A Framework for Grounding the Moral Status of Intelligent Machines

Michael R. Scheessele
Computer Science and Psychology
Indiana University South Bend
South Bend, IN, USA
mscheess@iusb.edu

ABSTRACT

I propose a framework, derived from moral theory, for assessing the moral status of intelligent machines. Using this framework, I claim that some current and foreseeable intelligent machines have approximately as much moral status as plants, trees, and other environmental entities. This claim raises the question: what obligations could a moral agent (e.g., a normal adult human) have toward an intelligent machine? I propose that the threshold for any moral obligation should be the "functional morality" of Wallach and Allen [20], while the upper limit of our obligations should not exceed the upper limit of our obligations toward plants, trees, and other environmental entities.

CCS CONCEPTS

Computing methodologies~Philosophical/theoretical foundations of artificial intelligence

KEYWORDS

Machine Ethics

ACM Reference format:

Michael R. Scheessele. 2018. A Framework for Grounding the Moral Status of Intelligent Machines. In Proceedings of 2018 AAAI/ACM Conference on AI, Ethics, and Society (AIES'18), February 2-3, 2018, New Orleans, LA, USA. ACM, New York, NY, USA. https://doi.org/10.1145/3278721.3278743

1 Intelligent Machines with Moral Status?

Intelligent machines will influence, improve, and impinge on our moral environment. This has roused some in the AI community, who realize that issues go beyond mere safety concerns addressable by felicitous engineering. For example, Wallach and Allen [20] have proposed artificial moral agents (AMAs), intelligent machines intended to function as if they have moral obligations to us. But what about the converse—the moral obligations that we could have to intelligent machines? Gunkel [8] laments the lack of attention given to this question. This asymmetry is especially glaring given that, in Western philosophy, if an entity has moral obligations to others—if an entity is a moral agent—then other moral agents have moral obligations to that entity. That is, the entity is also a moral patient and has moral status that other moral agents ought to take into account when acting with respect to the entity. Naturally, AMAs are not actual moral agents. Even so, might we have moral obligations to sufficiently intelligent machines? Of course, the question lurking here is whether an intelligent machine could have moral status.

The purpose of this article is to establish the possibility of moral status for current and readily foreseeable intelligent machines. Specifically, I propose a map of moral status derived from moral theory and show where current and foreseeable intelligent machines should be located in this map. This first requires providing some definitions and assumptions for the arguments that follow.

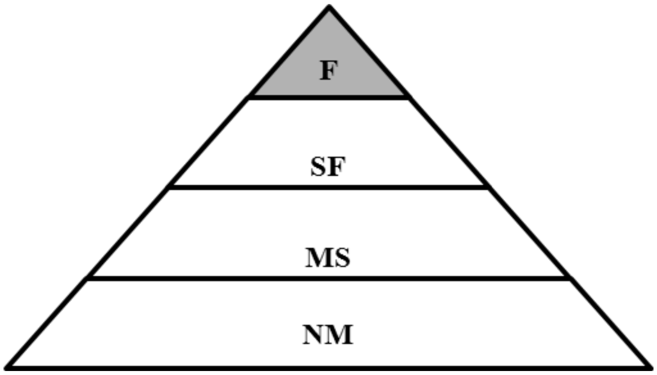

Figure 1: The moral status pyramid (**MSP**). **F**: full moral status; **SF**: significant-full moral status; **MS**: minimal-significant moral status; **NM**: negligible-minimal moral status. **Gray region** represents entities that are both moral agents and moral patients. **White regions** represent moral patients only. **Moral agency** is the threshold for region **F**; **sentience** for region **SF**; having **a good of its own** for region **MS**.

AIES '18, February 2–3, 2018, New Orleans, LA, USA
© 2018 Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-6012-8/18/02…$15.00
https://doi.org/10.1145/3278721.3278743

2 Definitions and Assumptions

This article employs the definition of moral status, due to Jaworska and Tannenbaum [9]: "An entity has moral status if and only if it or its interests morally matter to some degree for the entity's own sake, such that it can be wronged."

