-the-lord-of-the-rings/

Arkin, R. C. and J. Borenstein. (2016). Robotic nudges: the ethics of engineering a more socially just human being. *Science and Engineering Ethics* 22 (1), pp. 31-46. doi: 10.1007/s11948-015-9636-2.

Becker, R. **(**2019). Why hospitals shouldn't use telepresence robots to deliver devastating news. *The Verge*. https://www.theverge.com/2019/3/11/18256929/telepresence-robot-hospitals-patient-information-medical-news-ethics-kaiser

Cusano, D. (2019). A telemedicine 'robot' delivers end of life news to patient: Is there an ethical problem here, Kaiser Permanente? http://telecareaware.com/a-telemedicine-robot-delivers-end-of-life-news-to-patient-is-there-an-ethical-problem-here-kaiser-permanente/

Dellot, B. and F. Wallace-Stephens. (2017). *What is the difference between AI & robotics?* https://medium.com/@thersa/what-is-the-difference-between-ai-robotics-d93715b4ba7f

Doctor told California man he was dying via robot video call. 2019. CBS News. http://www.kcfj570.com/2019/03/09/doctor-told-california-man-he-was-dying-via-robot-video-call/

Giubilini, A. and J. Savulescu. (2018). The artificial moral adviser. The "ideal observer" meets artificial intelligence. *Philos Technol* 31(2), pp.169-188. doi: 10.1007/s13347-017-0285-z.

Hansson, S. O. (2007). The ethics of enabling technology. *Cambridge Quarterly of Healthcare Ethics* 16(3), pp. 257-267.

Jacobs, J. (2019). Doctor on video screen told a man he was near death, leaving relatives aghast. *New York Times*. https://www.nytimes.com/2019/03/09/science/telemedicine-ethical-issues.html.

Jasanoff, S. and S.-H. Kim. (2013). Sociotechnical imaginaries and national energy policies. *Science as Culture* 22 (2), pp. 189-196. https://doi.org/10.1080/09505431.2013.786990

Klincewicz, M. (2016). Artificial intelligence as a means to moral enhancement. *Studies in Logic, Grammar and Rhetoric* 48 (61), pp. 171-187.

Klincewicz, M. (2017). Challenges to engineering moral reasoners: time and context. In: P. Lin, K. Abney, R. Jenkins (eds.). *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence.* Oxford: Oxford University Press, pp. 244-257.

Klincewicz, M. and L. Frank. (2018). Making metaethics work for AI: Realism and anti-realism. In: M. Coeckelbergh, J. Loh, M. Funk, J. Seibt, M. Nørskov (eds.). *Envisioning Robots in Society – Power, Politics, and Public Space, Proceedings of Robophilosophy*. TRANSOR 2018. Series: Frontier in Artificial Intelligence and Applications, Amsterdam, IOS Press. pp. 311-318.

Moor, J. H. (2006). The nature, importance, and difficulty of machine ethics. *IEEE Intelligent Systems* 21 (4), pp. 18-21.

Moor, J. H. (2009). Four kinds of ethical robots. *Philosophy Now. A Magazine of Ideas* 72 https://philosophynow.org/issues/72/Four_Kinds_of_Ethical_Robots.

Nickel, P. J. (2013). Trust in technological systems. In: M. J. de Vries, S. O. Hansson, A. W. M. Meijers (eds.). *Norms in Technology: Philosophy of Engineering and Technology*, vol. 9. Dordrecht: Springer, pp. 223-237.

Savulescu, J. and H. Maslen. (2015). Moral enhancement and artificial intelligence. Moral AI? In: J. Romportl, E. Zackova, J. Kelemen (eds.) *Beyond Artificial Intelligence. The Disappearing Human—Machine Divide.* Springer, pp. 79-95.

Serafimova, S. (2020). Whose morality? Which rationality? Challenging artificial intelligence as a remedy for the lack of moral enhancement. *Humanit Soc Sci Commun* 7, 119 https://doi.org/10.1057/s41599-020-00614-8.

Serafimova, S. (2020a). How to assign weights to values? Some challenges in justifying artificial explicit ethical agents. *Balkan Journal of Philosophy* 12 (2), pp. 111-118.

Tingley, B. (2019). A robot, a dying man, and the cold, heartless future of AI medicine. https://mysteriousuniverse.org/2019/03/a-robot-a-dying-man-and-the-cold-heartless-future-of-ai-medicine/

Wallach, W. and C. Allen. (2010). *Moral machines: teaching robots right from wrong.* Oxford: Oxford University Press.