Trends in
Cognitive Sciences

CellPress
OPEN ACCESS

Science & Society

What will society think about AI consciousness? Lessons from the animal case

Lucius Caviola[1,4,*], Jeff Sebo[2,4], and Jonathan Birch[3]

How will society respond to the idea that artificial intelligence (AI) could be conscious? Drawing on lessons from perceptions of animal consciousness, we highlight psychological, social, and economic factors that shape perceptions of AI consciousness. These insights can inform emerging debates about AI moral status, ethical treatment, and future policy.

AI consciousness as an emerging societal issue

AI is rapidly advancing, becoming more human-like and embedded in everyday life. Imagine having a video call, only to discover later you were speaking to an AI dialogue agent designed to mimic human behavior, expression, and emotional nuance with remarkable accuracy. We increasingly interact with AI systems playing roles such as assistant, coworker, coach, therapist, or even romantic partner.

As AI systems become more complex in their architectures, a pressing question arises: could they one day become conscious? In this article, we define consciousness as subjective experience – what philosophers call 'phenomenal consciousness'. This means that there is 'something it is like' to be that system. Think, for example, of the feeling of seeing red or experiencing happiness. While today's AI systems likely lack this capacity, the idea of AI consciousness is gaining serious attention in both academic and public discourse [1–6].

Regardless of whether AI systems ever become conscious (Box 1), a further pressing question arises: will society perceive them as conscious? Will society come to accept AI systems as conscious beings, reject the idea, or remain in a state of disagreement and uncertainty? Either way, what will follow for our lives and societies?

These questions matter because consciousness is widely viewed as a key condition for moral status. According to this view, if AI systems were conscious, they would have interests, raising important ethical questions. Whether society accepts, denies, or disagrees about AI consciousness could shape norms and laws in profound ways. Fortunately, our treatment of animals offers a useful, if imperfect, starting point for understanding what may lie ahead.

Attitudes about AI may resemble attitudes about animals

AI systems and animals are both nonhumans that share important parallels in how they resemble us and relate to us. What can our past with animals tell us about our future with AI?

Appearance and behavior

Humans are more likely to attribute consciousness to non-human animals who look and act like us, like great apes, than to those who do not, like insects [7]. This bias extends to features such as big heads, big eyes, four limbs, furry skin, and symmetrical faces. As a result, we respond to some animals with affection and empathy and to others with aversion and antipathy.

In the future, we can expect that humans will similarly be more likely to attribute consciousness to AI systems that look and act like us, so long as they are not perceived as uncanny or robotic. This includes systems that have faces and voices, speak natural language, make natural gestures, and engage in conversational turn-taking. Those without a visual avatar or the ability to directly communicate with users are less likely to elicit attributions of consciousness.

Social and economic roles

Humans are more likely to attribute consciousness to animals whom we use as companions (or other similar roles) than to animals whom we use as commodities (or other similar roles). For example, although dogs and pigs are mammals with complex cognition, we tend to attribute richer mental lives to dogs. These differences likely stem from the emotional bonds we form with pets, in contrast to the instrumental roles assigned to 'livestock' [8].

In the future, we can expect that humans will similarly be more likely to attribute consciousness to AI systems used as companions – or in other roles that involve social and emotional bonds, such as assistants or therapists – than to those used as commodities or in other roles that involve the performance of rote tasks without the need for a social interface.

Moral biases

Humans tend to assign greater moral worth to some animals than to others based on species membership alone, a tendency known as speciesism [7]. This moral devaluation is typically accompanied by lower attributions of consciousness. We exhibit a range of other biases that affect our treatment and perceived consciousness of animals as well, including a limited ability to empathize with distant strangers and a limited ability to appreciate the significance of large numbers.

Analogous biases may emerge toward potentially conscious AI systems. In addition to exhibiting a kind of speciesism against AI systems because they fail to count as human, we might exhibit substratism: favoring carbon-based beings over silicon-based ones based on material substrate alone. We

Box 1. Could AI become conscious?

Disagreement and uncertainty about AI consciousness persist among philosophers, scientists, and technical experts [4]. While current systems are unlikely to be conscious, there is wide agreement that future systems with more advanced and integrated cognitive architectures – such as those enabling recursive reasoning, global information access, or self-monitoring – could be more likely to be conscious. For example, in a 2020 survey, 39% of philosophers accepted or leaned toward the possibility of future AI consciousness, whereas 27% rejected or leaned against it [14]. And in a 2024 survey, the median AI researcher estimated a 25% chance of conscious AI by 2034 and a 70% chance by 2100 [15].

