**Consider the Xenobot: Moral Considerability for Intelligent Machines Revisited**

Joshua C. Gellers

Magdalena Hoły-Łuczaj

Abstract

Can intelligent machines be morally considerable? Should they be? For environmental philosophers, the answer to both questions is plainly in the negative. That is, because intelligent machines are not alive and not putatively ‘natural’ they do not warrant moral concern. But these conclusions are not shared by philosophers of technology, who have found artifacts in general and certain technological entities in particular to be eligible for moral status. In this essay, we seek to overcome this discrepancy by arguing that intelligent machines can and should be morally considerable. We show how environmental ethics has neglected important insights from philosophy of technology, inadequately assessed the status of technology in the Anthropocene, and failed to appreciate the normative obligation to adopt a pluriversal view of ethics. We illustrate the possibility and necessity of expanding the moral universe to include intelligent machines by presenting a ‘most likely’ case for moral considerability under both environmental ethics and philosophy of technology—the xenobot. By exposing the shortcomings of environmental ethics and the promise of philosophy of technology, we aim to address the intellectual incoherence that has paralyzed current theorizing about the moral status of nonhumans.

Key words: moral considerability, artificial intelligence, environmental philosophy, Anthropocene

**Introduction**

Might some forms of intelligent machines be morally considerable? To most students of environmental ethics, the answer is plainly ‘no.’ Technological artifacts are not alive, do not possess interests, and amount to “mere things” (Feinberg, 2013), not entities worthy of moral concern. While environmental ethicists continue to debate the moral status of individual biotic organisms and ecosystems (and their constituent elements), these observers remain united in their rejection of the idea that non-living, non-natural artifacts deserve a place in the moral universe. *Our* moral universe.

But this near-unanimous response on behalf of environmental ethicists to the so-called “machine question” (Gunkel, 2012)—that is, whether and to what extent machines qualify as moral agents and/or moral patients—rests on a faulty foundation. For one, following Gunkel (2018), this inquiry actually consists of two related questions—*could* machines be morally considerable and *should* machines be morally considerable?—which have not been sufficiently parsed in relevant scholarship. In addition, experts have failed to adequately engage with intellectual and ethical developments that cast doubt on the widely accepted notion that only natural beings or systems possess moral value. Once these factors are taken into consideration, we argue, the case for the moral considerability of intelligent machines becomes not just plausible but compelling.

The paper proceeds as follows. First, we review the literature on moral considerability, placing special emphasis on recent works that focus on artificial intelligence (AI). Second, we discuss philosophical, empirical, and normative interventions that support extending moral concern to intelligent machines. Third, we present a real-world example of technology that satisfies many of the criteria for moral considerability put forth by environmental ethicists, calling into question theories of moral status that ignore artifacts. Fourth, we explore the practical implications and potential criticisms of our argument. Finally, we conclude by offering areas for future research.

**A Brief Review of Moral Considerability**

To begin, it is important to explain what we mean by ‘moral considerability.’ Following Goodpaster (1978), we interpret the term as referring to the extent to which something is deemed worthy of practical respect. Another instructive way to clarify the core concept of moral considerability is to distinguish between moral patiency and moral agency. Moral agency entails the active capability to take responsibility for one's own actions, directed toward entities to which we owe ethical obligations and considerations. On the other hand, moral patiency denotes the passive state of deserving moral consideration. Importantly, one can be a moral patient without being a moral agent (i.e., an infant or mentally incapacitated individual), but all moral agents are also moral patients.

Hale (2011) later elaborated that moral considerability might be profitably thought of as pertaining to one of three questions whose combined answers determine moral *status*: (1) which entities fall within the moral scope (i.e., moral considerability), (2) what must we consider about such entities (i.e., moral relevance), and (3) how much weight should we assign these considerations (i.e., moral significance)? Distinguishing between these facets will prove to be crucial in examining the specific position of intelligent machines like xenobots. The distinction between grounds for admitting this biomachine into the circle of moral consideration and the aspects of this peculiar entity we need to morally consider when dealing with it will be unpacked in the final part of the essay. This endeavor will provide critical insights into the discussion on formulating the criteria of moral considerability in environmental ethics and philosophy of technology.

Over the course of several decades, environmental ethicists have offered a range of criteria intended to justify the moral consideration owed a natural entity (i.e., the individualistic perspective) or system (i.e., the holistic approach), most of which privilege biotic, living beings. Proposed qualifications have included, *inter alia*, “being alive” (Goodpaster, 1978, p. 310), “being in existence” (Hunt, 1980, p. 61), “[g]enuine goal-directedness” (Cahen, 1988, p. 215), and possessing “life processes [that] maintain the viability of the whole” (Johnson, 1992, p. 157). Many of these efforts explicitly touch upon the subject of technology, finding it not eligible for moral status. For instance, Hale (2009) writes that technological artifacts are always already morally considered when they exist in their natural form (which is independent of human will), so to deem such artifacts morally considerable would be to “double-count” their moral status. Such an account resonates with the strong tendency to portray artifacts as villains. This consists in arguing that their production requires exploiting natural resources, their use pollutes the environment, and when they are no longer needed, they litter the planet. Yet, underappreciated non-Western and Indigenous cultural worldviews and recent technological advances have arguably closed the artifactual world’s moral gap with the biotic world. We elaborate on this point later.

Despite these varied approaches, a great number of participants in this discourse seem to agree that for something to be considered on a moral basis, it needs to have interests. There are at least three ways in which interests manifest: (1) as one’s ability to be made better or worse off, (2) as the realization or frustration of one’s notion of the good, and (3) as the extent to which one’s existence or identity is fostered or hindered (Hoły-Łuczaj, 2019, p. 71). Some argue that the interests of the entity itself matter (Cahen, 1988; Taylor, 1981), whereas others urge that moral concern for nonhumans is ultimately epiphenomenal to the interests of humans (Kant, 2013). Still others identify interests of entire species or ecosystems (Johnson, 1992).

Although the precise kind(s) and hierarchy of interests that qualify an entity (or its surroundings) for moral consideration remain a source of contention, environmental ethicists generally maintain that artifacts do not *directly* possess interests of their own. This directness is one of the essential aspects of the concept of moral considerability (Birch, 1993; Gorke, 2003). Only direct attention, being focused on the particular type of things, guarantees sufficient respect for them, eliminating the threat that their well-being will be overlooked or ignored for sake of some other, more human-like being or nature as a whole (see Birch 1993).

