The limits of machine intelligence

Henry Shevlin, Karina Vold, Matthew Crosby, Marta Halina

Despite recent breakthroughs in machine learning, current artificial systems lack key features of biological intelligence. Whether the current limitations can be overcome is an open question, but critical to answer, given the implications for society.

Introduction

Many recognize that the concept of intelligence is both nebulous and potentially dangerous. Historically, it has been weaponized in various ways. 20thcentury eugenicists, for example, deployed early psychometric measures of intelligence as a means to oppress socially marginalized groups such as ethnic minorities. Undeterred by these controversies, however, employers, educators, and developmental psychologists continue to use measures of intelligence to assess cognitive potential and track individual progress. To give just one example, more than a million children in the United States are given IQ tests every year.

Despite there being little consensus on what intelligence is or how to measure it, the media and public discourse has become increasingly preoccupied with the concept due to recent accomplishments in machine learning. Governments and corporations are investing billions of dollars to fund researchers who are keen to produce an ever-expanding range of artificial intelligent (AI) systems. Governments representing over thirty countries have announced AI initiatives in the last three years (Dutton 2018). For example, the EU Commission pledged to increase the investment in AI to €1.5 billion by 2020 (from €500 million in 2017), while China has committed $2.1 billion towards an AI technology park in Beijing alone (ibid.). This global investment in AI is astonishing and prompts several questions: What are the true potentials and limitations of AI? What do AI researchers and developers mean by "intelligence" and how does this compare to our everyday notion of intelligence and how this term is used in the cognitive sciences? Can machine learning produce anything truly intelligent?

What is intelligence?

Though it may be hard to pin down, we have an intuitive grasp of what intelligence is. We have long associated it with capabilities such as solving difficult problems, reasoning consistently and reliably, and processing information quickly. Still, we recognise that there are different kinds of intelligence corresponding to varied abilities like mathematical ability, social and emotional reasoning, and imagistic and spatial skills. We should thus be open to the possibility that our intuitive notion of intelligence may not pick out a single neatly defined cognitive capability. With this in mind, it's reasonable to wonder what exactly investors and AI developers are striving towards and how their accomplishments measure up to our biological ones.

In a time when headline-making AI breakthroughs are an almost daily occurrence, it might seem that we are on the cusp of living with artificial systems that match or exceed human intelligence. In 1997 Garry Kasparov, head in hands, lost a chess match to IBM's Deep Blue. Almost exactly twenty years later, champion Ke Jie was defeated at Go by DeepMind's AlphaGo Master. The algorithmic accomplishments required to achieve these feats are breathtaking. And by many definitions they should count as intelligent.

In their now famous proposal, John McCarthy, Marvin Minsky, Nathaniel Rochester and Claude Shannon coined the term "artificial intelligence", defining it as a machine that behaves "in ways that would be called intelligent if a human were so behaving" (McCarthy et al. 1955). Although this is a useful umbrella definition, it fails to capture an important distinction between narrow and general intelligence. Artificial systems, such as Deep Blue and AlphaGo, excel at specific tasks, yet lack the ability to apply their resources outside fairly narrow domains. Such systems have what experts call Artificial Narrow Intelligence (ANI). Many of the headline-making accomplishments in AI today are intelligent in this way. Humans, on the other hand, possess general intelligence, or the ability to deploy the same core suite of cognitive resources on a wide range of different tasks.

We suggest that general intelligence in this sense arguably captures the key features of intelligence as a psychological concept, in particular its association with learning and flexibility. When we assert that a human is more intelligent than an artificial system like AlphaGo, this is not in virtue of the human possessing greater arithmetical abilities or faster processing, but because we are able to apply our information-processing capacities to a vastly broader set of tasks. And while we may think of general intelligence as best exemplified by humans, nature abounds with cases of animals with capacities well in advance of those found in current artificial systems. In confronting the surprisingly complex forms of communication and social learning in bees, the feats of robust long-range navigation in migratory birds, or the astonishing memory and tool use found in corvids, we naturally (and in our opinion, quite rightly) describe them in terms of intelligence. The term in this context serves to pick out not just the complexity of the tasks these creatures perform, but their versatility and adaptability, reflected in their ability to accomplish their goals in varying environments and facing different challenges.

This concept of general intelligence as involving cognitive flexibility aligns well with contemporary definitions of intelligence in computer science. Artificial intelligence researchers Shane Legg and Marcus Hutter, for example, define intelligence as a measure of "an agent's ability to achieve goals in a wide range of environments" (2001). A machine with such a capacity would be an example of

Artificial General Intelligence (AGI) and approach what many view as the *sine qua non* of biological intelligence: flexible, robust, innovative learning, reasoning and behavior.