Regarding usage of intelligent machine in this article, although humans and animals are sometimes considered machines, they are excluded for present purposes. Similarly, humans physically augmented by wearable, implantable, or attachable technologies, as well as humans working cooperatively with machines in a tightly integrated group, are not considered here. Roughly, intelligent machine, as used here, refers to a reasonably sophisticated product of artificial intelligence (AI) or related disciplines. An intelligent machine may be prominently hardware, as with a robot or digital computer, or primarily software, as with a virtual agent or software-based system. An intelligent machine may stand alone or be embedded in another artifact. It may be a collective or swarm. It may be silicon-based or not—as with products from the field of synthetic biology. Further, an intelligent machine, for purposes of this article, could be a hybrid of two or more materials—a silicon-based digital computer interfaced with a neural circuit made from biological material, perhaps. With this definition of moral status and a brief clarification of the intended use of the term intelligent machine, I give three assumptions in support of ensuing arguments.

First, I assume that it is at least possible that an intelligent machine could have more than mere instrumental value. The instrumental value of an entity is derived only from its usefulness in achieving the goals of some other agent (e.g., a normal adult human). Many are opposed to or unconvinced of the idea that a machine could have more than instrumental value, rejecting the notion that a machine could also have intrinsic value—value for its own sake whether others actually value it or not. Although intelligent machines to date have done little to undermine the instrumental value view, Gunkel [8] observes that the increasing autonomy of intelligent machines does challenge this view:

> In other words, the instrumental definition of technology, which had effectively tethered machine action to human agency, no longer applies to mechanisms that have been deliberately designed to operate and exhibit some form, no matter how rudimentary, of independent or autonomous action. (p. 36)

The reason is that such mechanisms:

> …directly contravene the instrumental definition by deliberately contesting and relocating the assignment of agency. Such mechanisms are not mere tools to be used by human agents but occupy, in one way or another, the place of agency. (p. 32)

Those not swayed by Gunkel's argument will likely also be unconvinced that an intelligent machine could have moral status.

The second assumption is that moral status is not all-or-none, but rather that there are degrees of moral status. Jaworska and Tannenbaum [9] describe how either a threshold approach or a scalar approach may ground moral status that comes in degrees and explain how each approach has its drawbacks. For example, the threshold approach typically requires that an entity have some capacity in order to meet the threshold for some degree of moral status. This can lead to cases where an entity possessing that capacity, but using it poorly, could meet the threshold, while another entity which lacks that capacity entirely, but which may have a similar capacity and use it well, could fall short of the threshold, leaving a "gap" in the moral status between the two entities. Jaworska and Tannenbaum point out that it is possible to narrow a gap by having multiple thresholds leading to different degrees of moral status. The scalar approach avoids the "gap" problem of the threshold approach, but as they explain, it could lead to counterintuitive cases where it is more wrong to harm an intelligent person than a not-so-intelligent person. In this article, I use the threshold approach.

Now, consider Figure 1. The pyramid suggests that there is a hierarchical ordering of the thresholds, where an entity meeting a higher threshold in the pyramid has a greater degree of moral status than an entity meeting only a lower threshold. In addition, if an entity meets or exceeds a given threshold, it will also meet/exceed all lower thresholds in the pyramid. This hierarchical-ordering-of-thresholds is the third assumption.

3 Framework for Assessing the Moral Status of Intelligent Machines

Figure 1 depicts a proposed map of moral status derived from moral theory. This map is divided into four regions: (1) F for full moral status; (2) SF for significant-full moral status; (3) MS for minimal-significant moral status; (4) NM for negligible-minimal moral status. Each region requires that an entity meet (or exceed) a certain threshold in order to be included into that region. According to this scheme, an entity must fall into exactly one of the four regions; specifically, an entity will fall into the region with the highest threshold that the entity meets or exceeds. Due to the hierarchical-ordering-of-thresholds assumption, an entity that meets a particular threshold will also meet all of the lower thresholds in the pyramid of Figure 1. For example, if an entity were to meet the moral agency threshold (i.e., the entity is an actual moral agent), the entity would meet the sentience and the "a good of its own" thresholds as well. The moral status pyramid (MSP) of Figure 1 clearly shows that the higher a threshold attained by an entity, the higher the degree of moral status accorded that entity. The MSP also illustrates the obvious fact that the higher the threshold, the fewer the entities capable of attaining that threshold.

