Leuphana Universität Lüneburg
Digitalisierung und Demokratie
Professor Doktor Alexander Weiß
Sommersemester 2019
31.07.2019

Rights for Robots: Should Sufficiently Intelligent Robots Receive Rights?

Collin Daniel Hooper
Matrikelnummer: 3029850
Student der Politikwissenschaft
Fachsemester: 6

Scharnhorststr. 1a WG13
21335 Lüneburg
collin.d.hooper@stud.leuphana.de

Table of Contents

1. Introduction

Technological advancements in the recent past have produced some amazing things that speak to the intensity of proliferation of technology by humans. Generally, immense technological gains are accompanied by discussions warning that humans use caution when applying new technologies to different problems. One need only think of the advent of nuclear technology; while a powerful source of energy that has the potential to change many different aspects of life, the destructive power of nuclear energy forced humankind to view the technology in a new light, and discussions regarding the morality and safety of the use of nuclear technology to solve problems intensified. One need only think about the Atomausstieg, in which certain countries such as Italy, Germany, and other European nations (though not all) decided to reduce reliance on nuclear energy over time. Not just in this case, but in others, the genesis of new technology often inspires debate regarding the future of the particular technology.

In recent years, much noise has been made about the possibilities of a new sort of technology, that is, artificial intelligence (from now on: AI). Much like nuclear technology, the constant reinforcement and improvement of the functionalities of AI have begun causing unrest in many different segments of the scientific community. In fact, "[…] our situation is deemed so worrisome, that several new research centers have been created with the explicit aim of reducing the potential dangers of AI." (Becker, 2018, p. 235). Branching off of this, the Fukoka World Robot Declaration (2004) claims:

> Confident of the future development of robot technology and of the numerous contributions that robots will make to Humankind, this World Robot Declaration is Expectations for next-generation robots: (a) next-generation robots will be partners that co-exist with human beings; (b) next-generation robots will assist human beings both physically and psychologically; (c) next-generation robots will contribute to the realization of a safe and peaceful society. (Fukoka World Robot Declaration, 2004)

With nations such as Japan and South Korea implementing robots in many social stations (Steinert, 2013), it makes sense for the purpose of this research paper to combine the terms robot and artificial intelligence. After all, "In general, the intelligence capabilities of robots follow the development path of artificial intelligence." (Tzafestas, 2018, p. 20). Of course, though the term robot is difficult to define, this paper uses the definition proposed by Robertson (2014): "A robot is an aggregation of

different technologies - sensors, software, telecommunication tools, actuators, motors, and batteries - that make it capable of interacting with its environment with some human supervision, like tele-operation, or even completely autonomously." (Robertson, 2014, p. 574).

Combining this information, the robots of today are equipped with artificial narrow intelligence (ANI), which allows them to carry out specific tasks very well, at a level that matches and even surpasses humans, such as Deep Blue's (an AI's) winning of a game of chess against the chess world champion, Garry Kasparov (Bostrom & Yudkowsky, 2014). That being said, under ANI, robots are unable to expand their functionality by themselves, and rely on close work with humans to carry out the tasks that they perform. However, it is not necessarily this particular AI that has the scientific community shaken. According to Tzafestas (2018), "By 2040, robots are expected to perform tasks compatible with 'artificial general intelligence' (AGI), i.e., they will be able to compete with humans across all activities, and perhaps convince humans that they are 'humans'." (Tzafestas, 2018, p. 20).

It is at this point that a discussion has been born regarding how humanity should conduct itself in the face of these new challenges. After all, it is important that these issues are addressed as well as they can be before the advent of these technologies, because the risks to humanity that stem from ever-improving AI, though they may seem very science fiction-esque, may come into realization, and making considerations about speculative future happenings could change the way that these technologies develop now, and it is the conviction of this research paper that making considerations about the future coexistence of AI/robots and humanity is an important use of time and resources, so that humanity might be prepared in the case of many possible events that could take place. This is all speculative, of course, but these concerns must be handled now, for in the future, it may be too late.

