Animals and Artificial Intelligence: Nonhumans as Moral Agents?

Zwierzęta, a sztuczna inteligencja: istoty pozaludzkie jako podmioty moralne?

Barbora Baďurová
Independent Researcher, Slovakia
ORCID https://orcid.org/0000-0002-3805-9038 • barborabadurovabb@gmail.com

Abstract: There is a widely debated issue regarding the status and impact of exponentially growing artificial intelligence. The article deals with the problem of the moral agency of animals, and artificial intelligence. The author addresses several criteria for moral agents and tries to find the answer to the question of whether we can treat animals and AI as moral agents. The author uses mostly method of philosophical analysis and comparative method. The author claims that moral agency is not a necessary condition for moral status and doubts the practicality of attributing full moral agency to animals and AI. Moreover, claims that moral agency comes in degrees and different kinds and therefore we have to consider the complex nature of moral agency when dealing with moral actions. For instance, even human moral agents are not all on the same level of development as suggested not just by empirical evidence but also virtue ethics.

Keywords: artificial intelligence, animals, humans, ethics, moral agency

Streszczenie: Szeroko dyskutowanym obecnie problemem jest kwestia dotycząca statusu i roli gwałtownie rozwijającej się sztucznej inteligencji. Niniejszy artykuł podejmuje temat podmiotowości moralnej zwierząt oraz sztucznej inteligencji. Autorka, odnosząc się do wybranych kryteriów definiujących podmiot moralny, próbuje odpowiedzieć na pytanie, czy zwierzęta oraz sztuczną inteligencję można traktować jako podmioty moralne. W swojej pracy, autorka posługuje się głównie metodą analizy filozoficznej oraz metodą porównawczą. Twierdzi ona, że podmiotowość moralna nie jest warunkiem koniecznym statusu moralnego i poddaje w wątpliwość praktyczne zastosowanie przypisywania pełnej podmiotowości moralnej zwierzętom, czy sztucznej inteligencji. Ponadto, autorka twierdzi, że istnieją różne stopnie i \ rodzaje podmiotowości moralnej i dlatego, gdy mamy do czynienia z kwestią oceny moralnej danych działań, musimy brać pod uwagę całą złożoność podmiotowości moralnej. Na przykład, jak sugerują nie tylko dowody empiryczne, ale także etyka cnót, nawet ludzkie podmioty moralne nie są na tym samym poziomie rozwoju moralnego.

Słowa kluczowe: sztuczna inteligencja, zwierzęta, człowiek, etyka, podmiotowość moralna

Introduction

The main aim of the article is to investigate whether we can consider only humans as moral agents or whether we can also include in this category animals or artificial intelligence and attribute them moral agency. The article starts with a characterization of artificial intelligence as one of the key words and quite a new term. The next part is focused on moral agency and the concept of moral agent in the relation to humans, animals and artificial intelligence. Next, the author sums up and highlights the main issues of moral agency of humans, animals and artificial intelligence and presents the conclusions based on the philosophical analysis and comparison of concepts related to moral agency.

In the contemporary era, we can experience things that were unintelligible many centuries ago - we can already lead conversations with artificial intelligence and ask it even philosophical or ethical questions.

For instance, Chat GPT[1] replied to the question of whether it considers itself a moral agent as follows:

> "No, GPT (Generative Pre-trained Transformer) and similar language models like GPT-3.5 are not moral agents. They are computer programs created by OpenAI to generate human-like text based on the input they receive. These models lack consciousness, self-awareness, and moral agency. While GPT can generate text that may seem like it has moral or ethical considerations, it does not possess a true understanding of morality, consciousness, or the ability to make moral decisions. It operates based on patterns learned from large datasets and does not have personal beliefs, values, or intentions. Moral agency involves the ability to make moral judgments, be aware of one's actions, and act by ethical principles. GPT, being a machine learning model, does not have these capabilities. Any appearance of morality in its responses is a reflection of the input data it has been trained on, and it should not be interpreted as genuine moral reasoning or understanding" ChatGPT (openai.com).

However, the questions regarding the moral agency of AI (but also animals and humans) are still subject to philosophical discussion. Large Language Models like ChatGPT despite their popularity and hopes placed on them by some users are not omniscient, they do not know everything and are even prone to mistakes therefore there still exists room for human philosophical investigations.[2]

AI ethics is a rapidly growing field as the technical development in the area of new technologies, and in particular of artificial intelligence, is exponential.[3] The theme of AI is

---

[1] Chat GPT is a natural language processing tool driven by AI technology.