AGI and contemporary artificial systems

We have suggested that biological intelligence still has a significant edge over AI. However, closing this gap is an explicit goal for many machine-learning researchers today. The Defense Advanced Research Projects Agency (DARPA), for example, recently invested over $2 billion towards the development of what it calls the "third wave" of AI technologies. Unlike narrow AIs that depend on handcrafted rules ("first wave") or domain-specific machine-learning systems trained on large data sets ("second wave"), the next wave of AI aspires to create machines that will "function more as colleagues than as tools" with capacities to "understand and reason in context" (DARPA). In comparison to most existing AIs, artificial systems with a high degree of general intelligence might be expected to perform well under a wider range of contexts and to be more robust outside of specialised training environments. Such systems might also exhibit a greater degree of operational autonomy, insofar as they possess more flexibility in dealing with novel or challenging situations without human input.

But has AGI come close to being achieved? Not yet. Deep Blue and AlphaGo are examples of first-wave and second-wave AI respectively. Although AlphaGo consists of a sophisticated blend of neural networks and Monte Carlo Tree Search, it learns Go by playing millions of games against itself and is unable to apply the knowledge and skills acquired to new domains (Lake et al. 2017). Programs such as these teach us that the activity of playing chess and Go do not require general intelligence after all. Indeed, most current benchmarks in artificial intelligence (including the ImageNet challenge, MNIST, Starcraft and others) measure performance on narrow tasks and are not indicative of general intelligence. While there are some benchmarks that have been traditionally aimed at general intelligence (such as the Turing Test, machine translation challenges, and the Winograd Schema), current systems that do well on these tests typically do so using techniques that are not generalisable to other problems. Thus, these are no longer considered good tasks for testing general intelligence. The quest for AGI has, thus far, not been achieved. Indeed, it appears that the field lacks adequate tests for evaluating progress in this direction.

Measuring progress towards AGI

Even if AGI remains currently out of reach, are we at least making progress towards it? Answering this question requires an account of how one should measure the general intelligence of machines. Whatever the difficulties involved in comparing intelligence across individual humans, they are dwarfed by the much greater challenges involved in assessing intelligence across non-human systems. Most neurotypical humans possess a broadly similar set of cognitive capacities such as episodic memory, working memory, and theory of mind, as well as similar sensory inputs and motor abilities, thus making it possible to develop suites of cognitive tasks that enable meaningful comparison of performance across a wide range of individuals.

Moving slightly away from the familiar human case, non-human animals vary dramatically in their cognitive and sensorimotor capabilities, making it extremely challenging to develop informative sets of tasks to compare abilities. Tests involving visual cues, for example, such as the standard mirror self-recognition test will be applicable only to individuals with reasonable eyesight, while tests of causal understanding that rely on spontaneous tool use face the difficulty that diverse non-human animals have quite different physical abilities to manipulate objects. While a primate, elephant or octopus can use their prehensile limbs or trunk to grasp an external object, creatures such as cetaceans, fish, or birds must manipulate objects via their mouths, thus in many cases requiring different task schema. To make matters more difficult, non-human animals differ in capacities such as inhibitory control that modulate performance across tasks. This makes it hard to know in many cases whether a system's failure on a task is due to lack of competence in the specific domain being tested or more general cognitive limitations such as lack of attentional control (Beran & Hopkins, 2018). Nevertheless, many tasks have been designed that can be translated across species based on the commonalities of visual processing, navigation, and motivation towards food sources.

Measurement of general intelligence in artificial systems presents a more daunting challenge even than in the case of non-human animals. Whereas animals share some similarities, artificial systems exist without any of these properties. Our newly launched competition, the Animal-AI Olympics, is attempting to find a common ground by providing a simulated environment, with realistic physics, where artificial agents receive visual inputs and can move around with the goal of retrieving food (Crosby et al. 2019). The competition includes tests of capabilities important for our biological understanding of general intelligence that are direct translations of those used in comparative psychology. A key element of the competition is that, like many animal cognition studies, the agents have no prior experience of any of the tests, just of the environment. The current capabilities of AI systems, especially in unseen situations, means that only the simplest tests from the animal cognition literature are included in the first iteration of the competition. For example, whilst there has been promising work in online and social learning, it is not yet at the stage where it can be tested in such a general setting. Solving the initial tasks will be an important first step toward systems with general intelligence in the way that we understand it in biological systems, but even then, there will still be a long way to go.

Learning is a further component of general intelligence that must be accounted for. An important idea is to attempt to group different systems into classes based on their ability to perform different kinds of learning (Dennett, 1996). For example, a wide range of biological organisms seem to be capable of associative learning. A relatively smaller set of organisms are able to learn from observations of conspecifics, and still fewer are able to use causal models to determine how to complete tasks (Jelbert et al. 2014). By classifying systems according to their possession of these broad capabilities, it may be possible to develop multi-dimensional 'intelligence profiles' of different cognitive agents and apply them to artificial systems.

