Moral Consideration of Nonhumans in the Ethics of Artificial Intelligence

Andrea Owe and Seth D. Baum
Global Catastrophic Risk Institute
https://gcri.org

AI & Ethics, vol. 1, no. 4 (November 2021), p.517-528, DOI 10.1007/s43681-021-00065-0
This version 22 February 2022

Abstract
This paper argues that the field of artificial intelligence (AI) ethics needs to give more attention to the values and interests of nonhumans such as other biological species and the AI itself. It documents the extent of current attention to nonhumans in AI ethics as found in academic research, statements of ethics principles, and select projects to design, build, apply, and govern AI. It finds that the field of AI ethics gives limited and inconsistent attention to nonhumans, with the main activity being a line of research on the moral status of AI. The paper argues that nonhumans merit moral consideration, meaning that they should be actively valued for their own sake and not ignored or valued just for how they might benefit humans. Finally, it explains implications of moral consideration of nonhumans for AI ethics research and practice, including for the content of AI ethics principles, the selection of AI projects, the accounting of inadvertent effects of AI systems such as via their resource and energy consumption and potentially certain algorithmic biases, and the research challenge of incorporating nonhuman interests and values into AI system design. The paper does not take positions on which nonhumans to morally consider or how to balance the interests and values of humans vs. nonhumans. Instead, the paper makes the more basic argument that the field of AI ethics should move from its current state of affairs, in which nonhumans are usually ignored, to a state in which nonhumans are given more consistent and extensive moral consideration.

Keywords: ethics, nonhumans, environmental ethics, artificial intelligence, intrinsic value, anthropocentrism

1. Introduction

The growing role of artificial intelligence (AI) technology raises important ethical questions about how AI systems should be designed and used. To date, initiatives for ethical AI have largely focused on human interests and values, such as in projects on "AI4People" [1] and "human-compatible AI" [2], two different initiatives on "AI for Humanity" [3], [4], the Partnership on AI (PAI) tenet "We will seek to ensure that AI technologies benefit and empower as many people as possible" [5], and governmental efforts such as a Chinese government report stating "The goal of AI development should be to promote the well-being of humankind" [6].

This paper advances the proposition that AI ethics should also consider the interests and values of nonhumans, including (but not necessarily limited to) nonhuman animals, the natural environment, and the AI itself. We do not argue that AI ethics should only consider nonhumans. Clearly, humans are also worthy of moral consideration. We also do not argue that all nonhumans merit moral consideration. Which particular nonhuman entities deserve moral consideration and how to weight humans vs. nonhumans are important questions, but they are also complex and controversial. Given that nonhumans have thus far gotten relatively limited attention in AI ethics, we believe it is a constructive first step to address the more basic proposition that nonhuman entities merit some nonzero, nontrivial moral

consideration, including in areas of AI ethics that currently give no moral consideration to nonhumans. By this we mean enough moral consideration to potentially merit some meaningful activity, and not a minuscule moral consideration so far down in the decimal points that it could simply be ignored. We believe this to be a widely acceptable proposition. It also sets the stage for the more difficult questions of how exactly to operationalize consideration of nonhumans in AI ethics, a matter that we leave for future work.

Moral consideration of nonhumans is an important topic in theoretical ethics, but it is also a practical issue for real-world AI systems. There are several matters at stake. First, AI can be applied for the advancement of nonhuman entities, such as for environmental protection. In a world of limited resources, there are decisions to be made about how much to invest in AI projects that benefit nonhumans. Second, AI can inadvertently harm the nonhuman world, such as via its considerable energy consumption or potentially via certain algorithmic biases. Arguably, where AI activities harm the nonhuman world, these impacts should be balanced against the benefits of AI. Third, the long-term prospect of strong AI or artificial general intelligence (AGI) may radically transform the world for humans and everything else. How an AGI should be designed and built could depend on the particulars of the moral consideration of humans as well as nonhumans, with potentially catastrophic implications for the wrong AGI design or build. This paper does not determine how exactly these various matters should be resolved. Instead, we seek to establish that these are matters that need to be resolved.

Some prior literature on AI ethics has considered nonhuman entities. A primary line of scholarship discusses the moral value of the AI itself and other computer systems [7]-[10]. Additionally, several studies applying Indigenous perspectives to AI ethics give moral consideration for nonhuman animals, the natural environment, and the AI itself [11]–[13]. Other relevant work discusses the role of AI in suffering endured by both humans and nonhumans [14] and in the design of AI systems with ethics frameworks based on ethical views held by both humans and nonhumans [15].

Whereas the literature referenced above addresses specific ethical issues and perspectives related to nonhumans, this paper addresses the more general question of the overall role of nonhumans within AI ethics. In other words, the original contribution of this paper is to provide a broad analysis of the role of nonhumans in AI ethics. The paper also informs discussions of the overarching ethical principles that should guide AI development and use. In recent years, many groups have published statements of AI ethics principles; a survey by Jobin et al. [16] identifies 84. This paper examines these and other statements of AI ethics principles in terms of their moral consideration for nonhumans. The paper also presents an argument for moral consideration for nonhumans in AI ethics, drawing on prior moral philosophy of nonhumans, especially from the field of environmental ethics, which has given extensive prior attention to the ethics of nonhumans [17]. Before turning to these matters, the paper first clarifies what we mean by moral consideration for nonhumans.

2. The Concept of Moral Consideration for Nonhumans

We use the term moral consideration to refer to the act of assigning intrinsic moral value or significance. The term moral consideration has been used in this way in prior literature, including on the ethics of AI and robotics [8] and the environment [18]. Intrinsic value is defined as that which is valuable for its own sake and not in reference to anything else [19]. It is contrasted with extrinsic value: that which is valuable for some other reason [20], [21]. One important type of extrinsic value is instrumental value: that which is valuable because it advances some intrinsic value. Often, intrinsic and instrumental value are treated as opposites and as the two main types of value in ethical discussion.

We define "nonhuman" as anything that is not human, though in doing so we do not mean to claim that all nonhumans merit moral consideration. Prior studies have argued for the intrinsic value of nonhuman animals [22], [23], living organisms [24], [25], including extraterrestrial life [26], ecosystems [24], [27], [28], abiotic nature, including in outer space [29], [30], technologically enhanced "posthumans" [31], relationships between sufficiently advanced moral agents, including advanced robots [8], AI [32], especially sentient AI [33], [34], information [35], and the universe itself [36]. The concept of "posthuman" speaks to fuzziness of the boundary between human and nonhuman: there is no definitive point at which an entity is sufficiently posthuman to no longer classify as human. As noted above, it is not our interest in this paper to adjudicate between these various arguments about which nonhumans are intrinsically valuable. We present a more general argument for intrinsically valuing nonhumans in Section 4.

The distinction between intrinsic and instrumental value is central for the ethics of nonhumans. Nonhuman entities may be considered valuable for their own sake or because they are valuable to humans. Clearly, nonhuman entities are instrumentally valuable to humans. Humans depend on natural environments for survival, such as for air, water, and food. Artifacts such as computers are also of obvious usefulness to humans. If humans are intrinsically valuable, then some nonhuman entities are instrumentally valuable. That is without question. The question is whether any nonhuman entities are also intrinsically valuable. This is perhaps the most fundamental question in the ethics of nonhumans.

