

AI and animal welfare

Appended below are links to research articles that have explored the potential impacts of AI on animal welfare and highlighted the absence of safeguards for the interests of nonhuman animals.

Among those, Singer and Tse (2023) contended that “many AI systems have significant impacts on animals, with the total number of animals affected annually likely to reach the tens or even hundreds of billions. We therefore argue that AI ethics needs to broaden its scope in order to deal with the ethical implications of this very large-scale impact on sentient, or possibly sentient, beings.”

As examples of AI systems, the article included computer vision, natural language processing, search engines, recommendation algorithms, robotics, and systems that use machine learning. It cited examples of such systems impacting animals including: chicken production units designed to perceive data about the chickens and the environment they are in, and then alter the environment and the lives of the chickens; dairy farms using AI controlled robotic systems to extract milk from cows; pet training systems; autonomous vehicles with variable capacity to protect animals that share the road; algorithms that may recommend videos showing cruelty to animals; and drones used by poachers and others to hunt and target animals. A specific example: “In New Zealand... a company called Aeronavics is developing a fully autonomous drone to identify where possums—a protected native animal in Australia but a feral animal considered harmful to forests in New Zealand—appear and then drop poisons to kill them.” The article also acknowledged ways that AI systems could potentially have beneficial impacts, such as reducing use of animals in toxicity testing.

The Coghlan and Parker (2023) paper noted that “Although AI ethics guidelines from organisations and governments advocate a range of ethical principles including beneficence, nonmaleficence, and, justice (Jobin et al., 2019), these are mostly formulated explicitly for humans (and to a limited degree the environment) or are anthropocentrically applied by default (Hagendorff, 2021).”

According to Singer and Tse (2023): “We found 71 AI ethics or computer science ethics courses in which we were able to assess the course materials. One course touched on the possible role of AI in wildlife preservation, but it was not concerned with the welfare of individual animals. None of the other courses discussed AI’s current or potential impact on animals at all.”

“We also reviewed 68 published statements on AI ethics from institutions, non-governmental organizations, governments, and corporations. The vast majority of the statements appeal to principles like ‘benefits to humanity’... Only 2 of the 68 statements can be said to include animals in their scope by mentioning impacts on ‘sentient beings’.”

One of those two, the 2018 [Montréal Declaration for a Responsible Development of Artificial Intelligence](#) stated: "The development and use of artificial intelligence systems (AIS) must permit the growth of the well-being of all sentient beings." and "AIS must allow individuals to pursue their preferences, so long as they do not cause harm to other sentient beings."

The other was: [Communications From the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Building Trust in Human-Centric Artificial Intelligence](#). The European Commission had set up a high-level expert group on AI tasked with drafting AI ethics guidelines and recommendations for broader AI policy, while at the same time, the European AI Alliance, an open multi-stakeholder platform with over 2700 members provided broader input. The Commission’s 2019 Communications document included:

“VI. Societal and environmental well-being

“For AI to be trustworthy, its impact on the **environment and other sentient beings** should be taken into account. Ideally, all humans, including future generations, should benefit from biodiversity and a habitable environment. Sustainability and **ecological responsibility** of AI systems should hence be encouraged. The same applies to AI solutions addressing areas of global concern, such as for instance the UN Sustainable Development Goals.”

However, in 2021, the European Commission proposed for the EU an Artificial Intelligence Act that lacked explicit attention to the welfare of nonhuman animals. The Act’s general approach was adopted by the European Council, and then by the European Parliament, which proposed many amendments. As of October 2023, final tripartite negotiations between the Commission, European Council and European Parliament are ongoing to settle on the final language that will be enacted.