[2] For instance, there is a risk of so-called hallucinations of AI when they generate responses that are fabricated without any grounding in reality or facts.

[3] The discussion on AI regulation is very vivid in the EU, especially regarding the AI ACT (Kouroupis 2023; Chiapetta 2023).

discussed in the realm of practical ethics especially the subfields such as ethics of technology, digital ethics, computer ethics, robot ethics, AI ethics[4], etc. However, relatively rarely dealt with by environmental ethicists or animal ethicists.[5] In the article, we would like to focus on a related theme as well namely the problem of moral agency. In this context, questions emerge such as: Who is to be responsible? What are the criteria for moral agency? As the field of AI ethics is on the one hand very new, and the technological development is very rapid, the ethical and philosophical understanding of it is also in the process of development, therefore, the ongoing debates keep bringing new insights.

1. Artificial Intelligence

The term "artificial intelligence" or "AI" is these days frequently used in media and it is becoming a subject of discussion led not only by experts but also by lay people. Speculations appear repeatedly (sometimes accompanied by emotions such as excitement or fear) that AI will replace humans and even conquer humanity and myths emerge about the actual skills possessed by AI.[6]

But, what really is AI? As it is with most terms in human natural languages, it is quite difficult to find one universal definition of artificial intelligence that would be widely accepted, since the discourse is quite rich, and we face various approaches to the topic. Intuitively, as the term implies, AI is related to intelligence and machines.[7.]

In most definitions of AI, as pointed out by some authors, the core idea is related to intelligent agency: "The main unifying theme is the idea of an intelligent agent. We define AI as the study of agents that receive precepts from the environment and perform actions. Each

---

[4] A valuable contribution to the discussion about AI ethics is for instance made by Grozdanoff, Popov, Serafimova (2023).

[5] An article by Peter Singer and Yip Fai Tse (2023) regarding AI and animal ethics has been published recently. The authors are dealing with the impact of AI on nonhuman animals. Similarly, Puzio (2024) dealing with eco-relational approach in robot ethics. Also, similarly there are articles dealing with AI as a helpful tool for environmental protection etc. (see e.g. Baum and Owe 2023; Nordgren 2022; McGovern, Ebert-Uphoff, Gagne, and Bostrom 2022).

[6] As suggested by anonymous reviewer, see for instance website of the World Economic Forum (2023).

[7] Professionals usually distinguish, the so-called weak AI and strong AI or even various levels of AI based on their level of independence of human decisions or commands. The "weak" or "narrow" AI is according to Mitchell (2020, 40) a system that can perform only one narrowly defined task (or a small set of related tasks). For instance, Google Translate can translate various languages but cannot play chess or tic-tac-toe, etc.

The "strong AI" or sometimes called "human-level," "general" or "full-blown" or AGI – artificial general intelligence is such that can do what "humans can do, and possibly much more" (Mitchell 2020, 41). No AI program however exhibits this general level of intelligence that can be equal or higher in its complexity as human intelligence. A collection of specialized intelligences will never equate to a general intelligence as it is not determined by the quantity of skills, but rather by the cohesion and integration among those skills.

such agent implements a function that maps percept sequences to actions, and we cover different ways to represent these functions…" (Russell and Norvig 2009, vii).

This seems to be a valuable characteristic, as one of the central concepts is the notion of agents interacting with the environment. On the other hand, it may be useful to point out a characteristic of AI that is interesting for our article too, presented by Bringsjord and Govindarajulu (2018): "Artificial intelligence (AI) is the field devoted to building artificial animals (or at least artificial creatures that – in suitable contexts – appear to be animals) and, for many, artificial persons (or at least artificial creatures that – in suitable contexts – appear to be persons)."

Thus, we can understand preliminary AI as intelligent systems that were created artificially by humans. What is interesting in the characteristics above also from the point of view of our article is the connection, overlap, or similarities between AI, humans, and animals. How close are AI, humans, and animals? Can AI replace animals? Can AI replace humans? Is there anything unique about animals or humans? We will deal with some of the related ideas in the following part of the article. As sketched above we will focus on the relation between animals, humans, and AI and explore the topics from the philosophical-ethical point of view.

The debate about differences between humans and animals and what makes us humans have been dealt with by philosophy for centuries. For instance, the field of philosophical anthropology[8] is trying to focus on unique human characteristics. They are often contrasted with those of animals. Some authors, for example, claim that only humans can truly think, have a mind, and have a sense of justice or morality. However, as was pointed out these characteristics are probably not that unique. Or are they?