More recently, the question of (direct) moral considerability has been applied to AI in general and social robots in particular by philosophers of technology. Here the debate has proceeded more in line with the trajectory of animal ethics, which has been largely dominated by a “properties” approach to moral status involving recognition of an entity’s possession of a trait like consciousness or sentience that stands in contrast to a “relational” approach in which moral engagements are governed by unique interactional contexts (Coeckelbergh & Gunkel, 2014). Given its emphasis on the importance of ontological attributes, scholarship on the moral considerability of AI has almost exclusively centered its analysis on individual entities.

Commentators in this space have proposed a range of qualifying factors, including consciousness and self-awareness (Neely, 2013), “being-in-the-technological-world” (Tavani, 2018), the extent to which AI cultivates an emotional response in a human (Elder, 2020), and sentience (DeGrazia, 2022). Most of these criteria are at least weakly anthropocentric given that they valorize human traits or the effects AI has on humans, yet they also exhibit the residue of environmental ethics’ bias towards living things. To be fair, the burgeoning empirical literature on robot rights suggests that popular support for the moral consideration of AI is positively associated with life-like qualities present at the individual level, such as biological substrate and possession of a soul (Pauketat & Anthis, 2022), so the theory resonates with contemporary opinion.

Despite the heavy influence that animal ethics has had on the discussion surrounding technological entities, a few interlocutors have presented arguments that deviate from a strictly properties-based account of moral consideration. Approaching the moral considerability of AI from an ecocentric perspective, Laukyte (2019) builds an uncommon bridge between environmental ethics and philosophy of technology. Extending Deep Ecology to the domain of technology, she submits that both living and putatively non-living entities could conceivably belong in the moral circle to the extent that they constitute part of our social environment. This conclusion follows from the premises that (1) Deep Ecology already finds non-living entities worthy of moral consideration, (2) artificial intelligences are similarly non-living forms of life, (3) the entire environment (including its constituent elements) is inherently valuable, (4) technology forms part of our environment, and (5) there is no ontologically meaningful separation between nature (and its living beings) and the artificial environment constructed by humans.[[1]](#footnote-1)

Hoły-Łuczaj (2019), although writing about artifacts in general, wades into the debate over the moral status of technology. She foregrounds her argument with two claims. First, there is a pro-environmental basis for finding artifacts morally considerable; doing so can increase the respect and care that humans show non-human others, which can have positive spillover effects on the ecosystem. Second, environmental philosophers have failed to adequately appreciate the moral dimensions of artifacts. For one, they have mistakenly treated artifacts as a monolith when there are important moral distinctions to be made with respect to their ontology. In addition, they have overlooked the contributions of philosophers of technology, who have taken a less rigid and more charitable view of the metaphysical nature of artifacts. Seeking to move beyond the shortcomings of environmental philosophy, Hoły-Łuczaj asserts that artifacts are morally considerable because of their “affectability,” or their ability to be influenced by humans. Far from being mere things, such entities possess an interest in maintaining their existence and identity grounded in their physical individuality. This means that only physical artifacts, not digital ones, are worthy of moral consideration. Because digital artifacts cannot have their physical integrity compromised, they do not possess morally relevant interests and are thus not morally considerable.

Coeckelbergh (2021) advocates in favor of extending indirect moral standing to social robots based on the way humans relate to them. More specifically, he articulates several circumstances under which such intelligent machines might warrant moral concern—when the virtue of a human is implicated by how it treats a robot, when a human develops feelings for a robot, when a robot is considered a partner in leisure activity or a collaborative enterprise, or when humans are uncertain about a robot’s moral standing (thus requiring that we err on the side of caution to avoid making a moral mistake). In these instances, ontological properties are, at best, indirectly related to the moral calculus; it is the nature of the relations between humans and robots that dictates the prospects for moral consideration.

The above survey of the literature on moral considerability allows us to reach several preliminary conclusions. First, environmental philosophy almost universally excludes non-natural entities and systems from the moral circle on the basis that only living things possess interests that humans should worry about protecting. Second, philosophy of technology exhibits a willingness to extend moral consideration to forms of AI using arguments native to the field and from animal/environmental ethics. Third, environmental ethics largely ignores non-Western and Indigenous perspectives when developing a framework for moral considerability. Fourth, philosophy of technology tends to couch its moral logics in anthropocentric language. Fifth, philosophy of technology (specifically AI ethics and roboethics) focuses mainly on moral concern for individual entities, not whole systems. Sixth, environmental philosophy mostly seeks to promote direct moral consideration for natural bodies whereas philosophy of technology trades in both direct and indirect rationales. Taken together, we can sort positions on moral considerability from environmental philosophy and philosophy of technology into a four-part classification scheme consisting of Western/non-Western, anthropocentric/non-anthropocentric, individual/system, and direct/indirect.

At this point, without critically examining the strength of the above arguments, we might be able to derive some criteria that should satisfy even the most stringent conditions for the moral considerability of AI according to environmental philosophy (which, as we have established, is more restrictive than philosophy of technology). To warrant moral consideration, the ideal technological candidate would be (1) an individual entity (2) that is alive and (3) pursues non-arbitrary biological goals (i.e., self-preservation, replication, etc.). Such a standard would immediately appear impossible for any artifact to obtain, which is precisely the point. This is a demanding set of qualifications that, if achieved, would compel environmental philosophers to admit such a being into the moral circle. Let us refer to this as the ‘*could* criteria,’ for if AI can meet the rigors of this moral recipe, it will demonstrate that intelligent machines *could* be morally considerable in a way that just about any environmental philosopher would have to accept.

Demonstrating that robots *should* be granted moral consideration, however, will require some heavy lifting. In particular, we will need to show that the approach undertaken by environmental philosophy is excessively stringent and that there are valid arguments that seriously undermine the legitimacy of this rigidity. The following section works towards this end by exploring three interventions that cast doubt on some of the assumptions used to justify the exclusion of technology from the moral universe.