A further valuable strategy for assessing intelligence in different systems may be to appeal to more abstract cognitive dynamics, such as the ability to transfer information from one domain to another, to retain information over extended periods, and to correct errors in performance. This approach is likely to be particularly useful in developing assessments of intelligence in artificial systems that differ considerably from biological systems in their more fine-grained capabilities. Many artificial systems are not situated sensorimotor agents, and hence it is not possible to examine, for example, whether they could arrive at strategies for copying the motor behaviours of other agents. However, we can still quantify the ability of such systems to transfer information from one task to another or to retain information over time without suffering from catastrophic forgetting, and compare this with equivalent capacities in biological organisms.

Current AI systems do not come close to existing biological entities on any of these metrics. We can, however, identify areas where progress is being made. For example, neural networks form the basis of many of the recent successes in AI, but, until very recently, have suffered from the problem that switching to learning a new (potentially very similar) task can cause catastrophic forgetting of solutions to the previous task. This can be overcome by locking in important parameters for solving certain tasks, making it possible to solve multiple tasks in sequence, leading to more generally applicable, less narrow, systems (Kirkpatrick et al. 2017). In few-shot (supervised) learning, a standard task involves a training set of categorised images, where the task is to learn a mapping from images to categories. At test time, a new category is introduced with only a few examples and the task becomes to label images even from the new set. The key problem here is how to generalise from the (sometimes single) new example(s). Solutions include, for example, meta-Learning methods, which attempt to learn how to learn from only a few examples (Finn et al 2017).

While each of these research areas may be taking us closer to AGI, it is still the case, as we have emphasised, that current AI falls far short of the kind of general intelligence we find in the biological world. It is easier to take a specific instance of a problem (such as a few-shot learning dataset), and focus on improving performance on that, instead of trying to build systems with truly diverse skill sets. Hence even the steps towards general intelligence capabilities just mentioned are narrow in the sense that they are not generally applicable without further work and they do not necessarily require flexible and innovative learning, reasoning or behavior. It will not be until many such advances have been made and can be combined into a single system that we will be approaching AGI. But will this require radical new approaches in AI or will it be possible with innovations to current methods?

It has been argued that "neither deep learning, nor other forms of second-wave AI, nor any proposals yet advanced for third-wave, will lead to genuine intelligence" (Brian Cantwell Smith Forthcoming). But perhaps AI is on the right track, and all the challenges between now and AGI are surmountable with the right innovations. In either case, we see no reason that AGI is in principle impossible. Biological general intelligence contains many examples of complex systems that are generally intelligent, and there is no principled reason to assume that such complexity is off-limits to AI.

The great Enlightenment philosopher David Hume claimed that all operations of the mind involved associations of ideas, a view that in various forms still has adherents among contemporary philosophers. Likewise the great comparative psychologist Edward Thorndike suggested that associative learning underpins all animal behaviour. These questions are still heavily debated in contemporary cognitive science. But if some broadly associationist picture of the mind turns out to be true, then incremental progress on our current techniques of reinforcement learning (AI's version of associative learning) might be able to get us most of the way towards general intelligence without requiring a fundamental paradigm shift. However, as the ongoing debate shows, we do not yet know how general intelligence is achieved in animals and, even if we did, the parallels to AI methods are not perfect. Simply put, it's still too early to tell if AI requires radical new approaches to reach generality. But the most fruitful way forward, in our view, is for computer and cognitive scientists to work together.

Thus far we have focused on the possibility of developing autonomous AGI. However, we would be remiss not to draw attention to a further possibility--that humans themselves may be part of our first ‘artificial’ general intelligences. The assumption of some degree of autonomous agency is often built into the definition of AI. And in the media and public discourse AI systems are mostly portrayed as autonomous and entirely distinct from their human counterparts. But there is also a great deal of machine learning used in specialised non-autonomous systems designed to support and enhance human cognitive capacities. Hence, although autonomous AGI might not be achievable, we may be able to achieve some of the associated results by augmenting our own capacities. While humans already enjoy general intelligence, our biological make-up entails many limitations, from low memory resources to the many built-in cognitive biases that psychologists have identified. Our cognitive history shows how even very simple technologies, like a pen and paper, have transformed our capacities, by ‘extending’ our memory and enabling us to perform complicated arithmetic. Given our past, one can only expect that machine learning techniques will push our cognitive boundaries even further--conferring sophisticated new mind modelling techniques, improving our conceptualization, learning, and abstraction skills, and ultimately improving our own abilities to flexibly achieve our goals.