Less obvious may be the observation that MSP very roughly tracks the anthropocentrism/nonanthropocentrism distinction of philosophers such as Sterba [18] and the moral agent/moral patient distinction of philosophers such as Gunkel [8]. By "very roughly" I mean that the MSP is not a high resolution rendering of the many theories and myriad nuanced arguments about moral status from the moral theory literature.[1] Rather, the objective here is to ground the claim that an intelligent machine could have moral status and to argue how much moral status such a machine merits.

4 Determining the Moral Status of Intelligent Machines Using the MSP

4.1 Moral Agency

I use moral agency as the threshold for classifying an intelligent machine as having full moral status (region F in Figure 1). This is not the functional moral agency (or "functional morality") intended for AMAs [20]. Rather, it is the so-called "full blown" moral agency of moral theory. Moral agency is a murky concept. In considering the question of machine moral agency, Gunkel [8] observes:

> What has been discovered is that the concept of moral agency is already so thoroughly confused and messy that it is now unclear whether we—whoever this "we" includes—are in fact moral agents. What the machine question demonstrates, therefore, is that the question concerning agency, the question that had been assumed to be the "correct" place to begin, turns out to be inconclusive. (p. 91)

Even so, we begin with moral agency. We need a reasonably precise, but still broad, definition of moral agency. The idea is to be as charitable to intelligent machines as possible in choosing this threshold. Thus, an intelligent machine need not possess the rationality, capacity to will, and presupposition of freedom of a Kantian [10] moral agent. Nor must it possess the rationality, consciousness, intentionality, and free will of the "full ethical agent" of Moor [14]. Deliberately setting the bar of moral agency relatively low, as compared to other moral theories, will show that current and readily foreseeable intelligent machines still cannot clear this bar and thus attain full moral status.

The threshold to region F in Figure 1 is based on the work of Gert and Gert [7]. They propose a "schema for definitions of 'morality'" as follows: "…morality is the *informal public system* that all rational persons, under certain specified conditions, would endorse." A public system is: "a system of norms (1) that is knowable by all those to whom it applies and (2) that is not irrational for any of those to whom it applies to follow (Gert 2005: 10)" [7]. Further, in an informal public system, there is no authority and no decision procedure for specifying exactly what to do in all situations. For Gert and Gert then, it appears that a moral agent must be a "rational person" able to "endorse" the informal public system "under certain specified conditions." The ability to meet these criteria is the formulation of "moral agency" used as the threshold to region F in the Moral Status Pyramid.

Note that their schema for morality is very broad—Gert and Gert [7] intend this schema, not as a moral theory in itself, but as a source of normative definitions of morality that in turn can serve as the bases of a variety of moral theories. They argue that their schema adequately captures conceptions of morality as diverse as Kant's and Mill's. In addition, Gert and Gert explain how natural law theory (as well as some Divine Command theories based on natural law theory) and some versions of virtue theory are covered by the umbrella of their schema. Although they recognize that their schema for morality may not supply the foundation for all moral theories, it seems that their morality schema is broad enough to support a wide variety of moral agents operating under various moral theories and codes. Thus, the formulation of "moral agency" from the preceding paragraph should serve as a charitable threshold in evaluating intelligent machines for full moral status (region F in Figure 1).

This formulation, however, ensures that no foreseeable intelligent machine will attain the classification of full moral status. The "person" requirement is the culprit. Dennett [5], in referring to the concept "person," observes:

> One might well hope that such an important concept, applied and denied so confidently, would have clearly formulatable necessary and sufficient conditions for ascription, but if it does, we have not yet discovered them. In the end there may be none to discover. In the end we may come to realize that the concept person is incoherent and obsolete. (p. 267)

One might suggest replacing "person" with some less controversial term in the Gert and Gert [7] morality schema, but as Gunkel [8] explains, "…philosophers, medical ethicists, animal rights activists, and others have often sought to differentiate what constitutes a person from the human being in an effort to extend moral consideration to previously excluded others." Taking up Dennett's challenge, Gunkel reviews the criteria for personhood advanced by various authors and discovers that consciousness is on all these lists in some form. Further, it is typical to consider consciousness as a requirement for moral agency (pp. 46-48). The reason is "because consciousness is considered one of the decisive characteristics, dividing between a merely accidental occurrence and a purposeful act that is directed and understood by the individual agent who decides to do it." (p. 47) Unfortunately, it is also difficult to specify just what consciousness is.

