

Intelligence in animals, humans and machines: A Heliocentric view of intelligence?

Halfdan Holm, Soumya Banerjee
University of Cambridge
Cambridge, United Kingdom

We argue that current conceptions of intelligence are anthropocentric. Looking at intelligence in both biological and artificial (human-engineered) systems can yield a more sobering and nuanced view of intelligence.

We will explore why we care about measuring the intelligence of AI systems, what we mean by the word “intelligence” in different contexts, and why AI systems may think very differently from humans.

Intelligence is a broad concept. A commonly used one defined by Legg and Hutter is “Intelligence measures an agent’s ability to achieve goals in a wide range of environments” [1]. Immanuel Kant had suggested that intelligence is the product of all cognitive faculties operating synergistically under a single higher-order unit of consciousness. Varela, Thompson, and Rosch have also suggested that intelligence is embodied: that is, there needs to be a body to have intelligence [10].

Francois Chollet used the No Free Lunch Theorem to argue that there is no such thing as general intelligence for all possible tasks [3]. The No Free Lunch Theorem states that “any two optimization algorithms (including human intelligence) are equivalent when their performance is averaged across every possible problem” [3]. Chollet points out a consequence of this being, that there is no such thing as a true universal intelligence. *We must define what tasks we care about if we want to compare the intelligence of different systems.*

Davis and Marcus point out how poorly GPT-3 performs on common-sense reasoning tasks [2]. Melanie Mitchell points out the more general statement about artificial intelligence (AI): “easy things are hard and hard things are easy” [6].

Let us take the example of image classifiers, which can be prone to adversarial examples. Adversarial examples are defined as samples where the change in input features does not lead to a change in the classification made by a human, but does lead to a change in the classification made by an AI system. This can be as simple as an image of a cow located on a beach (which is very uncommon and may fool an algorithm) rather than on a pasture, but also small changes on a pixel level that humans would not notice.

These AI systems are often criticized as being stupid, statistical, and basing their classification on the wrong parts of the image. In other words, if we want to minimize adversarial examples, we must build a system that basically *thinks like humans*.

Melanie Mitchell has a similar take: “it seems clear from AI system’s non-humanlike errors and vulnerability to adversarial perturbations that these systems are not actually understanding the data they process, at least not in the human sense of “understand” “ [6].

Hence machine intelligence may be a very different kind of intelligence compared to human intelligence.

It is not surprising that image classifiers with no prior knowledge trained on pictures of animals will pick up on the background (pastures) rather than the actual cow, while humans would not. Say a human who had never seen cows and crocodiles before was shown pictures of cows and crocodiles. If they were to use Occam's razor to decide how to classify the photos, they might focus on the animals because they already know what the concept of an animal is. However, a *tabula rasa* (blank slate) image classifier also applying Occam's razor might find the colour of the background to be the simplest feature to focus on.

That is, these two classifiers (human and artificial) have different feature representations. One could also argue that this is not about features and Occam's razor: rather humans understand what kinds of features other humans care about, and which are likely to be available when we are presented to out-of distribution examples. For example, the classic example of the cancer classification tool that based its classification on a ruler being present in positive cases and not in the negative cases. Here, a human would have figured out that it should have ignored that feature.

We can see some connections here to the age-old debate of nature vs. nurture. The differences in the feature representations above come from what world model humans already have. The image classifier has a blank slate, but an adult human already has a model that creates feature representations. Now, can these feature representations be reproduced in image classifiers? There are pretrained image classifiers that have learned useful feature representations; however, the feature representations do not necessarily end up being the same as the ones humans have.

Noam Chomsky argued that human children could not acquire human language in the fashion that they universally do, unless they were born with a "language acquisition device" [4]. In contrast, empiricists, such as John Locke, have taken something close to a "blank slate" or *tabula rasa* view, arguing that our knowledge comes from experience, delivered through the senses. Gary Marcus has been arguing that there are fundamental concepts that humans understand, that perhaps cannot be learnt from scratch [5].

The subject of how much of what we understand is innate, and how much do we learn over our lifetimes (including our childhood) has been debated for a long time. The answer to this question (as suggested by Immanuel Kant) is probably somewhere in between: understanding is both innate and learnt.

The philosopher Maurice Merleau-Ponty said that we cannot consider ourselves as disembodied beings living disconnected from the world. We need to consider ourselves as embodied and that we have embodied cognition. We are not disembodied beings (mind in a vat). Others like Kevin O'Regan [9] have also suggested that sensations like vision and touch give us a "feeling" because we can interact with these objects in an interactive manner. For example, we perceive an object as "soft" because we can squish that object in our hand. According to Edmund Husserl's phenomenological viewpoint, the way the world appears to us is a function of the way we experience it: our senses and the concepts we already have shape the raw data that come into us, and create an experience of an outside world.

Humans often like to compare their intelligence to that of animals. We like to claim that we control the world because we are more intelligent than all other animals. However, which of the above-mentioned definitions of intelligence are we using when we are making these claims? Are we claiming that we are better than all animals at all tasks? This is certainly not true. Consider, for example, how much better dogs are at processing sound and smells than we are. Birds such as cockatoos can also intelligently use multiple tools to retrieve food [7]. At the same time, we cannot claim that other animals are better than us in all tasks.

To make a comparison, we are, of course, forced to define what tasks we care about. *We think we are more intelligent than other animals because we have defined intelligence in terms of the tasks that we care about.*

Finally, we also need to look at intelligence in other biological systems (such as plants which are capable of intelligence) and complex systems. Complex systems such as cities, societies, organizations, and nation states exhibit complex intelligent behaviour. They have also existed for more time than artificial human created machines. Indeed, intelligence can be manifested on a variety of computational substrates: for example, on carbon (biological intelligence) or silicon (artificial intelligence) [11].