2. Who is a Moral Agent?

At first glance, doubts about the moral agency of humans are usually quite rare. To many people it seems self-evident that humans are capable moral agents[9]. For instance, people tend to save lives, and even children know that lying, stealing or killing is wrong[10]. We humans, are able to choose who should get more money, who should get organ transplantation, what to buy, whether to become vegetarians in order to reduce harm to animals, etc. We make moral choices, decisions, and judgements daily. We also take various moral actions. But what about animals

---

[8] See for instance contributions by philosophical anthropologist Max Scheler (2009).

[9] We will open question regarding human moral agency later.

[10] Although there can be a debate about the various cases when there can be a debate about the moral effect of e.g. lying.

or even AI? Before trying to find answers we need to start with some questions: What is moral agency in the first place? What are the requirements for it? Are they unique human characteristics?

Undoubtedly, moral agency is one of the central concepts in ethical theory. Morality is dealing with right and wrong actions (or omissions) that have the potential to harm those who have a moral status. A moral agent is a being that can be held accountable for their actions. Having sophisticated cognitive capacities seems to be (based on many arguments and definitions) a good criterion for moral agency. To be held accountable, moral agents have to understand the impact and consequences of their actions (or lack of action – omission to act when needed), understand the context, think about alternatives, and understand the concept of harm and hurting others that hold moral status. We humans seem to be very good examples of moral agents – as we possess the necessary abilities. We can understand the consequences, act intentionally, make moral decisions, etc. But are these abilities unique only to us? Moreover, are all humans capable of moral agency?

2.1. Animal Moral Agency

Various authors defend the intelligence and rationality of animals, while also advocating their moral status. We can mention, for instance, Peter Singer (2009), Tom Regan (2004), and Frans de Waal (2009) who present examples of various instances of animal rationality and intelligence, including emotional intelligence, or Jane Goodall (2012) contributing to understanding personhood[11] and rational agency of apes. Based on certain ethologists such as De Waal (2009) animals can act rationally and even morally – they have a sense of justice and fairness, they express love, or empathy, etc.

For instance, Frans de Waal writes the following:

"There is exciting new research about the origins of altruism and fairness in both ourselves and other animals. For example, if one gives two monkeys hugely different rewards for the same task, the one who gets the short end of the stick simply refuses to perform. In our own species, too, individuals reject income if they feel the distribution is unfair. Since any income should beat none at all, this means that both monkeys and people fail to follow the profit principle to the letter. By protesting against unfairness, their behavior supports both the claim that incentives matter and that there is a natural dislike of injustice" (De Waal 2019, 5).

---

[11] It is worth mentioning that the concept of rational being is related to the concept of person which was already claimed by Immanuel Kant (2018). However, in his approach it was only in relation to humans. The relation to animals was mostly presented by Peter Singer (2009) who claims that some animals are persons and that not all humans possess the necessary mental abilities to be considered persons.

Similarly, Mark Rowlands (2012) argues that animals have the capacity to act out of compassion, as evidenced by various instances. He gives an example of an elephant attempting to assist a dying family member to stand up, a captive gorilla rescuing an unconscious child who fell into her enclosure and carrying him to safety, or a golden retriever bravely intervening to protect a boy from a cougar, despite the peril it posed to herself.

Based on the examples mentioned above, we can claim that animals have a sense of justice and fairness[12] and they can express compassion. But does it make them moral agents? Does it mean they are true moral agents just like human moral agents? There seem to be important questions like: Can animals make moral judgments? Can they act otherwise? Can they make decisions about benevolent and malevolent actions? If so, what implications does it have?

It seems that certain animals can act intentionally and understand the consequences of their actions (Singer 2009). However, communication with them still seems to be quite difficult. Certainly, many pet owners, animal caregivers, or even animal scientists can observe many ways in which animals communicate similarly to humans. We can also communicate with many of them and empathize and sympathize with them just as with our own species, etc. Nevertheless, we do not use the same human languages, although apes can even learn sign language (Orlans 1998; Singer 2009). Some dogs or other species can communicate with humans with the assistance of a computer (Rodríguez-Vizzuett et al. 2023). Nevertheless, the vocabulary of non-human animals is still limited, and it is probably insufficient for interaction between moral agents.