Given this ongoing disagreement and uncertainty, experts increasingly endorse a pluralistic 'marker' method for assessing AI consciousness – examining systems for features that correspond to consciousness according to multiple scientific theories, with special focus on architectural features given the current unreliability of behavioral evidence in this context [6,11]. Experts increasingly also endorse a precautionary approach for AI systems that have a realistic possibility of being conscious based on the best current evidence [1,2,6].

might also display originism: preferring naturally evolved beings over artificially designed ones based on their causal origins alone. At the same time, our already-limited ability to empathize with large, distant populations (even when they are human) will likely continue to be a significant bias.

Attitudes about AI may also differ

Of course, there will be many differences between animals and AI systems as well, which could lead to divergences in how society responds to each.

Cognitive capacities

AI systems might have an advantage over non-human animals due to their cognitive capacities. In particular, many AI systems already have advanced language, reasoning, and decision-making, allowing them to engage with us on our own terms. In the future, this similarity between human and AI cognition could be even more striking, perhaps even approaching functional identity. This may make consciousness attributions to AI systems more likely, all else being equal.

By contrast, while many non-human animals have more advanced cognitive capacities than previously appreciated, the gap between human and animal cognition will likely remain stark. AI-assisted interspecies translation technologies could help narrow this gap by enabling limited communication with animals. But unless these technologies take major liberties with interpretation, the difference will persist.

Biological origins

At the same time, non-human animals may have an advantage over AI systems due to their biological nature. We share organic anatomies and evolutionary histories with other animals, making it easier to imagine they are conscious. In particular, these similarities in biology and origin allow us to assume a connection between external features like appearance and behavior and internal states such as consciousness [2].

By contrast, AI systems will both be made of different materials and result from different processes, at least for the foreseeable future. These factors not only understandably prompt skepticism about consciousness [9] but also make the search for consciousness harder [4]. AI systems may be designed to seem conscious, through facial expressions or tone of voice, without possessing mechanisms that plausibly support consciousness. Conversely, they may be designed to seem non-conscious despite possessing such features.

Practices and institutions

Human–animal relationships are deeply embedded in systems spanning food, labor, clothing, and companionship [10]. AI systems, by contrast, are not yet fully integrated into such systems. As a result, attitudes toward AI consciousness are still flexible and open to negotiation. Unlike attitudes towards animals, which are shaped by habit and custom, societal views on AI remain malleable and subject to influence by design, discourse, and policy.

Much will thus depend on how key actors respond moving forward. For instance, AI companies may have competing incentives: to play up AI consciousness as a way of promoting technological excitement, or to downplay it to avoid ethical scrutiny and regulation. Their marketing strategies, together with broader educational and media environments, could give companies substantial power to shape public perception.

The urgent need for more research

It remains difficult to predict how society will respond to the possibility of AI consciousness. Given the importance and difficulty of the topic and the presence of social and psychological forces pulling in opposing directions, societal disagreement and tension appear likely. The animal case suggests that, if acceptance of AI consciousness occurs at all, then it will be slow and partial. Long-term trajectories are even harder to forecast: eventual acceptance, eventual denial, or persistent uncertainty and polarization are all plausible outcomes.

Each of these scenarios carries substantial risks. To navigate them wisely, we urgently need interdisciplinary research regarding the possibility of AI consciousness, public attitudes about AI consciousness, and ethical and political norms for interacting with potentially conscious AI systems. Moreover, this research must guide public discourse and policy development. Across fields, it may be useful to build on past research about non-human animals, while remaining mindful of the many similarities and differences between animals and AI systems in this context.

Regarding the possibility of AI consciousness, scientists and philosophers can continue to investigate the nature of consciousness, with special focus on

Box 2. Social science research priorities

In addition to improving our understanding of whether and when AI systems could be conscious, and how to interact with potentially conscious AI systems, researchers can and should also improve our understanding of how societies will react to this issue. There are at least three general social science research priorities:

What drives attitudes about AI consciousness? To what extent will attitudes be shaped by intrinsic features like faces and voices, by relational features like social and economic incentives, and by contextual factors like corporate marketing? In each case, what exactly will the drivers and mechanisms be?

What is the trajectory of attitudinal change over time? As technology advances, will experts and the public become increasingly willing to accept that AI systems are conscious, increasingly unwilling, or remain in a state of persistent disagreement, agnosticism, and uncertainty? Is the debate becoming more polarized, and if so, how does that polarization affect policy options?