Another important distinction is between interests and values. An entity's interests are that which is good for the entity. An entity's values are that which the entity considers to be good. Unless the entity is completely selfish, its interests and values diverge. For example, someone might personally enjoy and be able to afford a life of leisure, but they nonetheless work hard to address important issues because they believe that is the right thing to do. Value systems can involve chains of moral agents valuing the values of other agents: agent 1 values the values of agent 2, who values the values of agent 3, and so on. Such chains can theoretically persist ad infinitum, though in practice they typically end with some valuation of interests.

Moral consideration of nonhumans can come from placing weight on nonhumans' values and/or interests. Likewise, AI systems can morally consider nonhumans in several ways. First, they can be preprogrammed to account for the values and/or interests of nonhumans. Second, they can learn to follow the values of humans who give moral consideration to nonhumans. This is consistent with certain conceptions of "value alignment" or "human compatibility" developed in the AI ethics literature [2], though the literature does not generally examine the role of nonhumans, a notable exception being [15]. Third, they can learn to follow the values held by any nonhumans that are sufficiently intelligent that they hold moral values. Potential examples include intelligent nonhuman animals, extraterrestrials, and advanced AI systems. This could also classify as "value alignment", though it may require different computational methods than can be used to align AI systems to human values.

3. Prior Attention (or Lack Thereof) to Nonhumans in AI Ethics

With the conceptual background of the previous section in mind, we can now take a closer look at treatments of nonhumans in AI ethics. We begin by reviewing two systematic studies of statements of AI ethics: the Jobin et al. [16] survey of AI ethics guidelines and the Baum [37] survey of the goals of AGI research and development projects. These surveys permit a more quantitative assessment of the extent of attention to nonhumans in AI ethics. We then dive into some of the data points, taking a closer look at a few notable treatments of AI ethics in academia, industry, and government, followed by a discussion of AI ethics research.

Though not comprehensive, the overarching trend observed is that the field of AI ethics gives extensive moral consideration of humans and a much smaller moral consideration of nonhumans. (The field also gives extensive attention to issues that are not specific to either humans or nonhumans, such as the trustworthiness of an AI system.)

3.1. AI Ethics Principles

Jobin et al. [16] present a systematic search of AI ethics guidelines, identifying 84. Jobin et al. [16] classified the guidelines in terms of the principles they contain. They report 11 types of principles: transparency (found in 73 guidelines), justice and fairness (68), non-maleficence (60), responsibility (60), privacy (47), beneficence (41), freedom and autonomy (34), trust (28), sustainability (14), dignity (13), and solidarity (6).

Some of the principles do not involve moral consideration for either humans or nonhumans. Guidelines for transparency mainly concern the usage of AI, such as in the need for trust, interpretability, and oversight of AI systems. Responsibility concerns matters of integrity, liability, and general attention to ethics by those involved in AI development and use. Trust concerns whether AI systems and the organizations that provide them can be counted on to behave as expected. These conceptions of transparency, responsibility, and trust involve a special role for humans as the users of AI systems, but they are compatible with moral consideration for both humans and nonhumans because humans can use the AI systems in ways consistent with ethical frameworks that give moral consideration to either humans or nonhumans. For example, a human using an AI system to protect biodiversity would want to be able to trust that the AI system is in fact accomplishing this goal.

All of the other principles included in Jobin et al. [16] are applicable to moral consideration for both humans and nonhumans, though specific treatments of the principles commonly neglect nonhumans. For example, principles of justice and fairness have been mainly (perhaps exclusively) applied to human issues such as bias and discrimination among humans, but there are also important issues of justice for nonhumans [38], [39]. Principles of non-maleficence have been mainly applied to domains associated with human interests, such as cyberwarfare and economic loss, but AI can also be used to harm nonhumans. Principles of privacy may be less relevant to nonhumans, except perhaps if the AI itself merits moral status such that its privacy should be respected. Treatments of freedom and autonomy emphasize matters such as empowerment, self-determination, and freedom from surveillance and manipulation; these matters can be highly relevant to nonhumans, such as if AI is involved in the treatment of nonhuman animals held in captivity. Treatments of dignity call for AI to enhance, or at least not diminish, human dignity; the same could be said for the dignity of nonhumans. Finally, treatments of solidarity emphasize labor disruption, such as in technological unemployment; this is perhaps less applicable to nonhumans, though one can speak of, for example, solidarity between human and AI laborers, or solidarity among biological organisms against the potential future threat of AI takeover.

The two principles in which nonhumans have gotten at least some moral consideration are beneficence and sustainability. Jobin et al. [16] observe that AI ethics guidelines typically do not define benefit. When they do, the definitions are mostly in terms of humanity, society, or other concepts specific to humans. However, five guidelines call for benefits to something distinctly nonhuman: the planet (2 guidelines), the environment (2), or all sentient creatures (1). Others are ambiguous, such as the six calls for AI to benefit "everyone." Regarding sustainability, five AI guidelines call for sustaining the AI itself, its data, and the applicability of the insights it produces. These principles do not give moral consideration to either humans or nonhumans and are compatible with both. Moral consideration for humans is apparent in calls for fair and equal societies (1 guideline), peace (1), and accountability with respect to potential job losses (1). Moral consideration for nonhumans is possible in calls for

environmental protection (3 guidelines), improving ecosystems and biodiversity (1), and reducing the environmental impact of AI systems (1). However, it is unclear whether these guidelines value nonhumans intrinsically or instrumentally.

To summarize, the Jobin et al. [16] data indicate that only a small portion of AI ethics guidelines give moral consideration to nonhumans. Five guidelines call for benefits to nonhumans. Five also call for some form of environmental sustainability, though these principles do not clearly distinguish between the intrinsic and instrumental value of the environment. There are two points of overlap between the two sets of five, so eight total guidelines give explicit consideration to nonhumans. The other 76 guidelines have no attention to nonhumans. Attention to humans is extensive.

3.2. AGI Projects

Baum [37] presents a systematic search of AGI research and development projects, identifying 45. AGI does not yet exist and remains a long-term research challenge, but there are active groups working on AGI, as documented by Baum [37]. Baum classifies the projects according to several attributes including their stated goals. The categories of goals map neatly to this paper's treatment of moral consideration. 23 projects state intellectual goals, either "the intellectual accomplishment of the AGI itself" or "using the AGI to pursue intellectual goals"; these are not specific to either humans or nonhumans. 20 projects stated the goal of benefiting humanity. Other goals include benefiting ecosystems (three projects), animal welfare (two projects), generating profit for the AGI builders (two projects), and benefiting sentient beings and robots (one project). Note that some projects stated multiple types of goals. A more recent survey of AGI projects by Fitzgerald et al. [40] finds similar trends. These data are similar to the Jobin et al. [16] data: many AGI projects give moral consideration to humans, and only a small minority give moral consideration to nonhumans.

3.3. Select Notable Examples of the Treatment of Nonhumans in AI Ethics

This subsection analyzes select AI ethics statements, with emphasis on statements that are in some way important or insightful to the paper's theme of nonhumans. The selection cuts across academia, industry, and government, with some statements including contributions from multiple sectors.