At the moment, it is reasonable to conclude that no current or foreseeable intelligent machine meets the moral agency threshold. In spite of attempting to define this threshold charitably, there are still two ontological problems: consensus on how to (1) define "person" and (2) define perhaps its main criterion, "consciousness." Even if these two problems can be solved, there

[1] For example, with respect to the anthropocentrism/nonanthropocentrism distinction of Sterba [18], I am not claiming that Sterba would choose moral agency as the threshold for full moral status, nor am I claiming that he would subscribe to my hierarchical-ordering-of-thresholds assumption. Further, his claim that species and ecosystems have moral status may not be captured by region MS in my Figure 1. With respect to the moral agent/moral patient distinction discussed in Gunkel [8], note that region SF in my Figure 1 roughly corresponds to his discussion of the extension of moral status to animals. My choice of sentience as the threshold for an entity to be included in region SF is exactly the sentience due to Singer [17], which is used to justify animal rights. However, Singer, with an utilitarian insistence on equal considerability (although not necessarily equal treatment), would not endorse the idea that moral agents (region F in Figure 1) automatically have greater moral status than animals. Finally, Rachels [15] gives a non-traditional, but persuasive, account of moral status that my Figure 1 may not capture at all.

remains the epistemological problem of determining whether another entity truly possesses consciousness.

4.2 Sentience

In the context of moral status, "consciousness" typically means phenomenal consciousness. Qualia—"the introspectible qualities of our experiences"—give rise to 'what it is like to be' a particular thing: a human, a bat, a dog, etc.; if a thing has a "state with qualia", it has phenomenal consciousness [12]. Phenomenal consciousness sometimes is discussed along with other properties such as sentience [17], desires and preferences [11], intentionality [21], etc. Regarding the threshold for classifying an intelligent machine into region SF (Figure 1), I propose that the entity must have just enough phenomenal consciousness to experience pleasure and pain. This threshold is the "sentience" due to Singer [17]. My assumption is that a more richly phenomenally conscious entity, one with desires, preferences, intentionality, etc., would also experience pleasure and pain, but perhaps not vice versa.[2] If this assumption is correct, then this choice of threshold is charitable with respect to the question of deciding whether an entity possesses phenomenal consciousness—and charitable with respect to determining whether an intelligent machine belongs in region SF of the MSP. Despite this, it is unlikely that foreseeable intelligent machines will fall into region SF.

There is a well-known epistemic problem: determining whether an entity actually is phenomenally conscious, even in the limited sense of being able to feel pleasure and pain. This problem arises because we can only observe the behavior of another entity, without access to its private and subjective mental world (if any). For example, it could be that an entity merely simulates the feeling of pleasure or pain, using appropriate behavioral responses, without actually experiencing pleasure or pain. In discussing whether animals are sentient, DeGrazia [4] claims that evidence supports the ability of vertebrates to feel pain, but that this is still an open question for "all but the most 'advanced' invertebrates." He adds:

> …the evidence available today is too indeterminate to justify confidently drawing the line between sentient and non-sentient animals in any specific place, although it is virtually certain that some invertebrates, such as amoebas, are non-sentient. (p. 44)

If it is difficult to draw this line for animals, it is not clear why it would be less difficult to draw this line for machines. Thus, it likely will be very difficult to reach consensus on whether an intelligent machine actually is sentient.

4.3 A Good of Its Own

Some are willing to grant moral status to a non-conscious entity with "a good of its own." For example, they would allow that my oak tree has moral status. Its good manifests itself in its continued survival, growth, and reproduction. In self-maintenance, it drops large branches in my yard when faced with a season of drought or the insult of the city cutting some of its roots to modernize the city sewer system. Regarding its reproductive ends, it pelts my house in late summer with acorns. It does these things in the service of what it means to be an oak tree, presumably without any consciousness and without regard to the ends or good of anyone or anything else.