**Three Interventions**

*Philosophical: Philosophy of Technology*

Environmental ethics and philosophy of technology, despite sharing room for common ground, have rarely engaged with one another (Epting, 2010). This lack of dialogue is unfortunate for both fields, as they stand to gain much from each other. We see this mutual abnegation figure prominently in the discussion about moral considerability. Environmental ethicists, be they biocentrists or ecocentrists, tend to possess a biotic, life-centered bias. As described earlier, anything not considered a teleological center of life is therefore a ‘mere thing.’ Philosophers of technology, on the other hand, adopt a more expansive view not of what constitutes biotic material or what is viewed as a living being, but of the kinds of entities to whom we hold ethical obligations. Crucially for these observers, moral considerability is determined more by the contours of our ethical systems than by an empirical accounting of the properties possessed by a given entity.

This difference between the two fields and the conclusions they reach about moral considerability can be seen in how they interpret the distribution of moral agency and patiency among beings. For environmental ethicists, humans are clearly moral agents and moral patients, but questions abound as to whether animals, natural bodies, and/or ecosystems are also moral patients. Consequently, in environmental ethics, artifacts, be they simple tools, machines, or intelligent robots, are neither moral patients nor moral agents. Instead, for philosophers of technology and feminist scholars of science and technology studies, agency is distributed across all actants (Verbeek, 2009) and can be a capability of all material substances down to the molecular level (Bennett, 2010). The human-centered understanding of “willful agency” (Cielemęcka & Daigle, 2019) is replaced with the more ontologically inclusive notion of “agentic capacity” (Coole, 2005). Furthermore, some philosophers of technology argue that robots deserve the status of moral patients, regardless of whether we find them to be moral agents or not (Gellers, 2020; Gunkel, 2018, 2022).

The point here is that while environmental ethics has largely ignored how concepts central to theorizing about moral considerability have been addressed by other subfields, philosophy of technology has directly challenged the taken-for-granted assumptions of environmental ethics. In order to adequately defend against such charges, environmental ethics would need to put forth a compelling argument as to why a system of ethics should not provide ontological cover for technological artifacts without resorting to begging the question.

*Empirical: The Anthropocene*

The concept of the Anthropocene, the geological era marked by human activity that has rapidly dominated natural processes on a planetary scale (Zalasiewicz et al., 2011), serves as a point of reflection that environmental ethicists heed superficially and philosophers of technology engage with seriously. One need only look to experts writing on earth system governance to determine that environmental ethics has failed to adopt some of the more paradigm-altering insights offered by the Anthropocene. Inattention to these empirically grounded claims, either by accidental omission or deliberate avoidance, has allowed environmental ethicists to evade the issue of incorporating technology into the moral frame.

The arrival of the Anthropocene has yielded a number of observations relevant to the present discussion. First, it signals the end of the Cartesian nature/culture divide, which alleges that ‘nature’ comprises all that is naturally occurring (except humans) and ‘culture’ separately refers to man-made ideas, practices, and artifacts like technology. While anthropologists have long strived to eliminate thinking in such dichotomous terms (Hui, 2017), the Anthropocene hastened pursuit of this project by permitting “an opening up of hitherto prohibitive epistemic ‘closures’” (Kotzé & Kim, 2019, p. 3). This turn has made possible a shift in orientation from discrete categories to integrative complex systems. Now, Earth has been technified as much as technology has been earthed, rendering their entanglement an ineradicable fixture of modernity (Lemmens et al., 2017). Practically speaking, this means it is no longer tenable to consider nature and technology fundamentally separate spheres of life and non-living substances, respectively. Whereas ecocentrists maintain that humans do not exist outside of nature but technology lies beyond the bounds of the ecological whole, philosophers of technology go further by including the synthetic realm in the environment as well.

Second, a related point involves the pervasiveness of technology during the Anthropocene and how it figures into the epoch conceptually. Some have argued that technical systems rise to the level of a major component of the Earth system, complementing the physical and biological worlds with a mental one called the “technosphere” (Haff, 2014), which consists of “all those parts made and/or modified by humans” (Kotzé, 2020, p. 80). Others have held that the causal significance of technology is so great that the Anthropocene is more accurately labeled the “Technocene,” an era in which technology is synonymous with “nature itself” (Cera, 2017, pp. 244, 247). Whether technology is interpreted as a layer of the Earth system or the defining factor of the current geological period or both, it stands to reason that its omnipresence must be taken into account and not simply discarded on the grounds that it is not living. Few environmental ethicists might take issue with including dead plants or future generations of humans in the moral universe, and yet technology, with its extensive reach and integration with natural systems, remains curiously outside the purview of moral considerability.

Third, the Anthropocene has witnessed the emergence of entities that belie facile either/or categorization for moral or other purposes. In line with the blurring of ontological boundaries and tentacular spread of technology in putatively ‘natural’ domains, the existence of hybrids casts fresh doubt on our ability to maintain an ethical divide between the synthetic and the biotic. The term “hybrid” owes an intellectual debt to Bruno Latour, who used it to describe the indissoluble entanglement of “both human and technological actants” (Conty, 2017, p. 308). It can also entail “natural-technological hybrids” such as “eco-mimetic bio-robotics” of the type discussed later (Blok, 2017, p. 138). The ontological indeterminacy exemplified by hybrids complicates the kind of ethical line drawing that environmental ethicists use to justify the exclusion of the artefactual from both biocentric and ecocentric moral universes.

*Normative: Post-anthropocentrism and Ontological Pluralism*

In addition to philosophical and empirical justifications for moral considerability, there is an equally strong normative case. The two main arguments offered to this end include the need to combat anthropocentrism and the importance of acknowledging and respecting Indigenous and non-Western perspectives. We address elaborate on both of these arguments in turn.

First, scholars writing about the Anthropocene have argued that the planet and all its inhabitants are in peril due to the human-centered nature of decision-making, systems, institutions, and values (Kotzé, 2020). By placing humanity outside of and above the environment in which we are always already enmeshed, we have fostered conditions ripe for interpersonal violence—as who is considered ‘human’ has varied over time—and sowed the seeds of ecological destruction—by placing our own immediate consumptive desires above the needs of the Earth and future generations (Celermajer et al., 2023). In order to prevent further devastation and chart a more ethical path forward we must address this “normative anthropocentrism,” which not only asserts the primacy of humans above all others but also deliberately structures society and its operations around this assumption (Mylius, 2018, p. 159).