That all being said, it is now time to introduce the purpose and structure of this research paper. First, one must understand a portion of ethics to really grasp what this work wishes to convey. First of all, "Rights in human society are counterbalanced by the need for commensurate responsibilities and duties. Consequently, the idea of artificial intelligence and robot rights necessitates a matching level of societal responsibility and duty." (Ashrafian, 2015, p. 318). By this logic, the more that humanity cedes responsibilities and duties of human persons (be that what they may) to robots, the more these robots will be deserving of certain rights, just like any other thing

that contributes to society. One may here be thinking that it takes more than the ability to carry out societal duties to be eligible for rights in a society, and they would not be wrong. However, once robots are equipped with AGI, and are on the same cognitive level as normal humans (or even higher), is it so crazy to think that they would be deserving of rights? In another article discussing the possibility of rights for robots, the authors state: “No influential modern secular account is plausibly read as committed to a principle whereby two beings can differ in moral status but not in any psychological or social properties, past, present, or future, actual or counterfactual.” (Garza & Schwtizgebel, 2015, p. 101). To them “Purely biological otherness is irrelevant unless nome important psychological or social difference flows from it.” (Garza & Schwitzgebel, 2015, p. 107).

Summing up the information that has been presented, if robots were to one day reach the level of AGI, putting them at the same standing (if not superior) as humanity in the consideration for moral status, then it would be unethical to deprive deserving moral agents of rights. That being said, considering the concerns that are being made about robots and AI in the scientific community, as well as the discussions regarding the handling of robots in the future, this paper asks: Do robots deserve rights? And if so, which rights should robots possess? Which rights should robots be forbidden? Now, it is assumed that when robot rights are being discussed, this work is also referring to robots/AI systems that could be deemed as possessing AGI. As has been stated before, the arrival of these robots is for now a hypothetical scenario, and this work is written in the case that this point does come to pass.

This research paper will next look at robot rights and roboethics, which will detail considerations about the moral status of robots and the conferring of certain rights upon them. The section after that will discuss how one goes about making an ethical robot from a theoretical standpoint, and details specifications for robots’ conduct for peaceful coexistence in human society. The second to last section shall look into a specific part of the robot rights debate, specifically who carries responsibility for the actions of a robot. After that section, a discussion will bring the information presented in this research paper full-circle, and briefly detail the ins and outs of robot rights, as well as containing a brief, abstract recommendation for the development of robots in the future.

2. Rights for Robots and Roboethics

This section shall discuss the ins and outs of the moral status of a robot, as well as how these considerations tie in to roboethics. To do this, one must first understand a few key aspects regarding the history of incorporation of particular groups into society. See, “A look at the history of ethical considerations and disputes in moral philosophy shows that the circle of ethical recipients has been continuously expanded.” (Steinert, 2013, p. 4). Steinert (2013) goes on the state: “This inevitably leads to the question whether the moral realm might someday also encompass advanced machines like robots.” (Steinert, 2013, p. 4).

Now, it may seem strange to observe robots in this way, as this is not the lens through which we view robots today, as they are not nearly as advanced as the robots that are discussed in the literature regarding rights for robots and roboethics. However, when one thinks about the non-personhood status that was historically conferred on undesirable or second-class groups, such as blacks under segregation and slavery in the United States, or Jewish people in the 1930s and 40s in Europe, for example, one could see how this circle has been expanding since the dawn of civilization, including more and more individuals into the fray of individuals possessing a moral status. As robots become more sophisticated, the literature states that it will be more and more unethical to deny them certain rights.

Now, just what this moral status is must also be understood, so that one is not confused as to why robots may one day be deserving of certain rights. Garza & Schwitzgebel (2015) state:

> We submit that as long as these artificial or non-Homo-sapiens beings have the same psychological properties and social relationships that natural human beings have, it would be a cruel moral mistake to demote them from the circle of full moral concern upon the discovery of their different architecture or origin. (Garza & Schwitzgebel, 2015, p. 106-7).

See, these authors follow the argumentation that it is on the inside what matters. That being said, these authors do convey the idea that denying a robot certain rights based on the way it looks or based on something other than its social/psychological characteristics is comparable to a sort of racism. Of course, it was this realization that humans are fundamentally equal that started the incorporation of marginalized groups into the discussion for civil rights. It is posited that once there is nothing that separates the social/psychological characteristics of a normal human and a normal

robot, then robots will be deserving of rights, because they will have a moral status on the same level as that of a human.