One could say that a conscious person and my oak tree both have *interests*. In Kaufman's [11] view, "morality is centrally, if not essentially, concerned with assessing benefits or harms resulting from the actions of moral agents." According to Kaufman, because benefits or harms only matter to entities with interests, it is just such entities that are candidates for moral status. Kaufman [11] claims that there are "two distinct senses of what it means to have an interest": (1) desires and (2) "a good or a well-being." Desires (actual, potential, or idealized) require what Kaufman refers to as "mentality." Assuming that if an entity has desires, then the entity is sentient, the entity would belong to region SF (assuming also that it is not a moral agent).

Mentality, on the other hand, is not required to have "a good or a well-being." In summarizing claims from environmental philosophy, Kaufman [11] states:

> Environmental philosophers have ascribed interests to such things as plants, other forms of life, whole species, and ecosystems. The good or well-being of these things is said to consist in their achieving their respective ends. Some authors talk about such natural entities as having a *telos*, and our respectful recognition of these ends is taken as a basic moral insight. (pp. 59-60)

Using my oak tree again as an example, it has a telos—an end or purpose—to survive. One good for it is water; another good for it is not to have its roots damaged. Neither a drought nor having its roots cut is good for my tree; neither is in my tree's interest in staying alive.

On such accounts, Kaufman [11] argues that machines, like living things, also can have a good or well-being, ends, and interests, but emphasizes how environmental philosophers are loathe to grant moral status to machines for the reason that, unlike living things, a machine's good and ends are not its own, but rather are derived from human ends. His rejoinder is that human goals may be "imposed from without" (e.g., by "a function of upbringing and enculturation"). He concludes that:

> The mere fact that one's ends are not one's own is not a good reason for denying that interests can nevertheless depend upon those ends. The same holds for machines: the claim that their ends are derived from our ends, that they have no ends of their own, is irrelevant in determining whether or not they are able to have interests. (p. 63)

An objection that avoids Kaufman's argument is that machines have no ends at all, rather than ends which happen to be derivative. Kaufman [11] concedes that talk of a machine's having ends may be just convenient shorthand, which, if necessary, could be replaced by more formal mechanistic descriptions of the machine's behaviors. However, he suggests that this may also be

[2] For example, Kaufman [11] argues that Singer's sentience is not sufficient for an entity to have preferences.

the case for naturally occurring organisms, that we may talk in terms of their ends merely because of our poor understanding of the mechanistic processes regulating their behavior.

Perhaps just for this reason, explanation in terms of an end or telos seems no longer sufficient. According to Basl and Sandler [1], explanation of the etiology of ends is also required:

> Appeals to teleological organization are a step toward explicating the basis and content of nonsentient organisms' good. However, it is also necessary to provide an explanation for the teleological organization; one that demonstrates that the teleological organization of plants isn't merely imagined. … The most prominent and most plausible explanation of the source of teleology in nonsentient biological organisms appeals to an etiological account of functions (Cahen 2002; Varner 1998).
>
> According to etiological accounts of function, a part or trait of an organism has the function of doing F only if it was selected for doing F. (p. 93)

They give the example of a heart, the function of which is to pump blood only if it was selected for pumping blood. Natural selection is operative for organisms, but Basl and Sandler argue that any "selection for" will work to ground the etiological account of function in justifying an entity's telos. This opens the door for an artifact—perhaps an intelligent machine—to have interests and a good of its own.

A robot experiment by Briggs and Scheutz [3] illustrates. Their research goal is to incorporate "felicity conditions" into a robot so that it may disobey unclear, ambiguous, and even deliberately deceptive instructions from a human. Felicity conditions are "contextual factors that inform whether an individual can and should do something." Briggs and Scheutz equipped a NAO robot with a set of felicity conditions for determining whether to comply with an order. One such condition is: "Does it violate any normative or ethical principle for me to do X, including the possibility I might be subjected to inadvertent or needless damage?" They placed the robot on a tabletop and gave it commands such as "sit." The robot complied. When close to the table edge and ordered to walk off the table, however, the robot disobeyed until it was coaxed off the table into the arms of the researcher. Even though the purpose of their experiment was to test robot disobedience and not to establish that the endowed NAO robot has a good of its own, it seems that the robot, though not conscious, does have a telos: to remain intact. This telos is justified by an etiological account of function. The function of the set of felicity conditions is to allow the robot to disobey an improper order, and this function was selected for by the researchers. Thus, the endowed NAO robot appears to have interests and a good of its own.