But even if we consider such vocabulary as sufficient, could and should we judge animals for their actions? Historically, there have been various examples of animal court trials, especially in the Middle Ages (Dinzelbacher 2002). Of course, there is a difference between law and morality, however, dealing with moral issues implies taking into account similar mechanisms and ideally the law should match ethical rules. Thus, despite the historical instances of animal trials, it seems that it is not reasonable and practical to regard animals as full moral agents that can be dealt with just like human moral agents. Moreover, especially so-called wild animals have limited ways to do otherwise than they do, e.g. when they kill prey for food. Therefore, we cannot judge them as blameworthy although we can condemn the act of killing in itself. In some other cases, such as, a dog stealing a sandwich from its owner's table we often do and can say the dog is guilty and blameworthy. However, its moral agency is still

---

[12] De Waal's approach stems from an evolutionary perspective on morality.

probably on the lower level than that of humans and often its possibilities of undertaking free actions are limited by its owner.

For instance, Rowlands (2012) is skeptical about the ability of moral agency of animals although he admits that they are capable of moral emotions. He, however, grants them a special category of moral subjects. The status of a moral subject is placed in the intermediate realm between that of a moral agent and a moral patient. Moral subjects can act on moral grounds, but they do not fully meet the criteria of agents. Animals are evidently also moral patients, individuals whose interests merit consideration and who are legitimate subjects of moral concern. By establishing this threefold classification, Rowlands accomplishes two significant objectives: acknowledging that animals lack certain essential aspects of moral agency and dispelling one of the more questionable implications of the argument that animals exhibit moral behavior — the notion that they should be held accountable for their decisions. Thus, for various reasons mentioned above, it seems to be implausible to grant full moral agency to animals.

2.2. Moral Agency of Artificial Intelligence

But what would happen if we granted moral agency to AI robots? Let us explore the rational competencies of AI. Do machines think? Can AI be considered as agents and moral agents? These questions have several interesting aspects from the perspective of philosophy of mind and ethics. The term AI implies a certain form of intelligence. Can this intelligence be viewed as equal to human intelligence?

One of the views of agency of AI is a famous approach to machine minds by Alan Turing (1980), who focused on proving the possessing of intelligence via the concept of conversation. According to him, if machines behave intelligently undistinguishingly from humans, then they are as intelligent as them and this is demonstrated by their ability to lead meaningful and "natural" conversation.

What may be interesting is that his ideas did not just refer to the realm of thought experiments but were put into practice in real life. There are famous Turing test competitions that are used to measure the intelligence of AI. Many AI systems are already winning the contests in the sense that human observers cannot distinguish whether they are chatting with a real human person or an AI.

Based on the given approach, it may seem that AI is really intelligent and that it has a certain form of mind similar to human beings. Imagine for instance a case of e.g. Sophia (Greshko 2018), a humanoid robot able to have a meaningful conversation. Fortunately, or unfortunately, it seems to be more of a marketing trick than a real communication. However,

as mentioned by Turing test contests some AI can hold a meaningful conversation with humans who can be tricked into thinking that they talk to a real human person. There are various chatbots, that can lead quite meaningful conversations already in use too. Some are even able to use humor but whether they can be considered as really witty is an open question. Media frequently mention the so-called Large Language Models able to achieve general-purpose language generation, etc. Thus, based on Turing's argument, it seems that AI can think. And if it can think it can be put into the category of moral agents too. Or it cannot?

As we can see, the Turing test seems to prove that AI is already able to (looks like it can) think similarly to humans. Nevertheless, it is important to consider also some other arguments regarding thinking AI. Another highly cited approach to AI and its intelligence in the realm of philosophy is a thought experiment called Chinese Room presented by John Searle (1980). The main idea lies in the problem of interpretation and understanding in contrast to just repeating or imitation. And this approach quite undermines the one of Turing.

Searle's Chinese Room Argument from 1980, aimed at challenging the concept of "Strong" AI. It goes as follows: Searle finds himself inside a room, with native Chinese speakers located outside who are unaware of his presence within. Searle, both in this thought experiment and in real life, does not understand Chinese but is proficient in English. The Chinese speakers insert cards through a slot into the room, with questions written in Chinese on these cards. Through Searle's clandestine activities within the room, responses are generated and returned to the native Chinese speakers as output. Searle's responses are generated by consulting a rulebook, essentially a lookup table that instructs him on what Chinese responses to produce based on the input received. To Searle, the Chinese characters are nothing more than what he describes as "squiggle-squoggles." The fundamental concept here is that Searle, while inside the box, is meant to encompass everything a computer is capable of, but since he lacks an understanding of Chinese, the argument asserts that no computer could possess such comprehension. Essentially, Searle is mechanically manipulating Chinese symbols without true comprehension, and according to the argument, this is the fundamental essence of what computers do (Searle 1980). So Searl's thought experiment points out that AI cannot think just like humans do.