Observers have located within our predilection for human supremacy evidence of right-wing authoritarianism (Fortuna et al., 2023), speciesism (Kim, 2023), and white supremacy (Fitz-Henry, 2022), ideological stressors that frustrate the achievement of justice and the cultivation of morality. Throughout history these tendencies have fueled the exploitation, subjugation, and maltreatment of *all* those not deemed ‘human’—people of color, animals, and (even) robots. Crucially then, eliminating the grip that anthropocentrism holds on our current ethical orientation will necessarily invite moral consideration of a wider range of entities and systems, including forms of technology. Posthumanists and feminist materialists have made this very argument (Forlano, 2017; Ivic, 2022), which has yet to be adequately addressed in environmental ethics.

Second, whether by innocent omission or purposive disregard, much of the writing about the moral status of nonhumans neglects non-Western and Indigenous perspectives that challenge hegemonic ontologies. A charitable take on this curious shortcoming might be that such authors have limited academic training in and exposure to alternative worldviews. A less forgiving assessment counsels that non-Western and Indigenous cosmologies threaten the integrity and coherence of Western ontology, leading to their deliberate exclusion. Either way, the current situation places Western thinkers and those adopting exclusively Western concepts in a bit of a bind. That is, either they admit to ignoring non-Western and Indigenous perspectives for ethnocentric reasons (which is ethically indefensible) or they agree to learn about and incorporate such views into their own scholarship at the risk of undermining their own arguments and conclusions (which could hurt their careers and complicate their research agendas).[[2]](#footnote-2)

Taking non-Western and Indigenous perspectives into account when evaluating the moral status of technology leads to a diversity of views, some of which run directly counter to the Western conceit found among many environmental ethicists that non-living matter doesn’t matter morally. A few examples prove illustrative. In Confucian philosophy, specifically Confucian virtue ethics and role ethics, robots might qualify for moral considerability because of how our treatment of them affects our capacity to develop desirable traits like empathy or due to our interest in the respecting the roles robots play in our social lives (Elder, 2020). Shintoism holds that all entities natural or artefactual possess their own spirit (Kitano, 2006), which “come[s] alive” when brought into productive use by humans (Mitsukuni et al., 1985, p. 90). Thus, even otherwise inert “things…have the same ontological status as living entities” (Vallverdú, 2011, p. 178). For one Plains Cree scholar, “all things have a place in our circle of kinship” (Lewis et al., 2018, p. 7), including AI and machines. However, reactions to this interpretation by members of other Native tribes ranges from qualified support to vehement opposition. Finally, in terms of the African philosophical perspective of *ubuntu*, experts are split. On the one hand, a robot could enjoy moral standing if it engages in actions that would qualify it for personhood, including relating well to others and fulfilling social roles (Jecker et al., 2022). On the other hand, we should not afford moral consideration to robots if doing so would impede human relations and thus our ability to become fully actualized humans (Friedman, 2023).

*The End of Exclusion*

The three interventions discussed in this section—philosophy of technology, the Anthropocene, and post-anthropocentrism and ontological pluralism—pose a serious challenge to the life-centered bias of environmental ethics. That is, environmental ethicists would have to disprove or otherwise reject on the basis of logic, empirics, or ethics all three arguments in order to successfully defend their position that technology is not morally considerable. Specifically, this would require demonstrating that (1) philosophy of technology is wrong about the moral status of technology; (2) the insights derived from the Anthropocene are incorrect (i.e., the nature/culture divide remains intact, technology is not an ordering principle of modernity, and hybrids do not present a difficulty for ethical assessment); and (3) there are no ethical imperatives mandating that we de-center humans in the moral universe or consider as potentially valid non-Western and Indigenous ontologies. While we await responses to this call, we tentatively conclude that the three interventions described above offer sufficient support, at least in the abstract, for the moral considerability of technology. More directly, some technological entities such as robots *should* be morally considerable. In the next section, we revisit and defend the idea that intelligent machines *could* be morally considerable using a concrete example.

**Enter the Xenobot**

Earlier we derived a series of conditions that, if met, would almost certainly have to be accepted by environmental ethicists as proving that at least some forms of technology *could* be morally considerable. To reiterate, under an individualistic perspective, a technological being worthy of elevated moral status would have to be, at a minimum, (1) an individual entity; (2) that is alive and hence pursues non-arbitrary biological goals (i.e., self-preservation, replication, etc.); and (3) is natural and as such entered existence without human intervention. Alternatively, under a holistic approach, the technology would have to play an integral role in or even constitute the ecosystem. Of course, the second and third criteria should serve to immediately disqualify from moral candidacy anything that is artefactual. However, what if there existed a kind of biological entity that was designed intentionally using technology, but which autonomously pursued the fulfilment of biological imperatives? Wouldn’t such a creature satisfy the moral requirements of even the stingiest bio-supremacist?

A team of researchers working at the intersection of biology and computer science has developed an organism that arguably fits the bill. The process began with a blueprint. Utilizing an evolutionary algorithm, a computer iterates and optimizes the design of tiny machines based on the extent to which they successfully perform certain tasks. The best designs are then manufactured using the stem cells of living tissue from *Xenopus laevis* (African clawed frog) embryos. The scientists describe the output of this bioengineering procedure as a “living, 3D approximation of the evolved design” (Kriegman et al., 2020, p. 1854).

Enter the xenobot, which, as Ball (2020) notes, is a term that appropriately describes the entity in question while honoring its etymological roots. The contemporary word “robot” was popularized by Karel Čapek’s 1921 play, *R.U.R.: Rossum’s Universal Robots* (Gunkel, 2023, p. 3). However, the term was actually coined by his brother, Josef Čapek, who adapted the Czech word *robota*, meaning ‘forced labor’ (corvee). Čapek originally wanted to call the robots ‘laborers,’ but in the end this name sounded too bookish to him and he borrowed the name from his brother to describe synthetic humans. Far from the machines made of metal typified by later science fiction, industry, and popular imagination, these initial robots were the artificial products of *chemical* processes. Frustrated by the manner in which “robot” had been coopted and contorted, Čapek asserted in a 1935 column that his *robot* was in fact a kind of “organic substance…something like another alternative to life” (Ackerman, 2024). Given their creation and material substrate, xenobots more closely adhere to the original definition of robot than do their anthropomorphic mechanical counterparts.