Conclusion

Our goal in this paper has been to suggest what general intelligence means and how we might measure progress towards it. One of our key claims is that even the most ingenious artificial systems still fall dramatically short of the wide-ranging general intelligence found in many animals. However, we believe that the next decade is likely to prove crucial in settling the question of whether major new conceptual leaps are required for AGI, as researchers probe and push the limits of existing paradigms in machine learning. If flexible, robust, and versatile forms of behaviour like those found in animals turn out to be possible via tweaks of existing models and the application of more compute, the gap between biological and artificial minds may narrow before our eyes. If, by contrast, our artificial systems continue to fail to match up to biological organisms in these respects, we may have reason to think that nature is still concealing some of her best tricks from us. In such a case, the hunt for bold new paradigms drawn from neuroscience will become critical, as will the use of hybrid human-AI systems. Either way, the fate of ongoing machine learning research will surely bear on longstanding debates in cognitive science concerning the structure and function of minds and perhaps the future of intelligence itself.

References

1. Dutton, T. (2018). An overview of national AI strategies. *Medium, June*, *28*.
2. McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. AI magazine, 27(4), 12-12.
3. Legg, S., & Hutter, M. (2007). Universal intelligence: A definition of machine intelligence. *Minds and machines*, *17*(4), 391-444.
4. Beran, M. J., & Hopkins, W. D. (2018). Self-control in chimpanzees relates to general intelligence. *Current Biology*, *28*(4), 574-579.
5. Crosby, M., Beyret, B., & Halina, M. (2019). The Animal-AI Olympics. *Nature Machine Intelligence*, *1*(5), 257.

6. Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and brain sciences*, *40*.
7. Dennett, D. C. (1996). *Kinds of Minds*. Basic Books.
8. Serrà, J., Surís, D., Miron, M., & Karatzoglou, A. (2018). Overcoming catastrophic forgetting with hard attention to the task. *arXiv preprint arXiv:1801.01423*.
9. Defense Advanced Research Projects Agency. "AI Next Campaign." Available online: <https://www.darpa.mil/work-with-us/ai-next-campaign> Accessed on 5 August 2019.
10. Brian Cantwell Smith. (Forthcoming) The Promise of Artificial Intelligence: Reckoning and Judgment, *MIT Press.*

Recommended Reading List:

On comparisons between artificial and biological forms of intelligence:

1. Buckner, C. (2019, May 20). The Comparative Psychology of Artificial Intelligences [Preprint]. Retrieved August 5, 2019, from http://philsci-archive.pitt.edu/16034/
2. Hernandez-Orallo, J. (2017). The Measure of All Minds: Evaluating natural and artificial intelligence. Cambridge University Press.
3. Godfrey-Smith, P. (2017). *Other Minds: The Octopus and the Evolution of Intelligent Life*. HarperCollins UK.
4. Shevlin, H., & Halina, M. (2019). Apply rich psychological terms in AI with care. *Nature Machine Intelligence*, *1*(4), 165–167.

On Animal Cognition:

5. MacLean, E. L., Hare, B., Nunn, C. L., Addessi, E., Amici, F., Anderson, R. C., & Boogert, N. J. (2014). The evolution of self-control. *Proceedings of the National Academy of Sciences*, *111*(20), E2140-E2148.
6. Jelbert, S. A., Taylor, A. H., Cheke, L. G., Clayton, N. S., & Gray, R. D. (2014). Using the Aesop's fable paradigm to investigate causal understanding of water displacement by New Caledonian crows. *PloS one*, *9*(3), e92895.

On AI methods:

7. The Animal-AI Olympics: http://animalaiolympics.com/

8. Finn, C., Yu, T., Zhang, T., Abbeel, P., & Levine, S. (2017). One-shot visual imitation learning via meta-learning. *arXiv preprint arXiv:1709.04905*.

9. Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., ... & Hassabis, D. (2017). Overcoming catastrophic forgetting in neural networks. Proceedings of the national academy of sciences, 114(13), 3521-3526.

10. Yao, Y., & Doretto, G. (2010, June). Boosting for transfer learning with multiple sources. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 1855-1862). IEEE.

On extended intelligence:

11. Clark, A. (2008). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension.* New York: Oxford University Press.

12. Hernandez-Orallo, J. & Vold, K. (2019). AI Extenders: The Ethical and Societal Implications of Humans Cognitively Extended by AI. AAAI /ACM Conference on Artificial Intelligence,. Ethics, and Society (AIES 2018), Honolulu, Hawaii, USA. January 27-28, 2019.

On the possibility of Artificial General Intelligence (AGI) and the risk of superintelligence:

13. Bostrom, N. (2017). *Superintelligence*. Dunod.
14. Shanahan, M. (2015). The technological singularity. MIT Press.