However, is the ability to establish this type of interaction sufficient to become a moral agent? It seems that it is not. We expect moral agents to really understand the situation and make deliberate choices, not just mechanical decisions. So, in this regard, AI appears to fail.

Another famous approach to AI and its thinking abilities has been proposed by Hubert Dreyfus. This philosopher claimed that human intelligence is unique in the sense that it uses

fast intuitive reasoning, and this cannot be replaced by machines and described in formal rules. The critique, put by Dreyfus (1972; 1992) in the second half of the 20th century has been widely debated. In simplified terms, this critique suggests that human expertise does not rely on explicit, detached, mechanical manipulation of symbolic information, such as logical formulas or Bayesian probabilities. According to this viewpoint, AI's attempts to create machines with human-like expertise are destined to fail when rooted in the symbolic paradigm (Bringsjord and Govindarajulu 2018). AI does not use intuitive judgment, does not use emotions for judgments only pre-programmed algorithms.

Many contemporary authors continue to speculate about the moral agency of AI despite the arguments mentioned above. There are still many open questions surrounding this topic. AI systems are already making ethical decisions in some cases. However, the term "ethical decision" in this context should be understood in a descriptive manner rather than a normative one. For example, AI is used to help decide who should receive a loan, a job, certain medical treatments, and court decisions. These moral decisions are based on previous human decisions and processed by AI often through deep learning.

But can AI be regarded as a moral agent[13] in the same or similar way as a human moral agent?

James Moor (2006) distinguishes the following types of machine agents:

- ethical impact agents (e.g., robot jockeys),
- implicit ethical agents (e.g., safe autopilot),
- explicit ethical agents (e.g., using formal methods to estimate utility),
- full ethical agents (who can make explicit ethical judgments and generally is competent to reasonably justify them) (Müller 2020).

An average adult human can be put into the category of the full ethical agent. Nevertheless, contemporary AI is not (yet) on such a level. There are speculations that it will possibly achieve it: "Several ways to achieve 'explicit' or 'full' ethical agents have been proposed, via programming it in (operational morality), via 'developing' the ethics itself (functional morality), and finally full-blown morality with full intelligence and sentience" (Müller 2020).

---

[13] There emerges also an often-mentioned problem of singularity – AI will become superintelligent and overcome humankind. However, as some pointed out – "People worry that computers will get too smart and take over the world, but the real problem is that they're too stupid and they've already taken over the world" (Domingos 2015, 286).

The idea of morality via ideal calculus and precise deontic logic that could be used by AI seems to be tempting. However, with ethics and morality, it seems to be not so easy. There are numerous exceptions, counterexamples, or intuitions that tend to hinder such calculations. For instance, variations of a famous trolley problem, etc. seem to show various moral intuitions shared by many people that are not so easily put into for instance basic utilitarian cost–benefit analysis (Lillehammer 2022).

Moreover, the decisions made by AI with the help of machine learning or deep learning are based on previous human decisions, that is on a sample belonging to the realm of descriptive ethics. However, as pointed out similarly already by Socrates this approach to ethics is not ideal. Also, from the point of view of logical fallacies - just because many people hold certain ideas it does not mean they are valid or right. We can also observe nowadays that AI is often multiplying the bias made by humans e.g. racial, gender, etc. (O'Connor and Liu 2023).

However, there emerge similar questions regarding AI as a moral agent as those related to animal moral agents, i.e. what practical implication would admitting AI moral agency have? It seems that communication with AI can be in some ways possible which makes it probably easier than even dealing with animals. Nevertheless, can we hold AI moral agents accountable? Can we make trials with AI? We can doubt so far that AI can reach the level of strong AI with general intelligence.

The key question is whether robots, when acting, should bear the responsibility, liability, or accountability for their actions. Alternatively, should the emphasis be on prioritizing the distribution of risk over discussions of responsibility? The conventional allocation of responsibility is already evident in various domains; for instance, a car manufacturer is accountable for the technical safety of the car, a driver is responsible for driving, a mechanic is responsible for proper maintenance, and public authorities are responsible for the technical conditions of the roads. This traditional distribution of responsibility is commonplace (Müller 2020). Decisions or actions influenced by AI often result from numerous interactions involving various stakeholders, including designers, developers, users, software, and hardware. This distributed agency naturally leads to distributed responsibility. The problem of distribution is not specific to the context of AI but it seems to gain great urgency (Taddeo and Floridi 2018; Nyholm 2018; Müller 2020).