Another dimension that must briefly be talked about in the run up to a moral status for robots is the fact that as the responsibilities and duties of robots in society increases, then the robots will be deserving of more rights. The best way to explain this is once again with the non-personhood example that was mentioned earlier. Though harsh, once marginalized groups were objectively carrying out societal duties, generally for the benefit of the greater good, they began receiving rights in small increments. It is thought that robot development will parallel this trajectory, that is, the more responsibilities and duties robots perform, or better said, the more society relies on robot capabilities, the stronger the moral status of the robot, and the more the robot will be deserving of rights (Ashrafian, 2014).

Yet another facet that must be covered in this research paper is the instrumentality of robots in general. Generally, robots are thought of as doing that which humans are either incapable of doing, or unwilling to do. This has lead talking heads in the roboethics community to claim "[…] one could argue that the continued development of AI as systems that are entirely subjected to our every wish and command would amount to reintroducing slavery." (Becker, 2018, p. 236). After all, the word robot stems from the Czech word *robota*, which means servitude or forced labor (Robertson, 2014, p. 573). When one thinks about it, giving robots the tasks and duties that humanity does not wish to do, due to difficulty or laziness or something else, is comparable to slavery, especially if the robots share a level of moral status with humanity. One could think about a nightmarish situation where robots get sick of doing human work, and decide to put down their wrenches and pick up weapons, however, that is just a little too much science fiction for this research paper. Of course, one might now ponder the entire usefulness of robots. That being said, "The primary objective of roboethics is to motivate the moral design, development, and use of robots for the benefit of humanity." (Tzafestas, 2018, p. 3).

It cannot be denied that intelligent robots are on their way, though to what extent this intelligence goes cannot at this time be known. Having conveyed that, it is important to realize that at that time when the Fukoka World Robot Declaration comes into realization, "Rules would […] be needed for a world where some intelligent players are machines." (Risse, 2019, p. 9). Risse (2019) goes on to state that "It is not impossible that, eventually, the Universal Declaration of Human

Rights would have to apply to some of them." (Risse, 2019, p. 9). According to one professor in the field of roboethics: "If the machine had the capacity to feel pain, if it had a psychological awareness that it was a slave, then we would want to extend rights to the machine." (Tzafestas, 2018, p. 21). Jo Bell, an animal rights activist, says "As we have incorporated other races and people - women, the disabled into the category of those who can feel and think, I think if we had machines of that kind, then we would have to extend some sort of rights to them." (Tzafestas, 2018, p. 22). With the information presented within this section, it should be clear to a reader that the advent of AGI-equipped robots will pose many not-easy-to-answer questions in the future, and it is advantageous that these aspects are being discussed and debated today, so that humanity has a head start on the moral and ethical development of robots deserving of a moral status on par with that of a human's.

Robots will likely deserve considerations regarding their moral status in the future, due to the ever increasing scope of the circle of ethical recipients, as well as humanity's increase of exportation of jobs/responsibilities/duties to robots. That being said, there is still much to be discussed regarding robot rights, starting with the making of a robot that deserves moral status. This information shall be presented in the next section.

3. The Making of a Robot Deserving of Moral Status

Not just any robot is worthy of moral status consideration. "At minimum, moral/ethical robots need to have: (i) the ability to predict the results of their own actions or inactions, (ii) a set of ethical rules against which to evaluate each possible action/consequence; and (iii) a mechanism for selecting the most ethical action." (Tzafestas, 2018, p. 4). This sets a rather high bar for moral status consideration of robots, and weeds out robots like the Roomba for consideration of a moral status. These requirements cover aspects of many branches of the scientific community, and "It is emphasized that for roboethics to be assured, the joint commitment of experts of different disciplines […] to design ethics-based robots, and adapt the legislation to their issues (technological, ethical) that arise from the continuous advances and achievements of robotics, is required." (Tzafestas, 2018, p. 2-3). The disciplines required for roboethics are diverse, and include the technical side of robotics, such as electrical, mechanical, and computer engineers, and the less technical side regarding the intelligence structure that involves the work of psychologists, philosophers, and artificial intelligence scientists.

