



Helping and not Harming Animals with AI

Simon Coghlan^{1,2} · Christine Parker^{2,3}

Received: 9 January 2024 / Accepted: 2 February 2024 / Published online: 13 February 2024
© The Author(s) 2024

Abstract

Ethical discussions about Artificial Intelligence (AI) often overlook its potentially large impact on nonhuman animals. In a recent commentary on our paper about AI's possible harms, Leonie Bossert argues for a focus not just on the possible negative impacts but also the possible beneficial outcomes of AI for animals. We welcome this call to increase awareness of AI that helps animals: developing and using AI to improve animal wellbeing and promote positive dimensions in animal lives should be a vital ethical goal. Nonetheless, we argue that there is some value in focusing on technology-based harms in the context of AI ethics and policy discourses. A harms framework for AI can inform some of our strongest duties to animals and inform regulation and risk assessment impacts designed to prevent serious harms to humans, the environment, and animals.

Keywords Artificial Intelligence (AI) · Ethics · Nonhuman animals · Harm framework

1 Introduction

Ethical discussions about Artificial Intelligence (AI) often overlook its potentially large impact on sentient nonhuman animals (hereafter, animals). Leonie Bossert is one of the few to have challenged this anthropocentric focus (Bossert & Hagendorff, 2021, 2023; Owe & Baum, 2021; Singer & Tse, 2022; Ziesche, 2021). Bossert's (2023) commentary on our recent paper in this journal, *Harm to Nonhuman Animals from AI: a Systematic Account and Framework* (Coghlan & Parker, 2023), reminds us of AI's potential to improve animal lives and wellbeing, including by adding to the positive dimensions

✉ Simon Coghlan
simon.coghlan@unimelb.edu.au

¹ Centre of AI and Digital Ethics (CAIDE), School of Computing and Information Systems, The University of Melbourne, Victoria, Australia

² Australian Research Council Centre of Excellence for Automated Decision Making and Society, Melbourne, VIC, Australia

³ Melbourne Law School, The University of Melbourne, Victoria, Australia

of their lives. Arguing that going beyond ‘do no harm’ is important, Bossert (2023) proposes expanding our harms framework to a harm-benefit framework to better illuminate ethical responsibilities to the numerous animals potentially affected by AI.

We welcome this call to increase awareness of AI’s ability to help nonhuman animals and improve their lives. A ‘do no harm’ principle is a partial and ultimately inadequate account of our ethical duties to nonhuman animals in general and in relation to AI’s impact in particular. Nonetheless, for several reasons, we think it helpful to clearly articulate the possible pathways to harm. Below, we briefly recap our harms framework, discuss positive dimensions of animal wellbeing, and argue that there is some value in focusing on animal harms in the context of AI ethics and policy discourses.

2 The Harms Framework for AI and Animals

Drawing on David Fraser’s work (2012), our framework for AI identifies various pathways by which AI may harm sentient animals. First, there are intentional harms, both illegal or condemned and legal or socially accepted. For example, AI might be misused to facilitate killing endangered animals or allow farmed animals to be crammed into still smaller spaces at their expense. Second, there are unintentional harms, both direct and indirect. For example, AI or robot ‘caretakers’ might estrange humans from animals and cause us to care less about them.

Third, there are foregone benefits. While this category apparently goes beyond harming, Bossert correctly observes that it does not necessarily reflect all the possible benefits for animals that AI might bring. ‘Foregone benefits’ tends to focus on ways we might use AI to *avoid harms*—especially more severe and extensive harms—which humans currently cause animals. Examples include using AI to replace harmful scientific uses of animals and the human-driven cars that, like some science, also kill many millions of sentient beings each year. Not building such systems maintains a status quo in which animals are harmed by human activity, often on massive scales.

Identifying these harm pathways is important. As Fraser (2012) argues, it is often easier to ignore or miss certain harms than others. For instance, we may be more attuned to AI that facilitates intentional illegal violent treatment of animals than AI that facilitates unintentional harms. Similarly, we may more readily perceive immediate and direct harms to animals than we do distant indirect harms, even though the latter can be very large. At the same time, some intentional harms, such as the harm done to billions of animals on factory farms (Singer, 2023), may be associated with particularly grievous injustices.