Two recent academic works are explicitly calling for human-centric AI. The initiative "AI4People" [1] is, as it is mainly oriented toward human concerns. However, it also calls for "use of AI technologies within the EU that are socially preferable (not merely acceptable) and environmentally friendly (not merely sustainable but favourable to the environment)" [1, p.704]. The emphasis on favoring the environment strongly suggests it intrinsically values the environment. In contrast, the concept of "human-compatible AI" developed by Russell [2] gives no explicit moral consideration to nonhumans. Instead, it calls for AI whose "only objective is to maximize the realization of human preferences" [2, p.173]. The reference to preferences is about human values, not human interests, and so the AI could give moral consideration to nonhumans to the extent that human preferences do the same, but Russell [2] does not explicitly consider this prospect or the prospect of accounting for the preferences of nonhumans.

Among AI companies, moral consideration for humans is typical. Google's AI ethics principles state, for example, "We will seek to avoid unjust impacts on people" [41]. OpenAI writes that it pursues AI that "leads to a good outcome for humans" and "Our mission is to ensure that artificial general intelligence benefits all of humanity" [42]. Microsoft's AI ethics principles state, for example, "AI systems should treat all people fairly" and "AI systems should empower everyone and engage people" [43]. Microsoft CEO Satya Nadella [44] has also published principles and goals for AI, including "AI must be designed to assist

humanity." Nadella also states that human empathy "will be valuable in the human–A.I. world," which might imply empathy for AI systems, though a more likely interpretation is empathy for other humans while developing and using AI systems. None of the above ethics principles give explicit moral consideration to nonhumans.

Microsoft does have an initiative that appears to be rooted in part in moral consideration for nonhumans. Its "AI for Earth" initiative supports a variety of environmental management projects [45]. Some projects are rooted in the environment's instrumental value for humans, such as Agrimetrics, which aims "to help create a more productive and sustainable food system" [46]. Other projects appear more rooted in the intrinsic value of the environment and nonhumans, such as Wild Me, which seeks to avoid the extinction of nonhuman species [47]. Microsoft's support for Wild Me is strongly suggestive of it giving some moral consideration to nonhumans. The nonprofit AI for Good, is another exception which seems rooted in both instrumental and intrinsic values of the environment and nonhumans, with its focus on AI and the UN Sustainable Development Goals [48].

The same trend is observed in recent government reports on AI governance. A Chinese report states, "The goal of AI development should be to promote the well-being of humankind" and that AI "should conform to human values and ethical principles (…) and serve the progress of human civilization." The phrase "serve the progress of human civilization" appears to express human interests, whereas the phrase "human values and ethical principles" is clearly about human values. That can include human values that give moral consideration to nonhumans, though this is not explicit in the report. A European Parliament report calls for AI risk assessment in terms of "human safety, health and security" and transparency on AI input in decisions impacting "one or more persons' lives." A French national AI strategy initiative is called "AI for Humanity"; its report includes attention to environmental issues, though it is unclear whether this has any motivation in the intrinsic value of the environment [3]. Finally, a United States report from the Obama administration calls for responsible AI in order to "benefit society," "improve people's lives," and advance the "public good." Interestingly, its discussion of "applications of AI for public good" includes applications for environmental protection, some of which appear to be motivated by moral consideration for nonhumans, such as "habitat preservation strategies to maximize the genetic diversity of endangered populations." Typically, "public good" refers to good for the human public; the Obama administration appears to have used a broader definition that includes nonhumans.

Finally, there are professional societies and multistakeholder entities that produce consensus statements on AI ethics. These entities can represent significant portions of the overall field of AI, and so their statements are worth considering more closely.

The Partnership on AI (PAI) is a multistakeholder consortium with members from industry, academia, and nonprofits. It has published a list of ethics tenets [5]. Some tenets give moral consideration to humans, such as "We will seek to ensure that AI technologies benefit and empower as many people as possible." The only reference to nonhumans is the preamble, which states "We believe that artificial intelligence technologies hold great promise for raising the quality of people's lives and can be leveraged to help humanity address important global challenges such as climate change, food, inequality, health, and education." Climate change is a threat to both humans and nonhumans, so concern about it is consistent with intrinsically valuing humans and/or nonhumans. Likewise, it cannot be determined whether the preamble gives moral consideration to nonhumans. Strictly speaking, the same holds for the other challenges listed: nonhuman animals also eat food; there are inequities that cut across species; members of other species can also struggle with health; and human education can be used to advance the interests of nonhumans. Nonetheless, when people speak of the issues of food, inequality, health, and education, they typically do so with

reference to human interests, and it is likely that PAI intended its statement in this way. The reference to climate change is more ambiguous given its status as a signature environmental issue.

The Japanese Society for Artificial Intelligence has published Ethical Guidelines [49]. The guidelines give frequent moral consideration to humans. For example, its preamble states the aim "To ensure that AI research and development remains beneficial to human society." Its first principle states "Members of the JSAI will contribute to the peace, safety, welfare, and public interest of humanity." The guidelines contain nothing that is at all suggestive of moral consideration for nonhumans.

The conference Beneficial AI 2017 produced a set of AI ethics principles [50]. The principles give moral consideration to humans such as by stating "AI should provide a shared benefit for as many people as possible" and "AI technologies should benefit and empower as many people as possible." There is no explicit attention to nonhumans. However, the principles call for AI "to align with human values" and "to accomplish human-chosen objectives." As discussed throughout this paper, some human values/objectives give moral consideration to nonhumans. It cannot be determined whether the reference to human values/objectives intended to include or exclude moral consideration for nonhumans.

Finally, the Association for Computing Machinery (ACM) is an academic and professional society for computer science and adjacent fields. It has published a Code of Ethics and Professional Conduct [51]. Though not specific to AI, the ACM Code is nonetheless applicable. Much of the code grants moral consideration only to humans, such as its first principle, that "a computing professional should contribute to society and human well-being, acknowledging that all people are stakeholders in computing." In some places, it recognizes nonhumans, such as its affirmation "an obligation of computing professionals, both individually and collectively, to use their skills for the benefit of society, its members, and the environment surrounding them." This phrasing appears to intrinsically value the nonhuman environment. On the other hand, the code also states "human well-being requires a safe natural environment" as a reason for computing professionals to "promote environmental stability." This phrasing clearly articulates the environment as an instrumental value.

3.4 AI Ethics Research

The AI ethics research literature is of course an important part of the overall field of AI ethics. Although it is too vast to systematically analyze within the space of this paper. Instead, we make some more anecdotal observations, drawing on two recent collections, and discuss the potential role of nonhumans in select issues addressed in AI ethics research.

Our primary observation is that AI ethics research includes a significant line of research giving moral consideration to the AI itself, but it generally neglects other types of nonhumans. That is apparent from the literature surveyed in the Introduction, which, for brevity, only references a small fraction of the literature on the moral status of the AI itself. It is also apparent from two recent collections, the Oxford Handbook of Ethics of AI [52] (henceforth "the Handbook") and Ethics of Artificial Intelligence edited by Liao (henceforth "Liao") [53]. 5 of the Handbook's 44 chapters and 2 of Liao's 17 chapters have the moral value of AI as a significant theme. None of the chapters have other types of nonhumans as a significant theme, though some give brief mention of moral consideration of other nonhumans: the collections each have 3 chapters mentioning nonhuman animals and 1 chapter mentioning nature. While these two works are not necessarily representative of the field of AI ethics research, their contents reinforce the observation that the field has a significant line of research on the moral value of the AI itself with much less on other types of nonhumans.