As pointed out by Nyholm (2018) AI robots such as, for instance, self-driving cars can have the status of moral agent only in a collaborative sense with a human supervisor not like/and not as an independent moral agent.[14]

Based on the current information mentioned also above, the contemporary AI seems to still be just like a tool, programmed, although to some extent independently trained on previous human decisions.[15] We can hold it responsible only in a descriptive not in a normative sense. It would not make sense to make trials with AI only in a sense to gain an explanation and understanding of the course of actions/events.

3. Animals, AI and Humans as Moral Agents?

We can assert that humans can think[16], albeit to varying degrees depending on age, genetics, and state of their health. The cognitive ability of a human newborn is obviously at a lower level than that of a fully developed, healthy adult (but also a fully developed, healthy pig). Additionally, upbringing and education can influence the ability to make justified moral decisions. A problem is that we often make choices and act without full deliberation, and it seems quite impossible to think over all the consequences of our (moral) actions or behavior. Ideally, however, we can develop our moral judgments and critical thinking skills regarding our daily moral dilemmas, present sound reasons for our decisions, and engage in communication and discussion with others.

Non-human animals can also think (of course, depending on species, age, genetics, etc.). However, their ability to articulate their moral decisions and justify them is extremely limited compared to adult humans, and their possibilities to explain their actions to humans are very restricted. Additionally, their choices are limited by environmental factors, etc. Many animals seem to be able to feel and experience emotions due to the endocrine system and act in, for

---

[14] Nyholm (2018), however, leaves it open for cases when it could be plausible to even achieve full moral agency for robots like friends, lovers, etc.

[15] We can agree with the idea "that a proper understanding of ethical issues in AI can teach us something valuable about ourselves, and what it means to lead a free and responsible ethical life, that is, being good people beyond merely 'following a moral code.' In the end, we believe that rationality must be seen to involve more than just computing and that value rationality is beyond numbers. Such an understanding is a required step to recovering a renewed rationality of ethics, one that is urgently needed in our highly technified society." (Génova, Moreno, and González 2023).

[16] In the article we decided to base moral agency primarily on rationality and logical thought, however, as suggested by an anonymous reviewer it can be also based on vulnerability or emotionality. However, although we think emotionality is important in dealing with moral actions as it is tied to motivation to act, we assume that it is still necessary to use reason to be able to process, articulate and discuss the intended actions and find and explain plausible solutions. Vulnerability seems to be also an important parameter in morality, however, we assume it is more tied to the idea of moral patients than moral agents.

instance, a compassionate way, but they can also act in ways that we consider wrong. Nevertheless, it does not seem fair to treat them as equal partners in moral decision-making.

AI can, in some sense, also think. As we observe nowadays, it can process very large amounts of data which would be impossible for a human. It can also learn from humans to some extent and communicate with us. However, it lacks emotions and an understanding of intuitive explanations. Communication with AI can help us assess large amounts of data relevant for moral decisions; however, it still lacks the level of strong general artificial intelligence needed to judge cases exactly like humans and explain them in a way that enables meaningful moral discussion. It seems that the most important criterion and issue is the ability to articulate justifications for moral decisions. This is related to the problem that there are various approaches to ethics, various ethical theories, and what seems most interesting is the ability of critical thinking. However, as mentioned above, perhaps AI can compose interesting justifications for decisions. But would it be sufficient for moral decisions? Intuitively, it seems not, and many biased decisions generated by AI are obviously implausible. It seems that animals and AI cannot be treated as full moral agents, but they can help us better understand the nature of human moral agency.

Conclusion

The article is aimed to open a discussion and sketch problems related to the criteria for the moral agency of, on the one hand, living beings, especially animals, and, on the other, artificial intelligence. The problems of moral agency open up many questions, particularly regarding what it means to be a moral agent. Usually, it means that we can hold the moral agent accountable and blame the holder of the status of moral agent, in many cases punish them, or otherwise prevent them from repeating harmful actions. It seems reasonable even in human society to admit that there are various degrees of moral agency, meaning that some people are more capable of understanding and judging moral dilemmas than others[17]. If we consider that for just punishing and preventing from doing harmful things (or omitting to act well) we need basic rational abilities, understanding of consequences, etc., we can maybe admit that even some animals and AI can pass the test. However, would it be just and meaningful to punish animals or AI for their wrongdoing? This seems to be problematic. But it also gives a new point

---

[17] There is of course also an old problem of akrasia – weakness of will.

of view on treating human moral agents, i.e., should we blame and punish human moral agents? Or, how should we treat their wrongdoing? How to even judge whether they have done wrong?