The European Parliament’s proposed amendments have encouraged increased attention to “environmental monitoring, the conservation and restoration of biodiversity and ecosystems” and would provide that “In order to achieve the objectives of this Regulation, and contribute to the Union’s environmental objectives while ensuring the smooth functioning of the internal market, it may be necessary to establish recommendations and guidelines and, eventually, targets for sustainability. For that purpose the Commission is entitled to develop a methodology to contribute towards having Key Performance Indicators (KPIs) and a reference for the Sustainable Development Goals (SDGs).”
https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html

Subsequent to completion of the Singer/Tse paper, in March 2023, the Republic of Serbia issued “Ethical Guidelines for Development, Implementation and Use of Robust and Accountable Artificial Intelligence” <<https://www.ai.gov.rs/tekst/en/459/ethical-guidelines.php>> that included provisions that “the artificial intelligence systems that are developed must be in harmony with the wellbeing of humans, animals and the environment.” and defined “ethics” as encompassing “human behaviour that is considered acceptable and moral from certain points of view, where this behaviour affects other humans, animals that may feel pain, suffering, fear and stress, and ecosystems.”

UNESCO produced the first global standard on AI ethics – the “Recommendation on the Ethics of Artificial Intelligence” in November 2021, a framework that was adopted by all 193 Member States. Full Text: <https://unesdoc.unesco.org/ark:/48223/pf0000381137>

UNESCO’s considerations specified within the document included AI technologies’ potential impact on “human dignity, human rights and fundamental freedoms, gender equality, democracy, social, economic, political and cultural processes, scientific and engineering practices, animal welfare, and the environment and ecosystems”

[...]

Central to the Recommendation are four core values:

1. Respect, protection and promotion of human rights and fundamental freedoms and human dignity
2. Environment and ecosystem flourishing
3. Ensuring diversity and inclusiveness
4. Living in peaceful just, and interconnected societies

According to the UNESCO recommendations, “Environmental and ecosystem flourishing should be recognized, protected and promoted through the life cycle of AI systems. Furthermore, environment

and ecosystems are the existential necessity for humanity and other living beings to be able to enjoy the benefits of advances in AI.

“All actors involved in the life cycle of AI systems must comply with applicable international law and domestic legislation, standards and practices, such as precaution, designed for environmental and ecosystem protection and restoration, and sustainable development.”

[...]

“In the event of possible occurrence of any harm to human beings, human rights and fundamental freedoms, communities and society at large or the environment and ecosystems, the implementation of procedures for risk assessment and the adoption of measures in order to preclude the occurrence of such harm should be ensured.”

[...]

“The value of living in peaceful and just societies points to the potential of AI systems to contribute throughout their life cycle to the interconnectedness of all living creatures with each other and with the natural environment.”

[...]

Member States should introduce incentives, when needed and appropriate, to ensure the development and adoption of rights-based and ethical AI-powered solutions for disaster risk resilience; the monitoring, protection and regeneration of the environment and ecosystems; and the preservation of the planet. These AI systems should involve the participation of local and indigenous communities throughout the life cycle of AI systems and should support circular economy type approaches and sustainable consumption and production patterns.

[...]

Some examples include using AI systems, when needed and appropriate, to: (a) Support the protection, monitoring and management of natural resources. (b) Support the prediction, prevention, control and mitigation of climate-related problems. (c) Support a more efficient and sustainable food ecosystem. (d) Support the acceleration of access to and mass adoption of sustainable energy. (e) Enable and promote the mainstreaming of sustainable infrastructure, sustainable business models and sustainable finance for sustainable development. (f) Detect pollutants or predict levels of pollution and thus help relevant stakeholders identify, plan and put in place targeted interventions to prevent and reduce pollution and exposure.

In April 2023, UNESCO announced the launch of “Women4Ethical AI, a new collaborative platform to support governments and companies’ efforts to ensure that women are represented equally in both the design and deployment of AI. The platform’s members will also contribute to the advancement of all the ethical provisions in the UNESCO’s Recommendation.”

<https://www.unesco.org/en/articles/artificial-intelligence-unesco-launches-women4ethical-ai-expert-platform-advance-gender-equality?hub=32618>

There has, to date, been no indication that an analogous collaborative platform of experts might be formed to promote flourishing of environment and ecosystems, while considering animal welfare.

In July 2023, four leading corporate developers of generative AI (OpenAI, Google, Microsoft and Anthropic) announced that they had formed a joint AI safety, group "Frontier Model Forum" with plans to add more members.

