Can Artificial Intelligence Do Everything That We Can?

Vincent Conitzer

A shorter version of this article appeared in The Wall Street Journal *under the title* Natural Intelligence Still Has Its Advantages*, on August 28, 2018 [2].*

The late Stephen Hawking has warned that AI could eventually "spell the end of the human race." Elon Musk has predicted that "robots will be able to do everything better than us." Meanwhile, AI systems are starting to outperform people in domains ranging from board games to speech recognition. Is humanity on the way out?

For those not working in AI, it can be difficult to interpret highly visible achievements in the field. Take, for example, Watson's 2011 victory over human Jeopardy champions Brad Rutter and Ken Jennings. This was a stunning achievement: while it should surprise nobody that Watson had access to an encyclopedic amount of knowledge, Jeopardy is a game that requires more than that. The hard part – at least for AI systems, but often also for humans – isn't having access to the relevant information, but rather understanding the clue well enough to link it to that. Even many AI researchers, myself included, thought this would remain beyond the capabilities of AI systems for a while to come. We were wrong.

But does this mean that Watson had obtained a human-level understanding of the world? No. Watson also produced some cringeworthy responses, for example "What is Toronto?" for a clue in the "US cities" category. This is part of a broader pattern of AI systems achieving superhuman levels of performance, and yet making blunders that leave us scratching our heads. For example, researchers from Carnegie Mellon were able to consistently fool a face recognition system that one of them, clearly a man, was actress Milla Jovovich, by wearing carefully designed eyeglass frames [3].

In both cases, what causes the mistake is that the AI system solves the problem in a way that is very different from how humans do it. Often, this involves picking up on some statistical pattern that can be used to surprisingly great effect, but that sometimes produces answers that lack any common sense. Moreover, if something changes about how the data is produced, performance may plummet. This is especially so when the change is intended to mislead the system, as in the case of adding the eyeglass frames.

This gives some insight into which jobs, or parts of jobs, the AI systems of today and tomorrow are likely to take over from us. Tasks that require

responding to the same kind of standardized input over and over again, with a clear measure of success, are a natural fit. Such tasks range from the diagnosis of medical images to flipping burgers. On the other hand, jobs that are messy and unpredictable and require understanding of people and the broader world – I like to think of kindergarten teachers – will likely remain safe for a long time. Driving a car in a busy city is probably somewhere in the middle.

Much progress has been made in AI in a short span of time, so it is not unthinkable that there will be further breakthroughs, especially if we think in terms of decades or a century. For now, humans remain unsurpassed in their broad, integrated, flexible, and robust understanding of the world. If AI starts to catch up with us on that, it will likely change our world beyond recognition, and some of most intractable problems in philosophy, such as the nature of consciousness, will become very pertinent. I personally am interested in questions about the nature of consciousness, and have even done some research on what I consider to be its more mysterious aspects [1]. (How could a lump of matter possibly give rise to *this* experience?) But there is currently no clear path towards building broadly intelligent systems. The AI systems we know how to build today are likely to be disruptive in many domains – the labor market, our social fabric, the nature of warfare. But they do not make humanity obsolete.

References

[1] Vincent Conitzer. A Puzzle about Further Facts. *Erkenntnis*, 2018. https://doi.org/10.1007/s10670-018-9979-6.

[2] Vincent Conitzer. Natural Intelligence Still Has Its Advantages. *The Wall Street Journal*, August 2018.

[3] Mahmood Sharif, Sruti Bhagavatula, Lujo Bauer, and Michael K. Reiter. Accessorize to a crime: Real and stealthy attacks on state-of-the-art face recognition. In *ACM Conference on Computer and Communications Security*, pages 1528–1540, 2016.