Subsequent versions of the xenobot have progressively illustrated its capacity for life-like capabilities. In 2021, researchers offered evidence that “the xenobots can navigate aqueous environments in diverse ways, heal after damage, and show emergent group behaviors” (Blackiston et al., 2021). Even more provocatively, a later iteration of the xenobot exhibited the ability to self-replicate when clusters of cells broke off from an organism early in its development (Kriegman et al., 2021). These cumulative innovations strongly suggest that this miniscule biomachine possesses attributes synonymous with living things. Significantly, humans and animals temporarily or permanently unable to engage in some of these same behaviors have still been deemed morally considerable. Surely, then, an entity made of living cells that can move autonomously, repair itself, work collaboratively, and reproduce evinces the kind of interests that even the most doctrinaire environmental ethicist would find worth protecting?

The philosophical implications of this development have not been lost on those responsible for its arrival. One of the xenobot researchers, a computer scientist and evolutionary roboticist named Josh Bongard, specifically notes how his team’s biological machines upends the dichotomous thinking that has long dominated Western science. He offers three reasons why the xenobot threatens to erode faith in Cartesian dualism—(1) the soft robots emerge from an feedback loop involving genes, form and function, and interactions with the environment; (2) their bodies belie binary categorization since they consist of more or less cells; and (3) selection for replication is informed by their shape (i.e., there is no formal instruction guiding evolutionary processes like there is in Turing’s tape). Bongard concludes with a call to “leave…dichotomous thinking behind, and instead learn to swim in the deeps, where real animals reside, and where really intelligent machines will reside” (Blackiston et al., 2023, p. 679).

This little creature, this human-conceived, AI-designed, animal cell-borne, environment-influenced biomachine, represents the kind of natural-technological hybrid that both philosophers of technology and environmental ethicists might agree warrants moral consideration. As to the former, xenobots come into existence by way of technological processes and thus constitute part of the technosphere. Given their status as artifacts, they can be affected by humans and hold an interest in maintaining their existence. As to the latter, xenobots are individual units, but they also work in swarms and can perform ecosystem services. Although they would not likely qualify as sentient, they are built from the living cells of an amphibian and exhibit traits associated with living beings. They also embark on non-arbitrary biological goals such as moving on their own volition, healing themselves, participating in group efforts, and self-replicating. Given that the xenobot satisfies many of the conditions necessary for moral considerability, whether scrutinized by the most imaginative philosopher of technology or the least forgiving environmental ethicist, what are some ethical implications of this real-world development and what kinds of critiques might our argument invite? We seek to address these questions in the following section.

**Implications and Critiques**

In advocating for the moral considerability of xenobots, the least controversial approach might entail following a holistic (ecocentric) perspective. Drawing on the words of Leopold (1949, p. 242), the pioneer of environmental ethics who stated that “a thing is right when it tends to preserve the integrity, stability, and beauty of the biotic community. It is wrong when it tends otherwise,” we might assert that xenobots *could* deserve moral considerability based on the extent to which they support the functioning of the ecosystem. Blackiston et al. (2023) explicitly sketch a vision of the future in which “xenobots clean up the environment” (p. 678). Were these xenobots to provide such an ecosystem service, they would then fall within the scope of moral considerability. Hoły-Łuczaj and Blok (2021) have established a precedent for this line of reasoning, finding that artificial coral reefs, as the emblematic case for hybrids and their functional (relational) value, could be morally considerable.

In advocating for the moral considerability of xenobots, philosophy of technology would likely extend its scope beyond their potential cooperation with the environment to include other scenarios where xenobots deliver medicine or combat cancer, as also envisioned by Blackiston et al. (2023). These kinds of engagements (Coeckelbergh & Gunkel, 2014) could serve as grounds for justifying the moral status of xenobots.

However, this line of argumentation is probably the clearest example of a gap between where an entity *could* be morally considerable and whether it *should* be under present conditions. The use of the subjunctive mood to discuss imagined or desired futures is not coincidental. As Blackiston et al. (2023) soberly underscore, it’s important to acknowledge that xenobots are not yet capable of performing the aforementioned activities; currently, they cannot solve any problems at all and may never become a viable technology. Therefore, it appears that, at least initially, xenobots *should not* be granted moral considerability based on the previously identified interactional contexts (Coeckelbergh & Gunkel, 2014). However, such consideration *could* be warranted in the future when xenobots are indeed capable of the described activities.

While their prospects for elevated moral status may seem remote at first, there remains another pathway to granting moral considerability to xenobots. We can again borrow from the environmental ethicist’s toolbox and employ the criterion of “being in existence” (Hunt, 1980) or its upgraded version expounded by philosophers of technology—challenging the belief in an ontologically meaningful separation between natural and artificial beings constituting the physical environment (Hoły-Łuczaj, 2019; Laukyte, 2019). Xenobots are concrete individuals existing in the technosphere whose integrity can be compromised or destroyed, thus warranting their moral consideration.

However, as emphasized earlier, the peculiarity of xenobots prompts us to explore the significance of distinguishing between the criterion for granting moral considerability to an entity and the aspects we need to consider about it, as proposed by Hale (2011). Specifically, the criterion of “being in existence” is typically applied to inanimate beings, yet xenobots exhibit traits traditionally associated with animate beings. Does this mean that the xenobot’s capabilities for self-repair, self-replication, and movement should be disregarded in ethical assessments? No, quite the opposite—these capabilities are of critical importance in the moral considerations of our actions toward xenobots.

To unpack this claim, we need to underscore that it is one thing to admit some being into the moral circle (or “club”, as Birch (1993) refers to its scope, criticizing its exclusivity) and another to consider this being within it. This is the difference indicated by Hale (2011) between what he refers to as (1) “moral considerability” and (2) “moral relevance.” Our understanding of this claim is that once some being succeeds in getting its foot in the door of the moral considerability club it deserves full consideration, without any *a priori* assumptions. The consideration given should not be based solely on the attainment of moral considerability, but rather on the specific capabilities and attributes of the entity in question.