Of course, the making of robots differs from place to place, as well as the ideas that capture the meaning of what a robot is and what a robot should do. Take, for example, Russian-American writer Isaac Asimov, author of books such as *I, Robot* and *Runaround*, who posed a global set of laws regarding the development of robots on Earth. One could think of these laws as priorities that robots must follow at all times. Though many iterations of these rules have been proposed, the original three laws (there are now four) capture a humanitarian view of intelligent robots and their capabilites, as well as what they should be allowed and forbidden to do. To understand these priorities, one need no more than read them in order:

Law 1: A robot may not injure a human being or, through inaction allow a human being to come to harm.

Law 2: A robot must obey orders that it receives from human beings except when such orders conflict with Law 1.

Law 3: A robot must protect its own existence as long as such protection does not conflict with Laws 1 and 2.

A law was later added by Asimov which takes priority over all the laws aforementioned. This is called Law 0. Law 0 states: “No robot may harm humanity or through inaction allow humanity to come to harm.” (Asimov, 1950, Tzafestas, 2018, p. 7).

These laws aim at the conduct of humans and robots as they coexist within a society together. These laws, briefly explained, put humanity first, and robots second, and attempt to create the framework for a world in which robots are the instruments of humanity. That being said, these laws are designed to protect humanity from robots/AI by setting the priority of the protection of humanity and humans as a robot’s highest goal. These laws also make some particular future developments impossible, such as the creation of an autonomous killbot that could be used by a government for times of war, or to silence political enemies and dissenters. The thought of using robots in warfare is troubling for a few reasons. First and foremost, for the countries or groups that are able to construct autonomous war robots, the human cost of waging war would sink drastically, making war a more viable option. When one does not have to worry about the death of soldiers and the effects that war has on families or a nation, it is likely that the frequency of war would increase. Regarding Asimov’s laws, the creation of a killbot would not be possible, and humanity would seemingly be better off because abiding by them.

Of course, Asimov is by far not the only author to have proposed something of a sort of code of ethics for robots. Tezuka Osamu, a physician who pursued a career in science fiction (and created Astro Boy), produced laws regarding the conduct of robots in society as well, however, Tezuka took a much different approach compared to Asimov, and wrote ten laws. Because of the difference in contemplations regarding robot conduct, these laws shall be listed here as well:

1. Robots must serve humankind.
2. Robots shall never kill or injure humans.
3. Robots shall call the human who creates them 'father.'
4. Robots can make anything, except money.
5. Robots shall never go abroad without permission.
6. Male and female robots shall never switch roles.
7. Robots shall never change their appearance or assume another identity without permission.
8. Robots created as adults shall never act as children.
9. Robots shall not assemble other robots that have been discarded by humans.
10. Robots shall never damage human homes or tools. (Schodt, 2007, p. 108).

As one can see, these laws proposed by Tezuka in the early 1950s differ significantly from those proposed by Asimov. For one, there are more of them that regard a robot's conduct than by Asimov, as well as some laws that one may question the validity and purpose of, such as laws 3, 5, and 8. Of course, there is an explanation as to why these two sets of laws for robots differ. The reason for this is the fact that Asimov's laws aim at the global regulation of the conduct of robots, while Tezuka's laws "[…] are synchronized with dominant Japanese social values and address the integration of robots into human (and specifically Japanese) society where they share familial bonds of kinship and perform familial roles." (Robertson, 2014, p. 584).

Summarily, "Both Tezuka and Asimov presaged the integration of robots in everyday life and work, and both drew up laws regulating human-robot interactions that have shaped current debates among roboticists, philosophers, and the public at large." (Robertson, 2014, p. 583). Obviously, rules of conduct between humans and robots will have to be put in place at some point. This begs the question of whether these rules of conduct will be global, or regional, or national? This particular question must be dealt with in the time leading up to the development of robots deserving a moral status, and deserves much contemplation on the side of humanity.

One thing that the Asimov and Tezuka seem to agree on is the fact that robots must serve humans. Ashrafian (2014) writes: “Whilst all their socio-cultural merits should be celebrated, the fundamental partiality to favor human needs should be maintained by artificial intelligence agents and robots.” (Ashrafian, 2014, p. 323). Another author complementarily states: “These intelligent players would have to be designed so they protect human rights even though they would be smart and powerful enough to violate them.” (Risse, 2019, p. 9).