A lot of AI ethics research is on specific issues raised by AI technology. Some of these issues are uniquely human issues, such that it would not make sense to consider nonhumans. Other issues also concern nonhumans, such that they could be addressed in the research. To illustrate this, we discuss two examples: algorithmic bias and autonomous weapons.

Algorithmic bias occurs when AI systems cause unfair biases, often by reproducing existing human biases found in data sets used to train the AI systems. Algorithmic bias research sometimes addresses issues in which nonhumans play no significant role, such as in algorithms used to evaluate job applications that are biased in favor of men over women [54]. Nonhumans do not apply for these jobs, so the bias is not relevant to nonhumans. In other issues, nonhumans are more significant. For example, research on language processing algorithms has found biases pertaining to human race and gender [55], [56]. Linguistic biases can also involve nonhumans, as documented in the field of ecolinguistics [57], [58]. A simple example is the convention of using "animal" to refer exclusively to nonhuman animals, when in fact humans are members of the animal kingdom. This can worsen the unfortunate tendencies for ontological and ethical anthropocentrism (Sections 4.2-4.3). Another example is the word "game", defined as animals hunted for food. It implies that nonhuman animals are good to the extent that they can be murdered for human benefit. Furthermore, "game" is an uncountable noun—one speaks of "game" in general, not "games" plural—which diminishes the individuality of the nonhuman animals classified as "game" [59]. These and other examples suggest that there could be algorithmic bias involving nonhumans. Likewise, there could be nonhuman algorithmic bias research, perhaps drawing on theories of justice for nonhumans [38], [39] similarly to human algorithmic bias research drawing on theories of social justice [56]. However, to the best of our knowledge, no AI ethics research has explored this issue, despite the proliferation of research on human-related algorithmic bias. It would appear that the study of algorithmic bias itself has a human-centric bias.

Autonomous weapons are systems that can make their own decisions of which targets to pursue and when and how to fire on them. Autonomous weapons are an important emerging issue in AI and military ethics. Autonomous weapons are generally targeted at humans and/or military infrastructure. They likewise mostly raise ethical issues that are specific to humans, such as questions of whether use of autonomous weapons violates human dignity [60]. Autonomous weapons may not raise significant issues regarding the natural environment. They do have some environmental impact, but so do other weapons technologies, and making a weapon autonomous may not significantly change its environmental impact. If there are any more distinctive issues raised, it may be if the AI in autonomous weapons is sufficiently advanced that the AI itself merits moral consideration. The possibility of moral consideration for a weapon system may be a novel issue for military ethics. Research on this possibility could operate at the interface of the literatures on autonomous weapons and robot rights.

To sum up Section 3, only a small minority of current treatments of AI ethics give any moral consideration to nonhumans, mainly research on the moral status of AI. It is not needed to build nonhumans into all work on AI ethics, but there is a clear role for nonhumans in a lot of work where it is currently neglected.

4. The Case for Moral Consideration of Nonhumans in AI Ethics

Thus far, we have explained what it means to give moral consideration to nonhumans and described the extent of moral consideration for nonhumans in existing work on AI ethics. In this section, we present an argument for why nonhumans merit moral consideration. We start with the example of biodiversity conservation, which is an especially clear case of nonhumans being intrinsically valued. We then argue against ontological anthropocentrism, which is the idea that humans are distinct from nature. We argue that humans are part of nature. Finally, we discuss different conceptions of ethical anthropocentrism, which is the

idea that humans are better than nonhumans. We argue that nonhumans have greater-than-zero intrinsic value and therefore merit moral consideration. We do not attempt to answer more difficult questions of the relative intrinsic value of humans and nonhumans.

4.1. A Preliminary Example: Biodiversity Conservation

The issue of biodiversity conservation is a good place to start because it is one in which moral consideration for nonhumans is already widespread. Biodiversity can have instrumental value to humans, such as for pharmaceuticals, plant breeding, and wildlife recreation [61]. However, recent research on the moral psychology of biodiversity conservation finds that people tend to care less about the instrumental value of biodiversity and more about its intrinsic value [62], [63]. Likewise, the Convention on Biological Diversity, an international treaty that entered into force in 1993, articulates both instrumental and intrinsic value of biodiversity. At the root of this is the moral intuition that it is fundamentally bad for another species to go extinct, even if the species is not important for humans. Those who might reject moral consideration for nonhumans should consider: do they think the extinction of a nonhuman species is unimportant unless it affects humans?

There are at least two ways that the intrinsic value of biodiversity can enter into AI ethics. One is via explicit articulations of this intrinsic value, such as in a principle "AI projects should work toward the goal of biodiversity conservation." Such projects could resemble the Wild Me project supported by the Microsoft AI for Earth program. The other is to call for AI activities to follow human values. Given that humans commonly value biodiversity for its own sake, this could, indirectly, give moral consideration to biodiversity. However, this indirect approach is less reliable. Not all humans intrinsically value biodiversity, and those who do typically also intrinsically value other things. AI activities can follow other human values and neglect biodiversity conservation. If biodiversity is to be intrinsically valued, it may be more effective to make this explicit.

In some cases, the intrinsic/instrumental value distinction is not important for biodiversity conservation. It can be worth conserving biodiversity because of its instrumental value for humans, regardless of whether it is of any intrinsic value. However, in other cases, the distinction matters. This can occur when something is of intrinsic value to humans but not to nonhumans, such that it is only worth pursuing if nonhumans are intrinsically valued. It can also occur when there are tradeoffs, i.e. something would be of intrinsic benefit to humans and intrinsic harm to nonhumans, or vice versa. For example, biodiversity conservation initiatives sometimes result in human populations being forcibly removed from a parcel of land in order to better protect the biodiversity [64]. These situations are complex, for example because the populations residing in that area are not the only humans affected. But setting these complexities aside, it follows that if biodiversity is only instrumentally valued, then such conservation initiatives would not be allowed, even if the harm to humans was just a minor inconvenience and the biodiversity conserved was enormous. Instead, arguably there should be a balance between humans and biodiversity, such that if enough biodiversity would be conserved, the conservation should proceed.

4.2. Against Ontological Anthropocentrism: Humans Are Part of Nature

Scholarship in environmental ethics often focuses on a matter that is ultimately about the nature of the world, i.e. how it is and not how it should be. This scholarship critiques the idea that humans are distinct from nature. This idea, known as ontological anthropocentrism or human/nature dualism, is seen as being at the heart of human mistreatment of nature [17], [24], [27], [65], [66]. It manifests as a failure to adequately value nature in both intrinsic and instrumental terms. By embracing the dualism, humans can damage nature in ways that ultimately hurt themselves.

