The limits of machine intelligence

Despite progress in machine intelligence, artificial general intelligence is still a major challenge

Henry Shevlin[1], Karina Vold[1], Matthew Crosby[2] & Marta Halina[1]

The concept of intelligence is both nebulous and potentially dangerous: historically, it has been weaponised in various ways. Twentieth-century eugenicists, for example, deployed early psychometric measures of intelligence as a means to oppress socially marginalised groups or 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.

....

> "*We should be open to the possibility that our intuitive notion of intelligence may not pick out a single neatly defined cognitive capability.*"

....

Despite there being little consensus on what intelligence is or how to measure it, the media and the public have become increasingly preoccupied with the concept owing to recent accomplishments in machine learning and research on artificial intelligence (AI). Governments and corporations are investing billions of dollars to fund researchers who are keen to produce an ever-expanding range of artificial intelligent systems. More than 30 countries have announced such research initiatives over the past 3 years [1]. For example, the EU Commission pledged to increase the investment in AI research 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 [1]. This global investment in AI is astonishing and prompts several questions: What are the true possibilities and limitations of AI? What do AI researchers

and developers mean by "intelligence"? How does this compare to the everyday concept of intelligence and how the term is other branches of cognitive science? And can machine learning produce anything that is truly "intelligent"?

What is intelligence?

Though it may be hard to come up with an exact definition, we do 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 such mathematical aptitude, 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 is reasonable to wonder what exactly investors and AI developers are striving towards and how the accomplishments of their creations measure up to our biological ones.

At 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 20 years later, Go champion Ke Jie was defeated by the AI company DeepMind's AlphaGo Master. AI is also being deployed to solve all sorts of pressing practical challenges. DeepMind uses AI to control their data centre cooling systems, YouTube and Facebook are using AI in advertising to optimise their recommender systems, and IBM Watson is

replacing humans in drug discovery. The algorithmic accomplishments required to achieve these feats are breathtaking. And by many definitions, they should count as intelligent. Indeed, the ultimate goal for many is to be able to create AI systems capable of solving all these problems at once. Artificial general intelligence (AGI) is AI that is capable of solving almost all tasks that humans can solve, and it would fundamentally change our society. To understand our current progress towards AGI, we must, however, first define artificial intelligence and general intelligence more clearly.

....

> "*At 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 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" [2]. 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 today are intelligent in this way. Humans, on the other

"Freedom", © 2016 Matt Dixon.

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 captures the key features of intelligence as a psychological concept, in particular learning and flexibility. When we assert that a human is more intelligent than an artificial system like AlphaGo, this is not by virtue of the human possessing greater arithmetical abilities or faster processing, but because humans are able to apply their 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 examples of intelligence 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 long-range navigation in migratory birds, or the astonishing memory and tool use found in corvid birds, we naturally—and, in our opinion, quite rightly—describe them in terms of intelligence. In this context, the term 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" [3]. A machine with such a capacity would be an example of AGI and approach what many view as the *sine qua non* of biological intelligence: flexible, robust, innovative learning, reasoning and behaviour.

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. In principle at least, AI systems with the same robustness and behavioural flexibility as animals have great commercial and scientific potential, whether in the form of reliable autonomous vehicles capable of sustained operation without human input or household robots that can safely navigate and interact with complex and varied domestic environments. The US Defense Advanced Research Projects Agency (DARPA), for example, recently invested more than US$2 billion towards the development of what it calls the "third wave" of AI technologies. Unlike narrow AIs that depend on hand-crafted 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" [4] (https://www.darpa.mil/work-with-us/ai-next-campaign). In comparison with most existing AIs, such 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. They 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.

> **"** *...the next wave of AI aspires to create machines that will "function more as colleagues than as tools"...***"**

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 its acquired knowledge and skills to new domains [5]. Such programs 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. There are some benchmarks that have been traditionally aimed at general intelligence—such as the Turing Test, machine translation challenges, and the Winograd Schema—but 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 is still searching for adequate tests to better evaluate progress.

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 challenge of 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 a meaningful comparison of performance across a wide range of individuals.

Moving slightly away from humans, other animals vary dramatically in their cognitive and sensorimotor capabilities, which makes it extremely challenging to develop informative sets of tasks to compare their abilities. Tests involving visual cues, such as the standard mirror self-recognition test, will be applicable only to species with reasonable eyesight, while tests of causal understanding that rely on spontaneous tool use face the difficulty that diverse 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 whether a system's failure on a task is owing to a lack of competence in the specific domain being tested or to more general cognitive limitations such as lack of attentional control. 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 an even more daunting challenge than for non-human animals. Whereas animals share some similarities through their common evolutionary heritage, artificial systems exist without any of these properties. Our newly launched competition, the Animal-AI Olympics, attempts to find common ground by testing artificial agents on tasks drawn directly from animal cognition research [6]. The competition tests the general problem-solving abilities of artificial agents in simulated environments with realistic physics. The agents are tasked with obtaining a positive reward where successfully doing so requires capacities such as overcoming physical obstacles, avoiding negative stimuli, planning, object permanence, functional generalisation, or causal reasoning. A key element of the competition is that, like many animal cognition studies, the agents are given no prior experience with the specific tasks on which they will be tested, but instead must transfer what they have learned from exploring a general physical environment (or "playground"). Demonstrating the ability to solve tasks under such conditions is an important first step towards developing systems with biological-like general intelligence, but even then, there is still a long way to go.