Having read all of that, it seems that there is some consensus when it comes to the human need for safety from robots, and an assurance of that fact as well. Apart from that, at least between Asimov and Tezuka, robots are largely considered instrumental, to be used by humanity for the gains of humanity. Still, many question remain open. Thinking back to the autonomy of robots, if they ever achieve autonomy, who is responsible for the actions of a robot? Is it the robot itself, is it the creator of the robot, or the group that created the robot? Or none of the above, or a mixture? This question shall be handled in the next section, which will detail discussions regarding the responsibility of the actions of a robot.

4. Who is Responsible for the Actions of a Robot?

This section shall attempt to provide context and discuss exactly who is responsible for the actions of a robot. As one may have been able to guess, there are a few different viewpoints within the scientific literature regarding the subject.

To begin, Alsegier (2016) writes: “People have the right to be safe, while at the same time, engineers and companies have the right to produce and profit from the development of robotic technology.” (Alsegier, 2016, p. 24). While both of these statements are true at the moment, there does seem to be a problem regarding what a robot does and who is responsible for it within society. Companies and manufacturers of robots/AI have the prerogative to profit from what they do, i.e., make robots and artificial intelligence. That being said, the prerogative is in opposition to the safety of a population. Though it is speculative, it should not be assumed that robots will integrate into societies without having a few integration problems. What these problems could look like is hard to imagine before the fact, however, it would be advantageous to answer this question before the integration of AGI-equipped robots into society. Alsegier (2016) goes on, stating: “[…] Nagenborg states that robots are products that will be placed in society and interact with people, so people have

the right to a set of rules that will control the market and use of these products." (Alsegier, 2016, p. 24).

Another reason that this question receives importance is the fact that "Individuals who live in a society have human rights that must be respected. The unguided development of robots puts many human rights in danger…" (Alsegier, 2016, p. 27). Again, the hypotheticals of the development of robots equipped with AGI tallies to an uncountable number, but a set of rules that regards the development of AI/robots does not seem so backwards. Also, basic sets of rules regarding human-robot interactions would also be necessary and helpful for the fluid running of society after the integration of robots.

The last point to be made in this section provides nuance to the information that was just presented. See, "According to [Roger K. Moore, a professor of computer science,] the problem is not the robots taking over the world, but that some people want to pretend that robots are responsible for themselves." (Tzafestas, 2018, p. 3). This viewpoint places the burden of responsibility on the humans that developed the robot/AI in the first place. He goes further to state: "Whoever owns and operates a robot is responsible for what he does." (Tzafestas, 2018, p. 4). In an abstract way, Moore is stating that humanity is responsible for the actions of robots, due to the fact that humans make robots to perform certain actions, and that humanity has the responsibility to deal with these hypothetical ethical questions before it is faced with real-time ethical problems. On this point, Alsegier (2016) writes: " Technology cannot be responsible in and of itself, but people must be responsible at its creators and controllers." (Alsegier, 2016, p. 28).

Again, this author states that it is ultimately up to humanity, and rather, the humans that work directly in the field of developing robots/AI, to try and consider all of the possibilities that become realized when fully functioning, intelligent robots become a reality. That being said, it is up to the human forces behind these developments to assure that ethical robots are made with humanity's best interest in mind.

5. Conclusion

Though much of this research paper bases itself off of hypothetical scenarios that may stem from humanity's intense technological proliferation, the problems discussed in this work have had

an abundant amount of ink spilled over them in literature regarding robots, artificial intelligence, and their likely future coupling. That being said, it is time to answer the research questions that were posed in this work.

To answer the question of whether robots deserve rights, one must take a few things into account. The first, and most important thing to remember in this regard is the fact that robots can and will vary in intelligence, duty, appearance, and in probably every other way as well. That being said, that means that while some robots will hypothetically be eligible for certain rights, this is not a generalizable characteristic that all robots will possess. At first, granting rights to robots will likely be a case-by-case ordeal. Having communicated this, to answer the question, it seems that robots will indeed be eligible for certain rights some day, given that they are intelligent enough and carry out a proper amount of societal duties in order to be eligible for those rights. One question that remains open in at what level rights for robots will be made. Will different countries have different rights for robots? Will an international organization be made to deal with this future scenario, should it ever come to pass? The generality of rights for robots is an important topic that deserves more exploring.