Ontological anthropocentrism has a long history in human thought and has been particularly dominant in the West since the Enlightenment, and it remains prevalent today, but it lacks scientific basis. Ontological anthropocentrism can be found, for example, in beliefs that Earth is the center of the universe[1] and that humans are above the animals. These beliefs have deep cultural, theological, and linguistic roots (Section 3.4), but they do not survive scientific scrutiny. Modern science is unambiguous in documenting that Earth revolves around the Sun (or, more precisely, the two revolve around the Sun-Earth center of mass, which is below the surface of the Sun) and that humans are members of the animal kingdom, composed of the same atoms and molecules as everything else. The evidence clearly implies that we humans are not "non-natural" or "super-natural." Even unresolved scientific questions, such as on the nature of consciousness, do not point to ontological anthropocentrism. At least some nonhuman animals are likely to also be conscious, such as our primate cousins. Ongoing cognitive science research characterizes forms of consciousness that may exist across a diverse range of animal species [67]. Other nonhuman entities, including AIs, may be capable of consciousness as well.

None of this is to deny the important differences between humans and other entities. Humans are an outlier species, at least for this period of life on Earth. Human activity has had an outsized impact on global climate, biodiversity, land surface usage, mineral deposits, and much more, such that some environmental scientists refer to this era of Earth's geological and biological history as the Anthropocene. Human technology is also without parallel on Earth. Chimpanzees, dolphins, and corvids may be highly intelligent, but they are not developing AI. Perhaps there are more intelligent and capable species elsewhere in the universe, and perhaps there could be more intelligent and capable species in future periods of Earth, or more intelligent artificial entities (i.e., AI systems), but for this period of Earth, humans are an outlier.

4.3. Ethical Anthropocentrism: The Moral Significance of Being Human

Related to the idea that humans are inherently distinct from nature is the idea that humans are inherently better than nature. The former is about ontology, or the ways in which things can exist. The latter is about ethics, or the intrinsic value of different things that do or could exist. Even if one accepts that humans are part of nature, one could still argue that only humans are intrinsically valuable, or that humans are more (or less) intrinsically valuable than other entities.

Ethical anthropocentrism is specifically the idea that humans are more intrinsically valuable because they are humans. There are other reasons why one might ethically favor humans, such as because humans are more intelligent than other entities, or if one considers humans as more capable of experiencing happiness than other entities. These reasons are not anthropocentric. This is apparent from considering hypothetical nonhuman entities that are more advanced than humans in these attributes (smarter, happier, etc.), such as an advanced AI or an extraterrestrial species. If humans are favored in the real world because of these attributes, then the AI or extraterrestrial should be favored in the hypothetical world [68]. If the human is still favored in the hypothetical world, then the underlying ethics are anthropocentric.

Ethical anthropocentrism is related to ontological anthropocentrism. Both maintain that humans are categorically distinct, and both provide reasons for morally favoring humans. But they are different reasons. If humans are ontologically distinct, then they could be morally favored due to them being ontologically distinct. This is not ethical anthropocentrism: anything else that is also ontologically distinct (perhaps an advanced AI or extraterrestrial)

[1] In isolation, this belief is strictly speaking geocentric, not anthropocentric. However, as the idea manifests, it relates strongly to ontological anthropocentrism.

would also be morally favored. In contrast, ethical anthropocentrism would favor humans even if humans are ontologically unremarkable. Ethical anthropocentrism favors humans because they are human, not because humans are ontologically special.

Literature on ethical anthropocentrism sometimes distinguishes between strong and weak forms [17]. Strong ethical anthropocentrism maintains that humans are the sole thing of intrinsic value. Weak ethical anthropocentrism places some intrinsic value on nonhumans, but still values humans more because they are human. Strong ethical anthropocentrism rejects moral consideration of nonhumans; weak ethical anthropocentrism does not.

Anthropocentrism and moral consideration touch on related but ultimately different aspects of valuation. Anthropocentrism is about bias in values that a moral agent holds. Moral consideration is about whether a moral agent gives any attention to something in the first place. Throughout this paper, we have emphasized moral consideration instead of anthropocentrism because the defining feature of work in AI ethics is the absence of attention to the intrinsic value of nonhumans. There is very little AI ethics work that explicitly argues against intrinsically valuing nonhumans. Given the evidence presented in this paper, it is entirely possible that AI ethicists generally reject strong ethical anthropocentrism and just have not yet thought to include nonhumans or taken the effort to do so.

Three major arguments against ethical anthropocentrism can be made. The first argument centers on the idea that species membership is morally irrelevant. Instead, intrinsic moral value should be rooted in other attributes such as subjective emotion (e.g., pleasure and pain), cognitive ability, or biological complexity. As long as some nonhumans possess these attributes, strong ethical anthropocentrism is mistaken, and those who favor strong ethical anthropocentrism should "expand their moral circle" to include the nonhumans that possess these attributes [22], [23]. Furthermore, if these attributes are the only sources of intrinsic value, then there is no reason to favor humans in any way, and so weak ethical anthropocentrism is also mistaken.

The second argument centers on the idea that intrinsic value should not be defined in terms of individuals of any type, human or otherwise. Instead, intrinsic value should be defined in terms of the holistic systems that individuals are part of, such as ecosystems. This perspective sees intrinsic value in the interdependent relations between members of the system and in the system itself. Because humans are at most one element of such systems, it follows that some nonhumans must also be intrinsically valuable, and therefore strong ethical anthropocentrism must be mistaken [24], [28]. Whether to adopt holistic conceptions of intrinsic value is a matter of philosophical debate. The problem with strong ethical anthropocentrism is that it requires that one rejects the holistic conceptions without even considering their merits. Furthermore, one can argue that humans have no special place within holistic systems, in which case, if such systems are the only source of intrinsic value, then weak ethical anthropocentrism is also mistaken.

The third argument pertains to social choice ethical frameworks in which moral views are derived from some aggregate of society's moral views. For example, democratic societies derive moral views from an aggregate of the views of voting citizens and often also their elected representatives. Likewise, AI ethics sometimes calls for AI systems to be "aligned" with or "extrapolated" from human values [2], [15]. Humans may not be the only beings to hold values, in which case the first argument above implies that the values of nonhumans should also be included. A social choice framework that gives equal consideration to all who hold values, human or otherwise, would go against weak ethical anthropocentrism. However, even if only human values are included, the derived moral view can still give moral consideration to nonhumans if some humans do. Indeed, moral psychology research finds that it is quite common for humans to intrinsically value nonhumans. Studies have found humans to place significant intrinsic value on nonhuman animals [69], wildlife [70], biodiversity [62],

[63], and ecosystems [71], and there is some evidence that some humans also intrinsically value AI and robots [72], [73]. To insist upon strong anthropocentrism requires privileging the moral views of strong ethical anthropocentrists over the views of everyone else. However, common arguments for using an aggregate of society's moral views emphasize that everyone's views should be included, in which case strong ethical anthropocentrism must be rejected.

A case for ethical anthropocentrism posits that the fact that we are human gives us special relations with other humans and moral reasons to favor humans over other species. Strong ethical anthropocentrism requires that we privilege human relations over all other factors, including other types of relations. Weak anthropocentrism only requires that we recognize human relations as one morally significant factor, potentially alongside other factors.