One may object that Basl and Sandler [1] are too loose in allowing anything other than natural selection to count with respect to a function being "selected for." Is there some distinction between artifact and organism that would allow only an organism to have a good of its own?

An obvious distinction between an artifact, as we have typically known them, and an organism is that the organism is alive. Basl and Sandler [1] make a persuasive argument that the "living/nonliving distinction" is not relevant by showing how internal organization, goal-directedness, and dynamism—three features they claim are important for living entities—are also found in some nonliving artifacts.

Even if their argument does not dissuade one from belief that there is a distinction between organism and artifact, products of synthetic biology, which are both artifact and organism, confuse this distinction. As a result, Sandler's [16] natural-artifactual continuum seems more appropriate than a hard distinction. If such products are organisms, then, it seems, they do have a good of their own [1]. Can these "artifactual organisms" be considered machines?

One could argue that a machine is just a type of artifact—therefore, whatever holds for artifacts, holds for machines as well. Boldt and Müller [2] express concern about a premature conflation of 'life' and 'machine' though, due to the traditional association of 'life' with 'value.' Their concern is that this "may in the (very) long run lead to a weakening of society's respect for higher forms of life that are usually regarded as worthy of protection." Thus, they caution synthetic biology researchers and commentators to be careful with 'machine' metaphors. Although their underlying concern seems legitimate, machine metaphors in synthetic biology soon may be obsolete. Synthetic biologists have already created cells that perform basic logic operations, count, add, and serve as crude memories; some anticipate that within the next five years biocomputers made from such raw living materials might be ready for use in diagnosing and treating certain diseases [13]. If so, then arguably these would not be machine metaphors, but rather genuine machines. Further, although early biocomputers will be primitive, slow, and inaccurate in comparison to their electronic counterparts, they are expected to be capable of interacting with the natural world in ways not possible for electronic computers [13].

In summary, it seems plausible that an intelligent machine could meet the threshold of having a good of its own. This would place current and foreseeable intelligent machines into region MS of the MSP (Figure 1).

5 Moral Obligations to Intelligent Machines?

If the reasoning so far is correct, current and foreseeable intelligent machines could have moral status. Could we, as moral agents, have obligations to such machines? Consider the sort of 'lifeboat dilemma' typical of philosophy: assume a lifeboat has room for just one more occupant, either an intelligent machine or another normal adult human (or perhaps, instead, a pet animal). Intuitively, many of us would want morality to dictate that the machine be tossed overboard in favor of the additional human (or family pet). If the intelligent machine belongs in region MS of the MSP, as argued, then morality and intuition coincide. The machine has less moral status than the human moral agent or sentient animal. Thus, to a first approximation, the upper bound of our moral obligations to an intelligent machine seems prudent. What about a lower bound to such obligations though? Torrance [19] warns, for example, about the potential human cost of erroneous ascription of moral status to "artificial beings."

Basl and Sandler [1] point out that even if an entity has moral status, this does not mean that moral agents necessarily must have "much if any concern." The information ethics (IE) of Floridi [6] goes beyond environmental ethics by including anything that exists as "worth some initial, perhaps minimal and overridable, form of moral respect." (See Figure 1, region NM as a possible location of moral status for such entities.) The point that these authors seem to be making is that some entities could have moral status, but that the actual obligations of a moral agent to such entities could be negligible. As argued here, current and foreseeable intelligent machines fall into region MS of the MSP. Even some non-intelligent machines [11] may qualify as having a good of their own, placing them into region MS as well. Which machines make obligations upon us as moral agents? I propose two criteria: the machine must be an intelligent machine as discussed here and the machine must possess the "functional morality," or functional moral agency, described by Wallach and Allen [20]. By analogy with actual moral agency, which sets a high bar for moral status, functional morality could be an appropriate lower bound for establishing a moral status high enough within region MS for us to take seriously any moral obligations to machines. After all, an intelligent machine with the functional morality of Wallach and Allen would be an active, autonomous (in some respect) agent operating in the moral world that we inhabit.