3 Positive Dimensions of Animal Wellbeing

Bossert argues that a framework that revolves around harms rather than the positive dimensions of wellbeing, which allow animals to flourish, might “perpetuate a rather reductionist perspective on nonhuman animals” (Bossert, 2023). Perhaps

Bossert is right to fear that effect, but it is worth appreciating that our conception of harm is deliberately broad enough to avoid what might be seen as reductionism about animal wellbeing. We shall briefly explain this point.

Appreciating our duties to animals requires some understanding of their general and species-specific interests. Earlier conceptions of animal welfare tended to highlight a narrow set of harms, such as pain and distress from hunger and thirst. Fortunately, animal welfare science has begun to better recognize a variety of harms and also positive dimensions of wellbeing, including various mental states animals can have (Mellor et al., 2020).

Bossert (2023) argues for a “normatively sophisticated understanding of the good life.” Likewise, we argued that having the right conception of animal wellbeing can be vitally important for protecting their interests (Coghlan & Parker, 2023). Like Bossert, we advocate for a sufficiently comprehensive understanding of animal wellbeing rather than an overly narrow one. The nature of wellbeing is philosophically disputed and there are competing theories. Nonetheless, it may be best to interpret theories of wellbeing in sufficiently rich ways—ways that even go beyond the less reductionistic definitions found in animal welfare science (Bossert, 2023).

For example, perhaps a sufficient *hedonist theory* of wellbeing would recognise not just obvious pains and pleasures, like physical discomfort and gratification, but also a variety of emotional and social sufferings and enjoyments animals can have. An adequate *desire theory* of wellbeing might accommodate animal desires well beyond the most elementary drives. A sufficiently rich *objective list theory* might stress the intrinsic importance to animal wellbeing of not only life, growth, and reproduction, but also other elements like play, social affiliation, cross-species relationships, and emotional expression (Nussbaum, 2007).

Evidently, there are various possibilities concerning the positive dimensions of animal wellbeing. As Bossert acknowledges, our paper suggested that the negative elements of wellbeing should include the *absences*—perhaps brought about by deprivation or death—of genuine positive dimensions of wellbeing, and not just overt negative states like pain and distress. Missing out on many key positive dimensions of wellbeing can make an animal’s life go poorly. In this way, a rich conception of animal harm necessarily depends upon a rich conception of animal good.

Because the positive and negative sides of wellbeing cannot be completely separated, a sophisticated description of harm need not in itself entail a “reductionist perspective” (Bossert, 2023) on the good life for animals. However, Bossert may believe that a harms framework still runs that risk of ‘wellbeing reductionism’ by not emphasizing the promotion of positive elements of animal wellbeing as an important additional goal for AI technology.

4 The Value of a Harms Framework

We started with a harms framework because it was a logical place to begin given the scarce attention animals have received in AI ethics. However, we agree with Bossert that supplementing a harms-based framework with a benefits framework would be valuable. In particular it is crucial to investigate and raise awareness of the

benefits that AI could bring animals. One example is improving veterinary health-care (Coghlan & Quinn, 2023), but there are many others.

A benefits framework would explain various pathways along which AI might positively improve animal lives. Bossert helpfully sketches one such framework based on categories in our harms framework. Improving animal lives by a variety of means, including via new technology, is not only potentially ethically good, but may in some cases be obligatory. That said, we shall now explain why a harms-focussed framework serves an important, and sometimes independent, role.

Harming an individual makes them worse off than they are or would otherwise be. Bossert advocates going beyond ‘do no harm’. This phrase recalls the medical oath *primum non nocere*, meaning ‘first or above all do no harm’ (Smith, 2005). Such wording implies that it can be especially irresponsible to make a patient who seeks professional help worse off. The duty of nonmaleficence is indeed a stringent duty in healthcare and, surely, in many other contexts.