The merits of weak ethical anthropocentrism is a more difficult matter and outside the scope of this paper. Our central argument in this paper is that nonhumans merit at least some nonzero, nontrivial moral consideration. This argument is consistent with either weak ethical anthropocentrism or ethical non-anthropocentrism, so we do not need to assess the merits of weak ethical anthropocentrism. As a point of information, we, the authors of this paper, reject weak ethical anthropocentrism, but it is not necessary for others to share this view in order to accept the arguments in this paper.

We do, for purposes of this paper, argue against strong ethical anthropocentrism. It is one thing to claim that being human gives us reason to favor humans. It is another thing to claim that being human gives us reason to not intrinsically value anything else. Each of us is more than just human. We are also members of, among other things, our families, our countries, our taxonomic kingdom (animals) and domain (eukaryotes), and our planet. Strong ethical anthropocentrism requires us to (1) privilege our species membership over our other memberships, especially our memberships in classes broader than species such as kingdom, domain, and planet, (2) reject holistic conceptions of intrinsic value without even considering the merits of such views, and (3) exclude the views of people who are not strong ethical anthropocentrists from aggregates of society's moral views. We can think of no good reason for doing these things, and so we reject strong ethical anthropocentrism.

As long as ontological anthropocentrism and strong ethical anthropocentrism are rejected, then nonhumans merit at least some nonzero, nontrivial moral consideration.

5. Implications of Moral Consideration of Nonhumans for AI Ethics

The precise implications of moral consideration of nonhumans for AI ethics depend on exactly what moral consideration is given. That includes which nonhumans get consideration. It also includes how to assess the importance of nonhumans relative to each other and relative to humans. As alluded to above, different ethical theories point in different directions on these matters, and there can be reasonable disagreement on them. Indeed, we, the authors of this paper, disagree amongst ourselves on these matters. How they should be resolved merits more attention than we are able to provide in this paper, and so we leave it for future work. Instead, here we outline some more general implications for AI ethics.

First, AI ethics research needs a robust study of the moral consideration of nonhumans. The field has thus far done little aside from the line of research on the moral status of the AI itself. One major need is to address the question of how to balance between humans and nonhumans. Another major need is to study the handling of the natural nonhuman world, including nonhuman animals and ecosystems. This has been a major blind spot in AI ethics. These topics are not unique to AI ethics, but AI technology does create distinctive challenges of how to operationalize the ethical issues in AI systems. A third major need is to consider the role of nonhumans in major AI ethics issues, such as algorithmic bias. Nonhumans could

factor significantly in these issues in ways that existing research has not adequately considered, to the extent that it has considered it at all.

Second, statements of AI ethics principles should give explicit attention to the intrinsic value of nonhumans. It is not enough to refer to human values on the grounds that some humans intrinsically value nonhumans. That leaves too much room for the intrinsic value of nonhumans being ignored, especially given how little attention nonhumans currently get in AI ethics. Exactly how to include nonhumans in the principles depends on which nonhumans are valued and how they are valued. For example, the Montréal Declaration for the Responsible Development of Artificial Intelligence includes the principle "The development and use of artificial intelligence systems (AIS) must permit the growth of the well-being of all sentient beings." This is a good example of a statement that clearly indicates the intrinsic value of sentient beings, which includes both humans and nonhumans. For illustration, an even stronger principle would be: "The main objective of development and use of AIS must be to enhance the wellbeing and flourishing of all sentient life and the natural environment, now and in the future."

Third, when selecting which AI projects to pursue, projects to advance the interests and values of nonhumans should be among the projects considered. That does not mean that those projects should always be selected. The balance of projects for humans vs. for nonhumans depends on the relative moral weight assigned to humans and nonhumans, but projects for nonhumans should sometimes be selected. The Microsoft AI for Earth program, and in particular its support of nonhuman-oriented projects like Wild Me, is a good example of how to operationalize moral consideration for nonhumans in AI project selection.

Fourth, when making decisions about which AI systems to develop and use, their inadvertent implications for nonhumans should be accounted for. This includes the material resource consumption of computer hardware and the energy needed to run AI systems. State-of-the-art AI techniques, such as deep learning, require large amounts of computing power, which in turn require large amounts of energy. Despite the growing emphasis on energy sources with low greenhouse gas emissions (mainly wind and solar, and to a lesser extent other renewables and nuclear), energy continues to come mainly from high-emission fossil fuel sources [74]. This drives global warming, which harms nonhumans. Recent attempts to quantify and raise awareness about AI energy consumption are constructive steps [75], [76]. Assessing the implications of energy consumption on nonhumans—and, for that matter, on humans—is a major undertaking. AI analysts should not take this on themselves, but instead should leverage existing work and expertise from fields such as environmental economics. AI groups should acknowledge that, in some circumstances, the resource and energy usage of an AI system may cause sufficient harm that it would be better to not use the AI system in the first place. Particular circumstances should be assessed on a case-by-case basis depending on the extent of resource and energy usage and other factors, and the extent of the benefits from the operation of the AI system.

Fifth, AI research should investigate how to incorporate nonhuman interests and values into AI system designs. How to incorporate human values is currently a major subject of study in AI, but some of the proposed techniques do not apply to nonhumans. For example, Russell [2] proposes for AI systems to derive human values from human behavior. Setting aside long-recognized problems with this approach even within the human context [77], it is clear that the approach does not straightforwardly apply for nonhumans that do not "behave" in the same sense as humans, such as ecosystems, inorganic matter, or inanimate human artifacts. Here lie compelling and challenging research questions at the intersection of philosophy, environmental science, and computer science.

AI ethics design is of particular importance for certain long-term AI scenarios in which an AGI takes a major or dominant position within human society, the world at large, and even

broader portions of outer space. Even the most well-designed AGI could be catastrophic for some nonhumans if it is designed to advance the interests of humans or other nonhumans. Furthermore, an AGI or other sufficiently advanced AI may merit moral consideration in ways comparable to humans, raising profound questions of how to balance the interests and values of humans and AIs. AGI projects should think especially carefully about which nonhumans to include in the AGI's value system, how to balance concern for humans and nonhumans, and how to operationalize these values in the AGI technology.

6. Conclusion

AI technology is important in many ways, including to both human society and to nonhumans. Whereas some prior work in AI ethics has considered specific topics related to nonhumans, this paper lays out more general considerations and calls for the whole field to move toward moral consideration for nonhumans. As AI becomes increasingly impactful across society, the extent to which AI ethics includes the nonhuman world will be important. Nonhumans merit moral consideration across all stages of the AI system life cycle, from data collection to design, deployment, and use. Further work is needed to explore which particular consideration to give nonhumans: which to include and how to include them. Some of this can draw on prior scholarship in moral philosophy, including on environmental ethics and computer ethics. However, AI ethics will need to do original work on how to value the AI itself and how to incorporate all of this into AI system design. Given the high stakes, this is important work to pursue.

Acknowledgments

Robert de Neufville, Tony Barrett, Migle Laukyte, Scott David, Kaj Sotala, Karim Jebari, and participants in a dialog hosted by the Global Catastrophic Risk Institute provided helpful feedback on an earlier version of this paper. McKenna Fitzgerald provided assistance in formatting the manuscript. Any remaining errors are the authors' alone. The views expressed here are those of the authors and do not necessarily reflect the views of the Global Catastrophic Risk Institute.