It is also evident that moral agency comes in various degrees, not just across species or types of entities, but also within them. For instance, as pointed out by virtue ethics theory, some human moral agents are more developed than others (Dzwonkowska 2020), and many are still evolving in their ability to make moral decisions and take moral actions[18]. Therefore, when judging the acts of moral agents, we need to take into account the complex nature of moral agency.

However, we must keep in mind that full moral agency is not a necessary prerequisite for moral status. For instance, if it were, taking care of children who are not yet full moral agents would not be obligatory, which is counterintuitive and leads to harmful consequences for mankind. Thus, animals should also be granted moral status even when they are not full moral agents. However, the case of the moral status of AI is still open.

The aforementioned approaches to animal agency and the agency of AI raise many questions about human moral agency. Do we have free will?[19] Are we also machines? Or are we just animals? Can we truly make decisions, or are we, in a sense, pre-programmed? And what about the new possibilities when human/machine/AI (or even animal)[20] abilities are blended? We will, however, leave these questions open for now, as a thought-provoking conclusion for the readers of this article.

Institutional Review Board Statement: Not applicable.

Conflicts of Interest: The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

---

[18] So-called moral perfectionism is related to virtue ethics.

[19] For instance, contemporary philosopher Sam Harris (2012) among others, promotes the idea that humans lack free will based on scientific research in neurosciences.

[20] E.g., implanting microchips like Neuralink promoted by Elon Musk, or cases of human/animal chimeras, etc.

References

Baum, Seth D., and Andrea Owe. 2023. "Artificial Intelligence Needs Environmental Ethics." *Ethics, Policy & Environment*, 26(1):139-143. https://doi.org/10.1080/21550085.2022.2076538.

Behdadi, Dorna. 2021. "A Practice-Focused Case for Animal Moral Agency." *Journal of Applied Philosophy*, 38: 226-243. https://doi.org/10.1111/japp.12486.

Bringsjord, Selmer, and Han Xiao. 2000. "A Refutation of Penrose's Gödelian Case Against Artificial Intelligence." In: *Journal of Experimental and Theoretical Artificial Intelligence*, 12: 307–329. https://doi.org/10.1080/09528130050111455

Bringsjord, Selmer, and Naveen Sundar Govindarajulu. 2018. "Artificial Intelligence." In *Stanford Encyclopedia of Philosophy.* Accessed February 18, 2024. https://plato.stanford.edu/entries/artificial-intelligence/

ChatGPT, Prompt: "Are you a moral agent?" Accessed January 29, 2023. https://chat.openai.com/auth/login?next=%2Fc%2F5aa8bd43-4190-40dd-91b1-9b018b41c7f3.

Chiappetta, Allessia. 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.

Coeckelbergh, Mark. 2020. *AI Ethics*. Cambridge, MA: The MIT Press.

De Waal, Frans. 2009. *Good Natured*. Cambridge, MA: Harvard University Press.

De Waal, Frans. 2019. *The Age of Empathy. Nature's Lessons for a Kinder Society.* London: Souvenir Press.

Dinzelbacher, Peter. 2002. "Animal Trials: A Multidisciplinary Approach." *The Journal of Interdisciplinary History* 32(3): 405–421. http://www.jstor.org/stable/3656215.

Domingos, Pedro. 2017. *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World.* London: Penguin Random House.

Dreyfus, Hubert. 1972. *What Computers Can't Do*. Cambridge, MA: MIT Press.

Dreyfus, Hubert. 1992. *What Computers Still Can't Do.* Cambridge, MA: MIT Press.

Dzwonkowska, Dominika. 2020. "Environmental Vices as Ethical and Anthropological Roots of the Environmental Crisis." *Studia Ecologiae et Bioethicae* 18(5): 171-80. https://doi.org/10.21697/seb.2020.18.5.15.

Floridi, Luciano, and Taddeo Mariarosaria. 2016. "What Is Data Ethics?" *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374(2083): 20160360. https://doi.org/10.1098/rsta.2016.0360.

Génova, Gonzalo, Victor Moreno, and Marcos Román González. 2023. "Machine Ethics: Do Androids Dream of Being Good People?" *Sci Eng Ethics* 29(10). https://doi.org/10.1007/s11948-023-00433-5.

Goodall, Jane. 2012. *Jane Goodall's Animal World, Chimps.* Byron Preiss.

Greshko, Michael. 2018. "Meet Sophia, the robot that looks almost human." *National Geographic*, May 18, 2018. Online: https://www.nationalgeographic.com/photography/article/sophia-robot-artificial-intelligence-science.