Of course, one might argue that the duty of beneficence for health professionals is *prima facie* as weighty as nonmaleficence. Beneficence is, after all, the primary goal of medical practice. However, the context of AI is much broader than medicine. All sorts of people and organisations design, engineer, build, sell, and use AI that could end up impacting on animals. Also, many parties associated with AI creation and implementation, such as many tech companies, are not part of professions or institutions whose primary goal is benefiting animals (or humans).

In some such cases, a stringent duty of providing benefit to animals (or indeed humans) may be lacking. But a duty to not harm sentient beings and make them worse off may nonetheless remain strong for these parties.¹ That is, even if an AI tech company or organization that uses AI does not have a specific duty to benefit animals, they would normally have an ethical duty, or so it may be argued, to ensure their AI products or tools do not harm animals. (Of course, further contextual details can matter; this makes it difficult to lay down blanket judgments about the nature of our responsibilities regarding AI.)

Another valuable feature of a harms framework relates to AI governance policy. Legal and policy responses to promoting safe and responsible AI are increasingly using risk assessment to identify and mitigate the potential harms of AI (AI Safety Summit, 2023). In this context, our focus on possible harms to animals from AI can be seen as a critical intervention in the otherwise anthropocentric development of AI risk governance.

The proposed European Union (EU) AI Act is a good example of this approach (European Parliament, 2023a, 2023b). The proposed Act is conditioned around the desirability of ‘promot[ing] the uptake of human centric and trustworthy artificial intelligence’ (European Parliament, 2023a, pp. 63, 68 Citation 1, Article 1). While it does seek to promote beneficial outcomes from AI, the primary regulatory intervention will be a requirement to conduct risk assessments to identify potential harms.

¹ But note that some moral theorists (e.g., utilitarians) may take general issue with the common belief that duties of beneficence are often weaker than nonmaleficence. Consider, for example, ongoing ethical debate about helping humans and animals in the effective altruism movement (Singer 2015).

The original draft proposed by the European Commission included consideration only of harms to humans. But due to interventions from NGOs and Green Members of the European Parliament (Chiappetta 2023), the EU Parliament's June draft of the Act also recognises that alongside ensuring AI systems are "safe, transparent, traceable, [and] non-discriminatory" for humans, they should also be "environmentally friendly" (News from EP, 2023).

At the time of writing, the new EU AI Act will require providers of AI systems deemed to be high risk to produce risk assessments that consider risks to not only humans but also the environment (European Parliament, 2023a, p. 55 Article 9.2a; European Parliament, 2023b). The providers of AI systems will also be required to make use of appropriate standards to reduce the environmental impact, particularly in terms of energy use in developing, training, and utilising these systems (European Parliament, 2023a, pp. 39–40 Article 28b.2(d)).

Our harms framework is well suited to informing and augmenting this type of policy attention to environmental risk assessment and reduction by highlighting the ways in which animals can be harmed by the material environmental impact of producing and running the hardware that supports AI systems. This includes the climate impact resulting from using enormous amounts of energy from fossil fuels and from the habitat destruction caused by many mining, manufacturing, and waste disposal processes connected with AI.

Importantly, our harms framework also outlines the ways in which the deployment of AI to assist otherwise legal economic activities, such as intensive animal agriculture or destructive mining, or to amplify illicit behaviours, such as illegal trade in wildlife or utilizing spectacles of animal cruelty for entertainment, may also harm animals in intended and unintended ways. Our harms framework can therefore suggest ways to extend both human and environmental risk assessments by considering impacts on animals. The framework also identifies a range of other harms to sentient animals, beyond those related to harms to humans and the environment, that should also be included in AI risk assessments (Coghlan & Parker, 2023).

5 Concluding Remarks

It is crucial that technologists, corporations, ethicists, scientists, and others become aware of how AI might be designed and deployed to help nonhuman animals as well as harm them. Nonetheless, we gave some reasons, related to ethical responsibilities and regulatory policy, for why it is important to have a framework that specifically details various pathways to animal harm.