Conflict of Interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

[1] Floridi, L. et al.: AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, *28*(4), 689–707 (2018). doi: 10.1007/s11023-018-9482-5

[2] Russell, S. J.: *Human Compatible: Artificial intelligence and the problem of control*. Viking, New York (2019)

[3] *AI for humanity*, AI for Humanity. https://www.aiforhumanity.fr. Accessed 27 August 2020

[4] *AI for humanity*, Mila. https://mila.quebec/en/ai-society/. Accessed 27 August 2020

[5] *Tenets*, Partnership on AI. https://www.partnershiponai.org/tenets/. Accessed 27 August 2020

[6] National Governance Committee for the New Generation Artificial Intelligence (NGCNGAI): *Governance principles for the new generation artificial intelligence-Developing responsible artificial intelligence*, China Daily. https://www.chinadaily.com.cn/a/201906/17/WS5d07486ba3103dbf14328ab7.html (2018). Accessed 27 August 2020

[7] Floridi, L. & Sanders, J. W.: On the morality of artificial agents. *Minds and Machines, 14*(3), 349–379 (2004). doi:10.1023/B:MIND.0000035461.63578.9d
[8] Coeckelbergh, M.: Robot rights? Towards a social-relational justification of moral consideration. *Ethics and Information Technology*, *12*(3), 209–221 (2010). doi: 10.1007/s10676-010-9235-5
[9] Gunkel, D. J.: *Robot Rights*. MIT Press, Cambridge, MA (2018)
[10] Danaher, J.: Welcoming robots into the moral circle: A defence of ethical behaviorism. *Science and Engineering Ethics*, *26*(4), 2023-2049 (2020). doi: 10.1007/s11948-019-00119-x
[11] Abdilla, A., Arista, N., Baker, K., Benesiinaabandan, S., Brown, M., Cheung, M., et al.: Beyond imperial tools: Future-proofing technology through Indigenous governance and traditional knowledge systems. In: Abdilla, A. & Harle, J. (eds.) *Decolonising the Digital: Technology as Cultural Practice,* pp. 67–81. Tactical Space Lab, Sydney (2018)
[12] Lewis, J. E., Arista, N., Pechawis, A., & Kite, S.: Making kin with the machines. *Journal of Design and Science*. (2018). doi: 10.21428/bfafd97b
[13] Lewis, J. E. et al.: Indigenous protocol and artificial intelligence position paper. Honolulu, Hawai'i: The Initiative for Indigenous Futures and the Canadian Institute for Advanced Research (CIFAR) (2020)
[14] Sotala, K. & Gloor, L.: Superintelligence as a cause or cure for risks of astronomical suffering. *Informatica*, *41*(4), 389–400 (2017)
[15] Baum, S. D.: Social choice ethics in artificial intelligence. *AI & Society*, *35*(1), 165–176 (2020). doi: 10.1007/s00146-017-0760-1
[16] Jobin, A., Ienca, M., & Vayena, E.: The global landscape of AI ethics guidelines. *Natural Machine Intelligence*, *1*(9), 389-399 (2019). doi:10.1038/s42256-0190088-2
[17] Curry, P.: *Ecological Ethics: An introduction*, 2nd ed., fully rev. and expanded. Polity Press, Cambridge, UK; Malden, MA (2011)
[18] Katz, E.: Is there a place for animals in the moral consideration of nature? *Ethics and Animals*, *4*(3), 74–87 (2011). doi: 10.15368/ea.1983v4n3.1
[19] Rønnow-Rasmussen, T. & Zimmerman, M. J.: *Recent Work on Intrinsic Value*. Springer Netherlands, Dordrecht (2005)
[20] Baum, S. D.: Value typology in cost-benefit analysis. *Environmental Values*, *21*(4), 499–524 (2012). doi:10.2307/41714206
[21] Bradley, B.: Extrinsic value. *Philosophical Studies*, *91*(2), 109–126 (1998). doi:10.1023/a:1004269309760
[22] Singer, P.: *Animal Liberation: Towards an end to man's inhumanity to animals*. Paladin Books, London (1977)
[23] Regan, T.: *The Case for Animal Rights,* 1st ed. University of California Press, Berkeley (1983)
[24] Rolston III, H.: *Environmental Ethics: Duties to and values in the natural world*. Temple University Press, Philadelphia (1988)
[25] Taylor, P.: *Respect for Nature: A theory of environmental ethics*. Princeton University Press, Princeton (1986)
[26] Cockell, C. S.: Originism: Ethics and extraterrestrial life. *Journal of British Interdisciplinary Society, 60*(4), 147-153 (2007)
[27] Jonas, H.: *The Phenomenon of Life. Toward a philosophical biology*. Harper and Row, New York (1966)
[28] Callicott, J. B.: *In Defense of the Land Ethic. Essays in environmental philosophy*. State University of New York Press, Albany (1989)

[29] Rolston III, H.: The preservation of natural value in the solar system. In: Hargrove, E. (eds.) *Beyond Spaceship Earth: Environmental ethics and the solar system*, pp. 140-182. Sierra Club Books, San Francisco (1986)
[30] Milligan, T.: *Nobody Owns the Moon: The ethics of space exploitation*. McFarland and Company, Jefferson, NC (2015).
[31] Buchanan, A.: Human nature and enhancement. *Bioethics, 23*(3), 141-150 (2009)
[32] Hubbard, F. P.: Do androids dream? Personhood and intelligent artifacts. *Temple Law Review, 83,* 405-474 (2011)
[33] Umbrello, S. & Sorgner, S. L.: Nonconscious cognitive suffering: Considering suffering risks of embodied artificial intelligence. *Philosophies*, *4*(24), 1-15 (2019). doi: 10.3390/philosophies4020024
[34] Ziesche, S. & Yampolskiy, R.: Towards AI welfare science and policies. *Big Data and Cognitive Computing, 3*(2), 1-13 (2019). doi:10.3390/bdcc3010002
[35] Floridi, L.: On the intrinsic value of information objects and the infosphere. *Ethics and Information Technology*, *4*(4), 287–304 (2002). doi:10.1023/A:1021342422699
[36] Lupisella, M.: Cosmological theories of value: Relationalism and connectedness as foundations for cosmic creativity. In: Milligan, T. & Schwartz, J. S. J. (eds.) *The Ethics of Space Exploration,* pp. 75–91. Springer International Publishing Switzerland, Switzerland (2016)
[37] Baum, S. D.: A survey of artificial general intelligence projects for ethics, risk, and policy. Global Catastrophic Risk Institute Working Paper 17-1 (2017). doi: 10.2139/ssrn.3070741
[38] Garner, R.: *A Theory of Justice for Animals: Animal rights in a nonideal world*. Oxford University Press, New York (2013)
[39] Higgins, P., Short, D., & South, N.: Protecting the planet: A proposal for a law of ecocide. *Crime Law Soc. Change*, 59, 251–266 (2013). doi: 10.1007/s10611-013-9413-6
[40] Fitzgerald, M., Boddy, A., & Baum, S.D.: 2020 Survey of artificial general intelligence projects for ethics, risk, and policy. Global Catastrophic Risk Institute Technical Report 20–1 (2020)
[41] Pichai, S.: *AI at Google: Our principles*. Google. https://blog.google/technology/ai/ai-principles/ (7 June 2018). Accessed on 12 March 2020
[42] *OpenAI Charter.* OpenAI. https://openai.com/charter/ (9 April 2018). Accessed on 11 March 2020
[43] *Responsible AI*. Microsoft. https://www.microsoft.com/en-us/ai/responsible-ai. Accessed on 12 March 2020
[44] Nadella, S.: Microsoft's CEO Explores How Humans and A.I. Can Solve Society's Challenges—Together. *Slate Magazine*. https://slate.com/technology/2016/06/microsoft-ceo-satya-nadella-humans-and-a-i-can-work-together-to-solve-societys-challenges.html (28 June 2016). Accessed on 27 August 2020
[45] *AI for Earth*. Microsoft AI. https://www.microsoft.com/en-us/ai/ai-for-earth. Accessed on 28 August 2020
[46] *Agrimetrics*. Microsoft AI. https://www.microsoft.com/en-us/ai/ai-for-earth-agrimetrics. Accessed on 27 August 2020
[47] *Wild me joins AI for Earth*. Microsoft. https://news.microsoft.com/2018/06/14/wild-me-joins-ai-for-earth/ (14 June 2018). Accessed on 27 August 2020
[48] *AI for Good Foundation*. AI for Good Foundation. https://ai4good.org/. Accessed on 22 February 2021
[49] The Japanese Society for Artificial Intelligence: Report. The Japanese society for artificial intelligence ethical guidelines. Available:

http://www.ai-elsi.org/wp-content/uploads/2017/05/JSAI-Ethical-Guidelines-1.pdf (2017). Accessed on 27 August 2020
[50] *AI Principles*. Future of Life Institute. https://futureoflife.org/ai-principles/. Accessed on 18 August 2020
[51] ACM code of ethics and professional conduct. *Association for Computing Machinery*. https://www.acm.org/code-of-ethics. Accessed on 28 August 2020
[52] Dubber, M., Pasquale, F., & Das S. (eds.): *The Oxford Handbook of Ethics of AI*. Oxford University Press, Oxford (2020)
[53] Liao, S. M. (ed.): *Ethics of Artificial Intelligence*. Oxford University Press, Oxford (2020)
[54] Dastin, J.: Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G (11 October 2018). Accessed on 27 August 2020
[55] O'Neil, C.: *Weapons of Math Destruction: How big data increases inequality and threatens democracy.* Crown Publishers, New York (2016)
[56] Noble, S. U.: *Algorithms of Oppression*. NYU Press, New York (2018)

[57] Fill A. & Penz, H. (eds.): *The Routledge Handbook of Ecolinguistics*. Routledge, New York (2018)
[58] Stibbe, A.: *Ecolinguistics: Language, ecology and the stories we live by*, (2nd edition). Routledge, New York (2021)
[59] Heugerber, R.: Overcoming anthropocentrism with anthropomorphic and physiocentric uses of language? In: Fill A. & Penz, H. (eds.) *The Routledge Handbook of Ecolinguistics*, pp. 342-354. Routledge, New York (2018)
[60] Sharkey, A.: Autonomous weapons systems, killer robots and human dignity. *Ethics Inf. Technol.*, 21, 75–87 (2019). doi: 10.1007/s10676-018-9494-0
[61] Paul, C., Hanley, N., Meyer, S. T., Fürst, C., Weisser, W. W., & Knoke, T.: On the functional relationship between biodiversity and economic value. *Science Advances, (6)*5, eaax7712 (2020). doi: 10.1126/sciadv.aax7712
[62] Berry, P. M., Fabok, V., Blicharska, M., Bredin Y.K., Llorente M.G., Kovacs, E., et al.: Why conserve biodiversity? A multi-national exploration of stakeholders' views on the arguments for biodiversity conservation. *Biodiversity and Conservation*, *27*(7), 1741–1762 (2018). doi: 10.1007/s10531-0161173-z
[63] Bugter, R., Harrison, P., Haslett, J., & Tinch, R.: Making a better case for biodiversity conservation: The BESAFE project. *Biodiversity Conservation*, *27*(7), 1549–1560 (2018). doi: 10.1007/s10531-018-1543-9
[64] Adams, W. M. & Hutton, J.: People, parks and poverty: Political ecology and biodiversity conservation. *Conservation and Society*, *5*(2), 147–183 (2007)
[65] Garrard, G.: *Ecocriticism*, 2nd ed. Oxon, Abingdon; Routledge, New York (2012)
[66] Morton, T.: *Dark Ecology: For a logic of future coexistence*. Columbia University Press, New York (2016)
[67] Birch, J., Schnell, A. K., & Clayton, N. S.: Dimensions of animal consciousness. *Trends in Cognitive Science*, *24*(10), 789–801 (2020) doi: 10.1016/j.tics.2020.07.007
[68] Baum, S. D.: Universalist ethics in extraterrestrial encounter. *Acta Astronautica, 66*(3), 617–623 (2010). doi: 10.1016/j.actaastro.2009.07.003
[69] Johansson-Stenman, O.: Animal welfare and social decisions: Is it time to take Bentham seriously? *Ecol. Econ.*, 145, 90–103 (2018). doi: 10.1016/j.ecolecon.2017.08.019
[70] Bruskotter, J. T., Nelson, M. P., Vucetich, J. A.: Does nature possess intrinsic value? An empirical assessment of Americans' beliefs. The Ohio State University, Columbus OH (2015). doi: 10.13140/RG.2.1.1867.3129

[71] Arias-Arévalo, P., Martín-López, B., Gómez-Baggethun, E.: Exploring intrinsic, instrumental, and relational values for sustainable management of social-ecological systems. *Ecol. Soc.*, *22*(4), 43 (2017). doi: 10.5751/ES-09812-220443
[72] Sommer, K., Nielsen, M., Draheim, M., Redshaw, J., Vanman, E. J., Wilks, M.: Children's perceptions of the moral worth of live agents, robots, and inanimate objects. *J. Exp. Child Psychol.*, 187 (2019). doi: 10.1016/j.jecp.2019.06.009
[73] Nijssen, S. R. R., Müller, B. C. N., van Baaren, R. B., Paulus, M: Saving the robot or the human? Robots who feel deserve moral care. *Soc. Cogn.*, *37*(1), 41-56 (2019). doi: 10.1521/soco.2019.37.1.41
[74] BP: *Energy Outlook: 2020 Edition*. BP. https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/energy-outlook/bp-energy-outlook-2020.pdf (2020). Accessed 26 April 2020
[75] Sorbaro, M., Liu, Q., Bortone, M., & Sheik, S.: Optimizing the energy consumption of spiking neural networks for neuromorphic applications. *ArXiv191201268 CsQ-Bio* (2020)
[76] Strubell, E., Ganesh A., & McCallum, A.: Energy and policy considerations for deep learning in NLP. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics,* 3645-3650. Florence, Italy (2019)
[77] Sen, A.: Behavior and the concept of preference. *Economica*, *40*(159), 241–259 (1973)