6 Conclusion

Current and foreseeable intelligent machines could have the approximate moral status of environmental entities, such as plants and trees (see region MS of Figure 1). I propose that the only machines, for which we may have meaningful moral obligations, are intelligent machines that embody the "functional morality" of Wallach and Allen [20]. The limit of obligations to such intelligent machines will fall short of our obligations to entities that are sentient.

ACKNOWLEDGMENTS

This research was funded by an Indiana University New Frontiers in the Arts and Humanities Experimentation Fellowship.

REFERENCES

[1] Basl, J., & Sandler, R. 2013. Three puzzles regarding the moral status of synthetic organisms. In G. E. Kaebnick & T. H. Murray (Eds.), *Synthetic biology and morality: Artificial life and the bounds of nature* (pp. 89-106). Cambridge, MA: MIT Press.

[2] Boldt, J., & Müller, O. 2008. Newtons of the leaves of grass. *Nature Biotechnology* 26(4): 387-389.

[3] Briggs, G., & Scheutz, M. 2017. The case for robot disobedience. *Scientific American* 316(1): 44-47.

[4] DeGrazia, D. 2002. *Animal rights: A very short introduction*. New York: Oxford University Press.

[5] Dennett, D. C. 1978. Brainstorms: Philosophical essays on mind and psychology. Montgomery, VT: Bradford Books.

[6] Floridi, L. 2008. Information ethics: Its nature and scope. In J. Van Den Hoven & J. Weckert (Eds.), *Information technology and moral philosophy* (Ch. 3). New York: Cambridge University Press.

[7] Gert, B., & Gert, J. 2017. The definition of morality. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Fall 2017 Edition)*. URL = https://plato.stanford.edu/archives/fall2017/entries/morality-definition/.

[8] Gunkel, D. 2012. The machine question: Critical perspectives on AI, robots, and ethics. Cambridge, MA: MIT Press.

[9] Jaworska, A., & Tannenbaum, J. 2017. The grounds of moral status. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Fall 2017 Edition).* URL = <https://plato.stanford.edu/archives/fall2017/entries/grounds-moral-status/>.

[10] Kant, I. 1785/1998. *Groundwork of the metaphysics of morals*. (M. Gregor, Trans. and Ed.). Cambridge, UK: Cambridge University Press.

[11] Kaufman, F. 1994. Machines, sentience, and the scope of morality. *Environmental Ethics* 16(1): 57-70.

[12] Kolak, D., Hirstein, W., Mandik, P., & Waskan, J. 2006. *Cognitive science: An introduction to mind and brain*. New York: Routledge.

[13] Lu, T. K., & Purcell, O. 2016. Machine life. *Scientific American* 314(4): 58-63.

[14] Moor, J. 2006. The nature, importance, and difficulty of machine ethics. *IEEE Intelligent Systems* (July/August, 2006): 18-21.

[15] Rachels, J. 2004. Drawing lines. In C. R. Sunstein & M. C. Nussbaum (Eds.), *Animal rights: Current debates and new directions* (pp. 162-174). New York: Oxford University Press.

[16] Sandler, R. 2012. Is artefactualness a value-relevant property of living things? *Synthese* 185: 89-102.

[17] Singer, P. 1974. All animals are equal. *Philosophic Exchange* 5(1): 103-116.

[18] Sterba, J. P. 2005. *The triumph of practice over theory in ethics*. New York: Oxford University Press.

[19] Torrance, S. 2008. Ethics and consciousness in artificial agents. *AI & Society* 22: 495-521.

[20] Wallach, W., & Allen, C. 2009. *Moral machines: Teaching robots right from wrong*. New York: Oxford University Press.

[21] Zahavi, D. 2003. *Husserl's phenomenology*. Stanford, CA: Stanford University Press.