In our case, this would mean that despite the fact that xenobots have entered the realm of moral considerability through the door typically reserved for inanimate beings, we need to consider all of their traits, including those typically associated with animate beings. That is to say, our moral consideration of xenobots is not limited to interfering with their physical integrity but extends to capacities such as self-repair, self-replication, and movement.

Such a direction of inquiry eliminates the requirement to classify a being as alive in order to consider the aforementioned capacities. It does not approach an entity by arguing upfront that it is alive, nor does it take a bottom-up approach. It might seem counterintuitive to start with the concept of aliveness when considering xenobots. Equally likely is the case that xenobots would fail the Duck test for being alive despite exhibiting vital activities (i.e., moving, healing, replicating). However, this does not mean that we should not morally consider such actions performed by xenobots. These vital actions seem to be of critical significance for their specific *identity*, and are therefore of greater moral importance than merely picking a single ontological property like consciousness or sentience and using it as the sole determinant of moral status.

This brings us to the following question: under what circumstances is it just to intervene and stop a xenobot from performing these activities, or even to annihilate it? Addressing this question seems indispensable when working within the boundaries of environmental ethics, which prioritizes non-interference with the flourishing (vigorous fulfilling of capacities) of non-human beings in the congenial environment (Drengson, 1992). Environmental ethics advocates in favor of refraining from disturbing non-human beings without supplying a well-justified reason for doing so.

What could be such a reason when it comes to xenobots? The very first thing that comes to mind is a situation in which a xenobot threatens or harms humans or natural beings. Of course, this assumes that the interests of human beings and natural beings have priority over xenobots. Exploring this problem would involve considering Hale’s (2011) third aspect of moral status, namely moral significance. As engaging with it would definitely go beyond the scope of the paper, for the purpose of our analysis, let us assume that xenobots are indifferent to the interests of humans or natural beings. What then? Can we freely disallow the xenobot from moving, self-replicating, etc.? This doesn’t seem right in light of granting the xenobot moral considerability. Thus our analysis inclines us to claim that *it is morally wrong to prevent xenobots* *from exercising their capacities without justification.*

**Conclusion**

The case of the xenobot is uniquely complex compared to other hybrids, such as the biomimetic or bioaugmented projects discussed by Hoły-Łuczaj and Blok (2019). This complexity arises from the extensive blending of features traditionally categorized within either the realm of nature or technology. Its classification as a biomachine already appears somewhat oxymoronic. It is therefore not surprising that we might find ourselves confused about its moral status.

In our paper, we developed an argument in favor of recognizing the moral considerability of intelligent machines. We offered several reasons why such entities *could* or *should* be found morally considerable. Using the example of the xenobot, a natural-technological hybrid, we explored how this argument might be applied to an existing entity. Drawing from both environmental ethics and philosophy of technology, we identified several pathways used to determine the xenobot’s moral status and found a range of possible responses. Given the limitations of pathways reliant on anticipated functions of xenobots, we found the strongest support for approaches that view the biomachine as a “being in existence” that is nevertheless animate and as an entity located within the larger physical environment.

Examining the practical implications of our findings, we concluded that, morally speaking, we should consider actions targeted at xenobots that might compromise their identity as individual entities or prevent them from exercising their capacities without a well-justified reason. This argument stems from the distinction between granting some entity moral consideration and indicating what we need to consider about it, as claimed by Hale (2011). For xenobots, this implies that despite being classified as artifacts and admitted into the circle of moral considerability, we need to consider all their traits, including those traditionally associated with animate beings.

As technology continues to evolve, so too should environmental ethics and philosophy of technology. Entities that don’t fit neatly into natural or artefactual categories will nevertheless require ethical deliberation. Recognizing this new reality, some have begun to call for a “posthuman ethics” (Ivic, 2022), “posthuman sustainability” (Cielemęcka & Daigle, 2019), and “critical posthumanism for robotics” (Dehnert, 2022). Such moves offer welcome and timely innovations. Whatever the next iteration of ethical thinking is called, it could very well inform the evolution of legal rights and responsibilities for both humans and the non-human world. As such, far from being a useless exercise conducted by those residing in privileged quarters (Birhane & van Dijk, 2020), rigorously debating the possibilities and scope of our moral obligations to intelligent machines is a task worth attending to, lest legislators and corporations write our future for us.

**References**

Ackerman, E. (2024, January 16). The Man Who Coined the Word “Robot” Defends Himself. *IEEE Spectrum*. https://spectrum.ieee.org/karel-capek-robots

Ball, P. (2020). Living Robots. *Nature Materials*, *19*(3), Article 3. https://doi.org/10.1038/s41563-020-0627-6

Bennett, J. (2010). A Vitalist Stopover on the Way to New Materialism. In D. Coole & S. Frost (Eds.), *New Materialisms: Ontology, Agency, and Politics* (pp. 47–69). Duke University Press.

Birch, T. H. (1993). Moral Considerability and Universal Consideration. *Environmental Ethics*, *15*(4), 313–332. https://doi.org/10.5840/enviroethics19931544

Birhane, A., & van Dijk, J. (2020). Robot Rights? Let’s Talk About Human Welfare Instead. *Proceedings of the 2020 Association for the Advancement of Artificial Intelligence/Association for Computing Machinery Conference on AI, Ethics, and Society*, 1–7.