"The Forum aims to help (i) advance AI safety research to promote responsible development of frontier models and minimize potential risks, (ii) identify safety best practices for frontier models, (iii) share knowledge with policymakers, academics, civil society, and others to advance responsible AI development; and (iv) support efforts to leverage AI to address society’s biggest challenges.

"The Frontier Model Forum will establish an Advisory Board to help guide its strategy and priorities."
<https://blogs.microsoft.com/on-the-issues/2023/07/26/anthropic-google-microsoft-openai-launch-frontier-model-forum/>

One of those founding members, Anthropic, has developed a constitution to help align its AI system's output (Claude.ai) with a set of ethical standards... "we want to emphasize that our current constitution is neither finalized nor is it likely the best it can be. We have tried to gather a thoughtful set of principles, and they appear to work fairly well, but we expect to iterate on it and welcome further research and feedback. One of the goals of this blog post is to spark proposals for how companies and other organizations might design and adopt AI constitutions.

"Our current constitution draws from a range of sources including the UN Declaration of Human Rights [2], trust and safety best practices, principles proposed by other AI research labs (e.g., Sparrow Principles from DeepMind), an effort to capture non-western perspectives, and principles that we discovered work well via our early research. Obviously, we recognize that this selection reflects our own choices as designers, and in the future, we hope to increase participation in designing constitutions."

In its current version, Anthropic's constitution lacks any attention to the interests of sentient nonhuman animals. <https://www.anthropic.com/index/claudes-constitution>

UNESCO, with its Recommendation on the Ethics of Artificial Intelligence, and Members of the EU tripartite completing work on the Artificial Intelligence Act for the EU, and the corporations that have established the Frontier Model Forum are all explicitly conveying their desire for ongoing expert input as standards and implementation procedures are refined. And in each of those cases there has been a dearth of attention to the interests of sentient creatures that should be remedied.

Research articles that have explored the potential impacts of AI on animal welfare and highlighted the absence of safeguards for the interests of nonhuman animals have included:

- Bendel, O. (2016). Considerations about the relationship between animal and machine ethics. *AI and Society*, 31(1), 103–108. <https://doi.org/10.1007/s00146-013-0526-3>
"Machine ethics almost exclusively concentrates on machine–human relationships rather than on machine–animal relationships.... this article contributes some basic considerations about the relationship between animal and machine ethics." \$39.95
- Bendel, Oliver. "Towards animal-friendly machines" *Paladyn, Journal of Behavioral Robotics*, vol. 9, no. 1, 2018, pp. 204-213. <https://doi.org/10.1515/pjbr-2018-0019>
"Semi-autonomous machines, autonomous machines and robots inhabit closed, semi-closed and open environments. There they encounter domestic animals, farm animals, working animals and/or wild animals. These animals could be disturbed, displaced, injured or killed. Within the context of machine ethics, the School of Business FHNW developed several design studies and prototypes for animal-friendly machines, which can be understood as moral machines in the spirit of this discipline." Free access
- Bossert, L., & 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>
"This paper is the first to give a comprehensive overview and to normatively explore the interconnection of animals and AI research as well as its application. The chapters aim to provide comments and tentative explorations of a new field, sometimes speculative first steps towards possible blind spots, which AI research as well as animal ethics communities might have overlooked." \$24.95
- Ziesche, S. (2021). AI ethics and value alignment for nonhuman animals. *Philosophies*, 6(2), 2. <https://doi.org/10.3390/philosophies6020031>
"This article is about a specific, but so far neglected peril of AI, which is that AI systems may become existential as well as causing suffering risks for nonhuman animals. The AI value alignment problem has now been acknowledged as critical for AI safety as well as very hard. However, currently it has only been attempted to align the values of AI systems with human values. It is argued here that this ought to be extended to the values of nonhuman animals since it would be speciesism not to do so." Free Access
- Owe, A., Baum, S.D. Moral consideration of nonhumans in the ethics of artificial intelligence. *AI Ethics* 1, 517–528 (2021). <https://doi.org/10.1007/s43681-021-00065-0>
"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." \$39.95
- Singer, P., Tse, Y.F. AI ethics: the case for including animals. *AI Ethics* 3, 539–551 (2023). <https://doi.org/10.1007/s43681-022-00187-z>
"This paper seeks to explore the kinds of impact AI has on nonhuman animals, the severity of these impacts, and their moral implications. We hope that this paper will facilitate the development of a new field of philosophical and technical research regarding the impacts of AI on animals, namely, the ethics of AI as it affects nonhuman animals." Free access
- Coghlan, S., Parker, C. Harm to Nonhuman Animals from AI: a Systematic Account and Framework. *Philos. Technol.* 36, 25 (2023). <https://doi.org/10.1007/s13347-023-00627-6>
"This paper provides a systematic account of how artificial intelligence (AI) technologies