The second question, which inquires as to what rights a robot should possess, is a bit more abstract, and more difficult to answer, due to the fact that there is still much to be discussed in this context. Though it is difficult to pinpoint exactly what rights a robot should and should not have, it would be advantageous to humanity if robots/AI possessed rights that aid in the benefitting of humanity. This is because of the fact that if humanity benefits from the contributions made by robots, the robots will not only solidify their standing in human society, but also incrementally be eligible for more full rights.

Lastly, answering the question of what rights intelligent robots should be denied seems to be easier from a protecting-of-humanity perspective, which this research paper advocates. Robots should be denied the right to harm humans, and humanity, just as Asimov and Tezuka posited decades ago. When one takes into account all forms of harm, in this case physical and psychological, and bans robots from being able to perform these harmful actions on humans, one can easily see how humanity would be better off than the alternative, in which robots do have the ability to harm humans/humanity.

As for other rights such as those listed in the Universal Declaration of Human Rights, humanity will have to possibly reconsider some of the wording in some of the rights so that robots may be taken into account one day. For example, Article 4 states: “No one shall be held in slavery or servitude; slavery and the slave trade shall be prohibited in all their forms.” (Universal Declaration of Human Rights, 1948). Questions will be raised regarding the personhood of robots, just as the literature states, and the burden is on humanity to define and refine what it means to have rights, so that they may be able to apply to certain machine entities in the future. With the scientific community worrying about happenings such as the singularity, which is described as the point when machines become smarter than humans with the help of artificial intelligence (the singularity could also be understood as the genesis of artificial general intelligence), it is important that these questions and the questions related to the literature in this work be discussed within society, especially at the level where these intelligent machine entities are being produced, but also at the other levels, so that the integration of intelligent robots into society can take place as smoothly as possible.

Bibliography

Alsegier, R. (2016). *Roboethics: Sharing our world with humanlike robots.* IEEE Potentials, Vol. 35, 1. 24-28.

Ashrafian, H. (2014). *Artificial Intelligence and Robot Responsibilities.* Science and Engineering Ethics, Vol. 21, 2. 317-326.

Ashrafian, H. (2014b). *AIonAI: A Humanitarian Law of Artificial Intelligence and Robotics.* Science and Engineering Ethics, Vol. 21, 1. 29-40.

Asimov, I. (1950). *I, Robot.*

Becker, S. (2017). *AAAI: An Argument Against Artificial Intelligence.* In: Philosophy and Theory of Artificial Intelligence 2017. Ed. Müller, V. C. 235-247.

Bostrom, N., & Yudkowsky, E. (2014). *The ethics of artificial intelligence.* In: The Cambridge Handbook of Artificial Intelligence. Cambridge University Press. 316-334.

Fukoka World Robot Declaration. (2004). Japan.

Garza, M., & Schwitzgebel, E. (2015). *A Defense of the Rights of Artificial Intelligence.* Midwest Studies in Philosophy, Vol. 39, 1. 98-119.

Risse, M. (2019). *Human Rights and Artificial Intelligence: An Urgently Needed Agenda.* Human Rights Quarterly, Vol. 41, 1. 1-16.

Robertson, J. (2015). *Human Rights Vs Robot Rights: Forecasts from Japan.* Critical Asian Studies, Vol. 46, 4. 571-598.

Schodt, F. L. (2007). Translation of: *Mishu purodakushon shiryōshū (1962-73).*

Steinert, S. (2013). *The Five Robots - A Taxonomy for Roboethics.* International Journal of Social Robotics, Vol. 6, 2. 249-260.

Tzafestas, S. G. (2018). *Roboethics: Fundamental Concepts and Future Prospects.* Information, Vol. 9.

United Nations. (1948). *Universal Declaration of Human Rights.*

Declaration of Originality

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

Lüneburg,
31 Jul. 2019

Signature