Blackiston, D., Kriegman, S., Bongard, J., & Levin, M. (2023). Biological Robots: Perspectives on an Emerging Interdisciplinary Field. *Soft Robotics*, *10*(4), 674–686. https://doi.org/10.1089/soro.2022.0142

Blackiston, D., Lederer, E., Kriegman, S., Garnier, S., Bongard, J., & Levin, M. (2021). A Cellular Platform for the Development of Synthetic Living Machines. *Science Robotics*, *6*(52), eabf1571. https://doi.org/10.1126/scirobotics.abf1571

Blok, V. (2017). Earthing Technology: Toward an Eco-centric Concept of Biomimetic Technologies in the Anthropocene. *Techné: Research in Philosophy and Technology*, *21*(2–3), 127–149. https://doi.org/10.5840/techne201752363

Cahen, H. (1988). Against the Moral Considerability of Ecosystems. *Environmental Ethics*, *10*(3), 195–216. https://doi.org/10.5840/enviroethics198810315

Celermajer, D., Schlosberg, D., Wadiwel, D., & Winter, C. (2023). A Political Theory for a Multispecies, Climate-Challenged World: 2050. *Political Theory*, *51*(1), 39–53. https://doi.org/10.1177/00905917221128833

Cera, A. (2017). The Technocene or Technology as (Neo)environment. *Techné: Research in Philosophy and Technology*, *21*(2–3), 243–281. https://doi.org/10.5840/techne201710472

Cielemęcka, O., & Daigle, C. (2019). Posthuman Sustainability: An Ethos for our Anthropocenic Future. *Theory, Culture & Society*, *36*(7–8), 67–87. https://doi.org/10.1177/0263276419873710

Coeckelbergh, M. (2021). Should We Treat Teddy Bear 2.0 as a Kantian Dog? Four Arguments for the Indirect Moral Standing of Personal Social Robots, with Implications for Thinking About Animals and Humans. *Minds and Machines*, *31*(3), 337–360. https://doi.org/10.1007/s11023-020-09554-3

Coeckelbergh, M., & Gunkel, D. J. (2014). Facing Animals: A Relational, Other-Oriented Approach to Moral Standing. *Journal of Agricultural and Environmental Ethics*, *27*(5), 715–733. https://doi.org/10.1007/s10806-013-9486-3

Conty, A. (2017). How to Differentiate a Macintosh from a Mongoose: Technological and Political Agency in the Age of the Anthropocene. *Techné: Research in Philosophy and Technology*, *21*(2–3), 295–318. https://doi.org/10.5840/techne2017102473

Coole, D. (2005). Rethinking Agency: A Phenomenological Approach to Embodiment and Agentic Capacities. *Political Studies*, *53*(1), 124–142. https://doi.org/10.1111/j.1467-9248.2005.00520.x

DeGrazia, D. (2022). Robots with Moral Status? *Perspectives in Biology and Medicine*, *65*(1), 73–88. https://doi.org/10.1353/pbm.2022.0004

Dehnert, M. (2022). Toward a Critical Posthumanism for Social Robotics. *International Journal of Social Robotics*, *14*(9), 2019–2027. https://doi.org/10.1007/s12369-022-00930-w

Drengson, A. (1992). The Long-Range Deep Ecology Movement and Arne Naess. *Trumpeter*, *9*, 43–45.

Elder, A. (2020). “How Could You Even Ask That?”: Moral Considerability, Uncertainty and Vulnerability in Social Robotics. *The Journal of Sociotechnical Critique*, *1*(1), 1–23. https://doi.org/10.25779/g8a3-f992

Epting, S. (2010). Questioning Technology’s Role in Environmental Ethics: Weak Anthropocentrism Revisited. *Interdisciplinary Environmental Review*, *11*(1), 18–26. https://doi.org/10.1504/IER.2010.034603

Feinberg, J. (2013). The Rights of Animals and Unborn Generations. In R. Shafer-Landau (Ed.), *Ethical Theory: An Anthology* (2nd ed., pp. 372–380). Wiley-Blackwell.

Fitz-Henry, E. (2022). Multi-species Justice: A View from the Rights of Nature Movement. *Environmental Politics*, *31*(2), 338–359. https://doi.org/10.1080/09644016.2021.1957615

Forlano, L. (2017). Posthumanism and Design. *She Ji: The Journal of Design, Economics, and Innovation*, *3*(1), 16–29. https://doi.org/10.1016/j.sheji.2017.08.001

Fortuna, P., Wróblewski, Z., & Gorbaniuk, O. (2023). The Structure and Correlates of Anthropocentrism as a Psychological Construct. *Current Psychology*, *42*(5), 3630–3642. https://doi.org/10.1007/s12144-021-01835-z

Friedman, C. (2023). Ethical Concerns with Replacing Human Relations with Humanoid Robots: An Ubuntu Perspective. *AI and Ethics*, *3*(2), 527–538. https://doi.org/10.1007/s43681-022-00186-0

Gellers, J. C. (2020). *Rights for Robots: Artificial Intelligence, Animal and Environmental Law*. Routledge.

Goodpaster, K. E. (1978). On Being Morally Considerable. *The Journal of Philosophy*, *75*(6), 308–325. https://doi.org/10.2307/2025709

Gorke, M. (2003). *The Death of Our Planet’s Species: A Challenge to Ecology and Ethics* (P. Nevers, Trans.). Island Press.

Gunkel, D. J. (2012). *The Machine Question: Critical Perspectives on AI, Robots, and Ethics*. MIT Press.

Gunkel, D. J. (2018). *Robot Rights*. MIT Press.

Gunkel, D. J. (2022). The Relational Turn: Thinking Robots Otherwise. In J. Loh & W. Loh (Eds.), *Social Robotics and the Good Life: The Normative Side of Forming Emotional Bonds with Robots* (pp. 55–76). Transcript Verlag.

Gunkel, D. J. (2023). *Person, Thing, Robot: A Moral and Legal Ontology for the 21st Century and Beyond*. MIT Press.

Haff, P. (2014). Humans and Technology in the Anthropocene: Six Rules. *The Anthropocene Review*, *1*(2), 126–136.

Hale, B. (2009). Technology, the Environment and the Moral Considerability of Artefacts. In J. K. B. O. Friis, E. Selinger, & S. Riis (Eds.), *New Waves in Philosophy of Technology* (pp. 216–240). Palgrave Macmillan.