Grozdanoff, Boris D., Zdravko Popov, Silviya Serafimova. 2023. *Rationality and Ethics in Artificial Intelligence.* Cambridge: Cambridge Scholars Publishing.

Harris, Sam. 2012. *Free Will.* New York: Simon & Schuster.

Kant, Immanuel. 2018. *Groundwork for the Metaphysics of Morals*. Yale: Yale University Press.

Kouroupis, Konstantinos. 2023. "The AI Act in light of the EU Digital Agenda: A critical approach." *Journal of Data Protection & Privacy* 5(3): 216-229. Accessed February 18, 2024. https://hstalks.com/article/7448/the-ai-act-in-light-of-the-eu-digital-agenda-a-cri/.

Lillehammer, Hallvard (ed). 2022. *The Trolley Problem.* Cambridge: Cambridge University Press.

McGovern, Amy, Imme Ebert-Uphoff, David John Gagne II, Ann Bostrom. 2022. "Why we need to focus on developing ethical, responsible, and trustworthy artificial intelligence approaches for environmental science." *Environmental Data Science* 1:e6. https://doi.org/10.1017/eds.2022.5.

Mitchell, Melanie. 2020. *Artificial Intelligence. A Guide for Thinking Humans.* New York: Pelican Press.

Moor, James H. 2006. "The Nature, Importance, and Difficulty of Machine Ethics." *IEEE Intelligent Systems* 21(4): 18–21. https://doi.org/10.1109/MIS.2006.80.

Müller, Vincent C. 2020. "Ethics of Artificial Intelligence and Robotics." In *Stanford Encyclopedia of Philosophy.* Accessed February 18, 2024. https://plato.stanford.edu/entries/ethics-ai/.

Nordgren, Anders. 2022. "Artificial intelligence and climate change: ethical issues." *Journal of Information, Communication and Ethics in Society* 21(1): 1-15. https://doi.org/10.1108/JICES-11-2021-0106.

Nyholm, Sven. 2018. "Attributing Agency to Automated Systems: Reflections on Human–Robot Collaborations and Responsibility-Loci." *Science and Engineering Ethics* 24(4): 1201–1219. https://doi.org/10.1007/s11948-017-9943-x.

O'Connor, Sinead, Helen Liu. 2023. "Gender bias perpetuation and mitigation in AI technologies: challenges and opportunities." *AI & Society.* https://doi.org/10.1007/s00146-023-01675-4.

Orlans, F. Barbara. 1998. *The Human Use of Animals: Case Studies in Ethical Choice*. New York: Oxford University Press.

Puzio, Anna. 2024. "Not Relational Enough? Towards an Eco-Relational Approach in Robot Ethics." *Philosophy and Technology* 37: 45. https://doi.org/10.1007/s13347-024-00730-2.

Regan, Tom. 2004. *Case for Animal Rights.* Oakland: University of California Press.

Rodríguez-Vizzuett, Liliana, Ismael Espinosa-Curiel, Humberto Pérez-Espinosa. 2022. "Digital Technology Supporting the Remote Human-Dog Interaction. Scoping Review." *Animals* 13(4):699. https://doi.org/10.3390/ani13040699.

Rowlands, Mark. 2012. *Can Animals Be Moral?* New York: OUP USA.

Russell, Stuart, and Peter Norvig. 2009. *Artificial Intelligence: A Modern Approach.* Saddle River, NJ: Prentice Hall.

Scheler, Max. 2009. *The Human Place in the Cosmos.* Evanston: Northwestern University Press.

Searle, John. 1980. "Minds, Brains and Programs." *Behavioral and Brain Sciences* 3: 417–424. https://doi.org/10.1017/S0140525X00005756.

Shapiro, Paul. 2006. "Moral Agency in Other Animals." *Theoretical Medicine and Bioethics* 27(4):357-373. https://doi.org/10.1007/s11017-006-9010-0.

Singer, Peter, Yip Fai Tse. 2023. "AI ethics: the case for including animals." *AI Ethics* 3:539–551. https://doi.org/10.1007/s43681-022-00187-z.

Singer, Peter. 2009. *Animal Liberation: The Definitive Classic of the Animal Movement.* New York: Ecco Book/Harper Perennial.

Turing, Alan. 1980. "Computing Machinery and Intelligence." *Creative Computing* 6(1): 44-53.

World Economic Forum. 2023. "These are the jobs most likely to be lost – and created – because of AI." Accessed April 04, 2024. https://www.weforum.org/agenda/2023/05/jobs-lost-created-ai-gpt/.