How can experts and policymakers support a positive trajectory? Given the importance of fostering a more informed, rational, and constructive societal discourse about AI consciousness with minimal unnecessary polarization, what can researchers, educators, and policymakers do to promote this? What lessons can we draw from other domains, such as discourse around animal consciousness?

whether particular computational functions may suffice for consciousness. They can also examine AI systems for consciousness-relevant features using an adapted 'marker method' that identifies properties corresponding to consciousness in humans and other animals [5,11] (Box 1).

Regarding public attitudes about AI consciousness, social scientists can use a wide range of methods to shed light on this issue: surveys to gauge public opinion, experiments to explore cognitive biases, qualitative interviews for deeper insight, and longitudinal studies to track how views evolve over time. Foundational research on mind perception can also reveal how people conceptualize consciousness and its dimensions [12], and expert forecasting and historical case studies can provide additional context [13] (Box 2).

Regarding ethical and political norms, philosophers and legal theorists can develop policy frameworks for considering and mitigating risks for humans, animals, and AI systems in a reasonable, proportionate manner. This will require estimating risks of further development and deployment for all stakeholders and navigating tradeoffs thoughtfully.

It may seem that society still has time to prepare. But AI is progressing at a rapid pace, and, as the case of animals shows us, public attitudes about AI could rapidly harden into deeply entrenched divisions shaped by culture, politics, and industry. The time to begin preparing is now.

Declaration of interests

No interests are declared.

[1]University of Oxford, Oxford, UK
[2]New York University, New York, NY, USA
[3]London School of Economics and Political Science, London, UK
[4]Equal contribution.

*Correspondence:
lucius.caviola@philosophy.ox.ac.uk (L. Caviola).

https://doi.org/10.1016/j.tics.2025.06.002

References

1. Sebo, J. (2025) *The Moral Circle: Who Matters, What Matters, and Why*, WW Norton
2. Birch, J. (2024) *The Edge of Sentience: Risk and Precaution in Humans, Other Animals, and AI*, Oxford University Press
3. Chalmers, D.J. (2022) *Reality+: Virtual Worlds and the Problems of Philosophy*, Penguin Books
4. Seth, A. (2025) Conscious artificial intelligence and biological naturalism. *Behav. Brain Sci.*, Published online April 21, 2025. https://doi.org/10.1017/S0140525X25000032
5. Bayne, T. *et al.* (2024) Tests for consciousness in humans and beyond. *Trends Cogn. Sci.* 28, 454–466
6. Long, R. *et al.* (2024) Taking AI welfare seriously. *arXiv*, Published online November 4, 2024. https://doi.org/10.48550/arXiv.2411.00986
7. Caviola, L. *et al.* (2021) Utilitarianism for animals, Kantianism for people? Harming animals and humans for the greater good. *J. Exp. Psychol. Gen.* 150, 1008–1039
8. Bastian, B. *et al.* (2012) Don't mind meat? The denial of mind to animals used for human consumption. *Personal. Soc. Psychol. Bull.* 38, 247–256
9. Ladak, A. and Caviola, L. (2025) Digital sentience skepticism. *PsyArXiv*, Published online January 30, 2025. https://doi.org/10.31234/osf.io/wvbya_v1
10. Bastian, B. and Loughnan, S. (2017) Resolving the meat-paradox: a motivational account of morally troublesome behavior and its maintenance. *Personal. Soc. Psychol. Rev.* 21, 278–299
11. Butlin, P. *et al.* (2023) Consciousness in artificial intelligence: Insights from the science of consciousness. *arXiv*, Published online August 22, 2023. https://doi.org/10.48550/arXiv.2308.08708
12. Weisman, K. *et al.* (2017) Rethinking people's conceptions of mental life. *Proc. Natl. Acad. Sci. U. S. A.* 114, 11374–11379
13. Caviola, L. (2025) The societal response to potentially sentient AI. *arXiv*, Published online March 29, 2025. https://doi.org/10.48550/arXiv.2502.00388
14. Bourget, D. *et al.* (2023) Philosophers on philosophy: the 2020 philpapers survey. *Philosoph. Impr.* 23, 11
15. Dreksler, N. *et al.* (2025) Subjective experience in AI systems: what do AI researchers and the public believe? *arXiv*, Published online June 13, 2025. https://doi.org/10.48550/arXiv.2506.11945