Why does this matter? Again, we will have to go back to why we care about measuring intelligence in the first place. Many may be concerned if we build AI systems that are more intelligent than us. However, by many definitions of intelligence, we are not smarter than animals. However, we have still “conquered the world”, which may be very concerning for many animals!

There may not be such a thing as truly universal intelligence, so we may be forced to define intelligence as something more specific. *This hints at the fact that there can be distinct definitions of intelligence, all used for different purposes.*

People are only concerned about AI becoming more intelligent than humans. *They are not talking about the different kinds of intelligence and are only concerned about intelligence defined over a scope of tasks that can affect humans.*

Finally, there are many different notions of intelligence, and any discussion of AI may become confusing when people use different definitions. It may be fruitful to use phrases like “an optimizer” or “a planner” rather than “an artificially intelligent system”; this would reduce the chance that people will compare these systems to human intelligence.

AI is neither artificial nor intelligent. It is a complex product of human intelligence. It may be better to invent other descriptions of AI and deflate the language while discussing AI. These are optimization functions that act on huge amounts of data collected by humans. There is no need to ascribe mysterious powers to AI.

Edsger Dijkstra once said “The question of whether a computer can think is no more interesting than the question of whether a submarine can swim.” The suggestion is that the concepts of “thinking”, and “intelligence” are very anthropocentric and the question of whether computers can think is meaningless. Additionally, intelligence is task-specific, as we have discussed before. Kevin O’Regan in his studies on sensory motor interactions [9] has

suggested that feelings like how we perceive a colour or how we sense the hardness of an object are constructed through interactions with the physical world. In this view, thoughts and feelings are not special objects or concepts: our brain constructs them.

In keeping with this, we lay the following provocative suggestions: 1) we would like to relinquish the claim that present day AI algorithms have any claim to some expanded notion of intelligence that we humans have defined, and 2) we suggest working towards a non-anthropocentric notion of intelligence that is a much more expansive concept of intelligence than humans are capable of.

We call for a more non-anthropocentric conception of intelligence. Indeed, the story of machine intelligence is a very short one compared to other forms of intelligence such as biological intelligence, intelligence in societies, intelligence in political economies, and ethical aesthetics. The challenge is to shift our perspective to a conception of intelligence that is far richer than human intelligence.

Our conception of intelligence also has implications for AI risk. Some people argue that we do not have to worry about AI risk and artificial general intelligence (AGI), because current systems are very bad at some tasks that humans find very easy. However, just because humans are vastly better at certain tasks than an AI system, does not mean that the AI system cannot be dangerous to humans. For example, killer robots are not highly intelligent, but could still pose a threat to society.

Additionally, biological and artificial intelligent systems maybe un-prestatable [8]: that is, their behaviour is an emergent property of all the complex interconnected components these systems have and cannot be stated in advance. This implies that in general it is nearly impossible to determine risk in these complex systems.

We have difficulty trusting what we do not understand. If we cannot understand intelligence, we will always misunderstand AI and misestimate AI risk.

A few centuries ago, we used to think that the entire Universe revolved around Earth. Then Copernicus and Galileo brought about a seismic shift in thinking. A similar seismic shift in thinking is required for intelligence. The world may not revolve around us: our current conception of intelligence may be very anthropocentric. We may need to redefine and rethink intelligence in both biological and artificial systems.

Ultimately humans and machines may have different kinds of intelligence. This has implications for how we view AI and AGI, and for how we manage AI risk.

Author contributions

HH and SB wrote the main text of the manuscript. SB directed the study. All authors reviewed the manuscript.

Conflict of interest

The authors declare that they have no conflicts of interest.

Funding

SB acknowledges funding from the Accelerate Programme for Scientific Discovery Research Fellowship.

Data availability

The manuscript has no associated data.

References

- [1] Marcus Hutter. Universal Artificial Intelligence: Sequential Decisions Based on Algorithmic Probability (Texts in Theoretical Computer Science. An EATCS Series). 2006.
- [2] Ernest Davis and Gary Marcus. Commonsense Reasoning and Commonsense Knowledge in Artificial Intelligence. In: Commun. ACM 58.9, pp. 92–103, 2015. issn: 0001-0782. doi: 10.1145/2701413. url: <https://doi.org/10.1145/2701413>.
- [3] Francois Chollet. On the Measure of Intelligence. In: ArXiv abs/1911.01547, 2019.
- [4] Noam Chomsky. Aspects of the Theory of Syntax, MIT Press, 1965
- [5] Gary F. Marcus. Innateness, AlphaZero, and Artificial Intelligence. In: ArXiv abs/1801.05667, 2018.
- [6] Melanie Mitchell. Why AI is harder than we think. In: Proceedings of the Genetic and Evolutionary Computation Conference, 2021.
- [7] <https://www.smithsonianmag.com/science-nature/like-humans-and-chimps-cockatoos-can-use-a-toolkit-to-get-a-meal-180981775/>, URL accessed Oct. 2023
- [8] Giuseppe Longo, Maël Montévil. Extended criticality, phase spaces and enablement in biology, Chaos, Solitons & Fractals, 55, 10, 2013
- [9] Kevin O'Regan. Why red doesn't sound like a bell: understanding the feel of consciousness, Oxford University Press, 2011
- [10] Francisco J. Varela, Evan Thompson, Eleanor Rosch. The Embodied Mind: Cognitive Science and Human Experience, MIT Press, 1992
- [11] Soumya Banerjee. A Roadmap for a Computational Theory of the Value of Information in Origin of Life Questions, Interdisciplinary Description of Complex Systems, 14(3), 314-321, 2016