could harm nonhuman animals and explains why animal harms, often neglected in AI ethics, should be better recognised." Free access

- Bossert, L.N. Benefitting Nonhuman Animals with AI: Why Going Beyond “Do No Harm” Is Important. *Philos. Technol.* 36, 57 (2023). <https://doi.org/10.1007/s13347-023-00658-z>
"When investigating the impact of AI on other animals, it is important to consider how these technologies can harm them. However, it is equally important to explore how they can be used to enable animals to live good lives and improve their wellbeing." Free Access
- Bossert, L. N., & Hagendorff, T. (2023). The ethics of sustainable AI: Why animals (should) matter for a sustainable use of AI. In: *Sustainable development* (pp. 1–9). <https://doi.org/10.1002/sd.2596>
"Most of the debate on sustainable AI takes place in an anthropocentric manner, focusing on human interests only and ignoring the interests of animals as well as the ways humans are currently treating them within the global animal industry. Within this article, we argued that an anthropocentric perspective on [sustainable development] and sustainable AI has many shortcomings." Free Access
- Hagendorff, T., Bossert, L.N., Tse, Y.F. *et al.* Speciesist bias in AI: how AI applications perpetuate discrimination and unfair outcomes against animals. *AI Ethics* 3, 717–734 (2023). <https://doi.org/10.1007/s43681-022-00199-9>
"We find that speciesist biases are solidified by many mainstream AI applications, especially in the fields of computer vision as well as natural language processing. In both cases, this occurs because the models are trained on datasets in which speciesist patterns prevail. Therefore, AI technologies currently play a significant role in perpetuating and normalizing violence against animals. To change this, AI fairness frameworks must widen their scope and include mitigation measures for speciesist biases." Free Access

Affiliations of the article authors:

- Baum, S.D, Global Catastrophic Risk Institute, PO Box 40364, Washington, DC, 20016, USA
- Oliver Bendel (oliver.bendel@fhnw.ch) School of Business, Institute for Information Systems, University of Applied Sciences and Arts Northwestern Switzerland (FHNW), Bahnhofstrasse 6, 5210, Windisch, Switzerland
- Leonie Bossert, Leonie.bossert@izew.uni-tuebingen.de) International Center for Ethics in the Sciences and Humanities, University of Tübingen, Tübingen, Germany
- Simon Coghlan (simon.coghlan@unimelb.edu.au) School of Computing and Information Systems and Centre for AI and Digital Ethics, The University of Melbourne, Melbourne, VIC, Australia & Australian Research Council Centre of Excellence for Automated Decision Making and Society, Melbourne, VIC, Australia
- Thilo Hagendorff (thilo.hagendorff@uni-tuebingen.de) Cluster of Excellence “Machine Learning: New Perspectives for Science”, University of Tuebingen, Tübingen, Germany,
- Andrea Owe (andrea.owe@gcrinstitute.org) Global Catastrophic Risk Institute, PO Box 40364, Washington, DC, 20016, USA
- Christine Parker, Australian Research Council Centre of Excellence for Automated Decision Making and Society, Melbourne, VIC, Australia & Melbourne Law School, The University of Melbourne, Melbourne, VIC, Australia
- Peter Singer, University Center for Human Values, Princeton University, Princeton, USA
- Yip Fai Tse (tseyipfai@gmail.com) University Center for Human Values, Center for Information Technology Policy, Princeton University, Princeton, USA
- Soenke Ziesche (soenke.ziesche@aipolicylabs.org) AI Policy Labs, New Delhi 110001, India