> **"** *Whatever the difficulties involved in comparing intelligence across individual humans, they are dwarfed by the much greater challenge of assessing intelligence across non-human systems.***"**

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 [7]. For example, a wide range of biological organisms seem to be capable of associative learning. A relatively smaller set of organisms including bees, lizards and dogs are able to learn from observations of conspecifics, and a much smaller set—perhaps limited to humans and birds—are able to use causal reasoning to determine how to complete tasks [8]. By classifying systems according to these broad capabilities, it may be possible to develop multidimensional "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. Many artificial systems are not sensorimotor agents, and hence it is not possible to examine, for example, whether they could arrive at strategies for copying the motor behaviours of others. 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.

The way ahead

Current AI systems do not come close to 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 task—even a very similar one—can cause catastrophic forgetting of solutions for 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 and less narrow systems [see recommended reading]. Another area where progress is being made is in supervised learning using only a small number of examples ('few shot' learning). In the case of visual classifiers, for example, a system might have to learn to accurately assign novel images to categories on the basis of just a few prior examples. This is a daunting challenge, but techniques such as meta-learning or 'learning to learn' hold considerable promise [see recommended reading].

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 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 behaviour. 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" [9]. But perhaps AI is on the right track, and all the challenges 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 artificial systems.

....

> " ...*even the most ingenious artificial systems still fall dramatically short of the wide-ranging general intelligence found in many animals*"

....

The 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 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. It may be that using bigger networks with more compute is all we need. 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 is 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 continue to work closely together.

Enhancing human intelligence

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 media and public discourse mostly portray AI systems 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. Contemporary digital personal assistants, such as Siri and Alexa, fuelled by speech recognition and natural language processing algorithms, already support our planning and decision-making, while the routing algorithm on apps like Waze help users navigate safely and quickly through dense urban traffic. But we can also imagine future systems that go far beyond this, for example, personalised systems that improve our attentional capacities by modelling our interests and goals in order to focus our attention on things we might otherwise overlook, or systems that model the interests and goals of our friends and help us anticipate their decisions, actions and emotions, thereby enhancing our emotional intelligence and mind-modelling capacities [10]. Hence, although autonomous AGI might not be achievable in the near future, we may be able to achieve some results by augmenting our own capacities. While humans already enjoy general intelligence, our biological makeup 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 conceptualisation, 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

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   Hernández-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: 165–167

On Animal Cognition:

5   MacLean EL, Hare B, Nunn CL, Addessi E, Amici F, Anderson RC & Boogert NJ (2014) The evolution of self-control. *Proc Natl Acad Sci USA* 111: E2140–E2148
6   Beran MJ & Hopkins WD (2018) Self-control in chimpanzees relates to general intelligence. *Curr Biol* 28: 574–579

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: 1709.04905*
9   Kirkpatrick J, Pascanu R, Rabinowitz N, Veness J, Desjardins G, Rusu AA & Hassabis D (2017) Overcoming catastrophic forgetting in neural networks. Proc Natl Acad Sci USA 114, 3521–3526
10 Serrà J, Surís D, Miron M & Karatzoglou A (2018) Overcoming catastrophic forgetting with hard attention to the task. *arXiv: 1801.01423*
11 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 the possibility of Artificial General Intelligence (AGI) and the risk of superintelligence:

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

On the abuse of intelligence in psychometrics and artificial intelligence:

14 Cave, S: "Intelligence: a history" https://aeon.co/essays/on-the-dark-history-of-intelligence-as-domination
15 Gould, SJ: The Mismeasure of Man. New York, NY: Norton

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 tweaking existing models and the application of more computing power, the gap between biological and artificial minds may narrow. 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 ML, Rochester N, Shannon CE (2006) A proposal for the dartmouth summer research project on artificial intelligence, August 31, 1955. AI magazine, 27(4): 12. https://doi.org/doi.org/10.1609/aimag.v27i4.1904
3. Legg S, Hutter M (2007) Universal intelligence: a definition of machine intelligence. *Mind Mach* 17(4): 391–444
4. 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
5. Lake BM, Ullman TD, Tenenbaum JB, Gershman SJ (2017) Building machines that learn and think like people. *Behav Brain Sci* 40: e253
6. Crosby M, Beyret B, Halina M (2019) The animal-AI olympics. *Nat Mach Intell* 1: 257
7. Dennett DC (1996) *Kinds of minds*. New York, NY: Basic Books
8. Jelbert SA, Taylor AH, Cheke LG, Clayton NS, Gray RD (2014) Using the Aesop's fable paradigm to investigate causal understanding of water displacement by New Caledonian crows. *PLoS ONE* 9: e92895
9. Smith BC (Forthcoming) *The promise of artificial intelligence: reckoning and judgment*. Cambridge, MA: MIT Press
10. Hernández-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