Hale, B. (2011). Moral Considerability: Deontological, not Metaphysical. *Ethics & the Environment*, *16*(2), 37–62. https://doi.org/10.2979/ethicsenviro.16.2.37

Hoły-Łuczaj, M. (2019). Artifacts and the Limitations of Moral Considerability. *Environmental Ethics*, *41*(1), 69–87. https://doi.org/10.5840/enviroethics20194116

Hoły-Łuczaj, M., & Blok, V. (2021). Hybrids and the Boundaries of Moral Considerability or Revisiting the Idea of Non-Instrumental Value. *Philosophy & Technology*, *34*(2), 223–242. https://doi.org/10.1007/s13347-019-00380-9

Hui, Y. (2017). On Cosmotechnics: For a Renewed Relation between Technology and Nature in the Anthropocene. *Techné: Research in Philosophy and Technology*, *21*(2–3), 319–341. https://doi.org/10.5840/techne201711876

Hunt, W. M. (1980). Are Mere Things Morally Considerable? *Environmental Ethics*, *2*(1), 59–65. https://doi.org/10.5840/enviroethics19802111

Ivic, S. (2022). Posthuman Ethics: The Priority of Ethical Over Ontological Status. In E. Tumilty & M. Battle-Fisher (Eds.), *Transhumanism: Entering an Era of Bodyhacking and Radical Human Modification* (pp. 175–187). Springer. https://doi.org/10.1007/978-3-031-14328-1\_10

Jecker, N. S., Atiure, C. A., & Ajei, M. O. (2022). The Moral Standing of Social Robots: Untapped Insights from Africa. *Philosophy & Technology*, *35*(2), 34. https://doi.org/10.1007/s13347-022-00531-5

Johnson, L. E. (1992). Toward the Moral Considerability of Species and Ecosystems. *Environmental Ethics*, *14*(2), 145–157. https://doi.org/10.5840/enviroethics199214231

Kant, I. (2013). We Have no Duties to Animals. In R. Shafer-Landau (Ed.), *Ethical Theory: An Anthology* (2nd ed., pp. 359–360). Wiley-Blackwell.

Kim, M.-S. (2023). Meta-narratives on Machinic Otherness: Beyond Anthropocentrism and Exoticism. *AI & SOCIETY*, *38*(4), 1763–1770. https://doi.org/10.1007/s00146-022-01404-3

Kitano, N. (2006). “Rinri”: An Incitement Towards the Existence of Robots in Japanese Society. *International Review of Information Ethics*, *6*, 78–83.

Kotzé, L. J. (2020). Earth System Law for the Anthropocene: Rethinking Environmental Law Alongside the Earth System Metaphor. *Transnational Legal Theory*, *11*(1–2), 75–104. https://doi.org/10.1080/20414005.2020.1776556

Kotzé, L. J., & Kim, R. E. (2019). Earth System Law: The Juridical Dimensions of Earth System Governance. *Earth System Governance*, *1*, 100110.

Kriegman, S., Blackiston, D., Levin, M., & Bongard, J. (2020). A Scalable Pipeline for Designing Reconfigurable Organisms. *Proceedings of the National Academy of Sciences*, *117*(4), 1853–1859. https://doi.org/10.1073/pnas.1910837117

Kriegman, S., Blackiston, D., Levin, M., & Bongard, J. (2021). Kinematic Self-replication in Reconfigurable Organisms. *Proceedings of the National Academy of Sciences*, *118*(49), e2112672118. https://doi.org/10.1073/pnas.2112672118

Laukyte, M. (2019). Against Human Exceptionalism: Environmental Ethics and the Machine Question. In M. V. D’Alfonso & D. Berkich (Eds.), *On the Cognitive, Ethical, and Scientific Dimensions of Artificial Intelligence* (pp. 325–339). Springer.

Lemmens, P., Blok, V., & Zwier, J. (2017). Toward a Terrestrial Turn in Philosophy of Technology. *Techné: Research in Philosophy and Technology*, *21*(2–3), 114–126. https://doi.org/10.5840/techne2017212/363

Leopold, A. (1949). *A Sand County Almanac and Sketches Here and There*. Oxford University Press.

Lewis, J. E., Arista, N., Pechawis, A., & Kite, S. (2018). Making Kin with the Machines. *Journal of Design and Science*, *3.5*. https://doi.org/10.21428/bfafd97b

Mitsukuni, Y., Ikko, T., & Tsune, S. (Eds.). (1985). *The Culture of ANIMA*. Mazda Motor Corporation.

Mylius, B. (2018). Three Types of Anthropocentrism. *Environmental Philosophy*, *15*(2), 159–194. https://doi.org/10.5840/envirophil20184564

Neely, E. L. (2013). Machines and the Moral Community. *Philosophy & Technology*, *27*(1), 97–111. https://doi.org/10.1007/s13347-013-0114-y

Pauketat, J. V. T., & Anthis, J. R. (2022). Predicting the Moral Consideration of Artificial Intelligences. *Computers in Human Behavior*, *136*, 107372. https://doi.org/10.1016/j.chb.2022.107372

Tavani, H. T. (2018). Can Social Robots Qualify for Moral Consideration? Reframing the Question about Robot Rights. *Information*, *9*(4), Article 4. https://doi.org/10.3390/info9040073

Taylor, P. W. (1981). The Ethics of Respect for Nature. *Environmental Ethics*, *3*(3), 197–218.

Vallverdú, J. (2011). The Eastern Construction of the Artificial Mind. *Enrahonar: Quaderns de Filosofía*, *47*, 171–185.

Verbeek, P.-P. (2009). Cultivating Humanity: Toward a Non-Humanist Ethics of Technology. In J.-K. B. Olsen, E. Selinger, & S. Riis (Eds.), *New Waves in Philosophy of Technology* (pp. 241–265). Palgrave Macmillan. https://doi.org/10.1057/9780230227279\_12

Zalasiewicz, J., Williams, M., Haywood, A., & Ellis, M. (2011). The Anthropocene: A New Epoch of Geological Time? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *369*(1938), 835–841. https://doi.org/10.1098/rsta.2010.0339

1. While Laukyte’s approach seems to be compelling and inspiring, Deep Ecology seems to be a rather difficult ally in arguing for granting moral considerability to robots. Its extension of moral consideration to inanimate beings is grounded in a radically non-anthropocentric assumption, that their moral status is warranted by their naturalness, defined as originating independently of human beings. [↑](#footnote-ref-1)
2. A related third concern is that Western scholars will mischaracterize, appropriate, or co-opt non-Western or Indigenous ideas for their own purposes. We contend, however, that this fear can be overcome by approaching research with humility, seeking advice from non-Western and Indigenous scholars, and positively engaging with marginalized groups whose perspectives they intend to address or employ so as to not disrespect or misrepresent them. [↑](#footnote-ref-2)