In closing, we might also note that too strong a focus on possible benefits flowing from AI could promote the expansion of AI usage without adequate consideration of harms. After all, there is a tendency among some AI developers and advocates to emphasize how profoundly beneficial AI will be, including for animals. AI may well be beneficial for animals and humans alike, but there is also a chance that the benefits will be overrated and the harms great. Given the preponderance of human activities and industries that currently cause severe harm to nonhuman animals, that possibility should not be underestimated.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

Declarations

Competing Interests Nil.

Ethical Approval N/A.

Consent to Participate N/A.

Consent to Publish Granted.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- AI Safety Summit. (2023). The Bletchley Declaration by Countries Attending the AI Safety Summit, 1–2 November 2023. <https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023>. Accessed 5 Nov 2023.
- Bossert, L. N. (2023). Benefitting nonhuman animals with AI: Why going beyond “Do No Harm” is important. *Philosophy & Technology*, 36(3), 57. <https://doi.org/10.1007/s13347-023-00658-z>
- Bossert, L. N., & Hagendorff, T. (2021). Animals and AI. The role of animals in AI research and application – An overview and ethical evaluation. *Technology in Society*, 67, 101678. <https://doi.org/10.1016/j.techsoc.2021.101678>
- Bossert, L. N., & Hagendorff, T. (2023). The ethics of sustainable AI: Why animals (should) matter for a sustainable use of AI. *Sustainable Development*, 31(5), 3459–3467. <https://doi.org/10.1002/sd.2596>
- Chiappetta, A. (2023). Navigating the AI frontier: European parliamentary insights on bias and regulation, preceding the AI Act. *Internet Policy Review*, 12(4). <https://doi.org/10.14763/2023.4.1733>
- Coghlan, S., & Parker, C. (2023). Harm to Nonhuman Animals from AI: A Systematic Account and Framework. *Philosophy & Technology*, 36(2), 25. <https://doi.org/10.1007/s13347-023-00627-6>
- Coghlan, S., & Quinn, T. (2023). Ethics of using artificial intelligence (AI) in veterinary medicine. *AI & Society*. <https://doi.org/10.1007/s00146-023-01686-1>
- European Parliament. (2023a). Artificial Intelligence Act - Draft. https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html. Accessed 1 Nov 2023
- European Parliament. (2023b). ‘Artificial Intelligence Act: Deal on Comprehensive Rules for Trustworthy AI’. <https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai>. Accessed 9 Jan 2024.
- Fraser, D. (2012). A “Practical” ethic for animals. *Journal of Agricultural and Environmental Ethics*, 25(5), 721–746. <https://doi.org/10.1007/s10806-011-9353-z>
- Mellor, D., Beausoleil, N. J., Littlewood, K. E., McLean, A. N., McGreevy, P. D., Jones, B., & Wilkins, C. (2020). The 2020 Five Domains Model: Including Human-Animal Interactions in Assessments of Animal Welfare. *Animals*, 10(10), 1870. <https://doi.org/10.3390/ani10101870>

- News from EP. (2023). EU AI Act: First regulation on artificial intelligence. <https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>. Accessed 1 Nov 2023
- Nussbaum, M. C. (2007). *Frontiers of justice: Disability, nationality, species membership*. In *Frontiers of Justice*. Harvard University Press.
- Owe, A., & Baum, S. D. (2021). Moral consideration of nonhumans in the ethics of artificial intelligence. *AI and Ethics*. <https://doi.org/10.1007/s43681-021-00065-0>
- Singer, P. (2015). *The Most Good You Can Do: How Effective Altruism Is Changing Ideas about Living Ethically*. Text Publishing.
- Singer, P. (2023). *Animal Liberation Now: The Definitive Classic Renewed*. Harper Perennial.
- Singer, P., & Tse, Y. F. (2022). AI ethics: The case for including animals. *AI and Ethics*. <https://doi.org/10.1007/s43681-022-00187-z>
- Smith, C. M. (2005). Origin and Uses of Primum Non Nocere—Above All, Do No Harm! *The Journal of Clinical Pharmacology*, 45(4), 371–377. <https://doi.org/10.1177/0091270004273680>
- Ziesche, S. (2021). AI Ethics and Value Alignment for Nonhuman Animals. *Philosophies*, 6(2), 31. <https://doi.org/10.3390/philosophies6020031>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.