proceedings

MDPI

Abstract

Singularities and Cognitive Computing †

Devdatt Dubhashi [1],

[1] Dept. of Computer Science and Engineering, Chalmers University, Sweden 1; dubhashi@chalmers.se
* Correspondence: dubhashi@chalmers.se; Tel.: +46-(0)31-772-1046
† Presented at the IS4SI 2017 Summit DIGITALISATION FOR A SUSTAINABLE SOCIETY, Gothenburg, Sweden, 12-16 June 2017.

Published: date: 9 June 2017

Abstract: Advances in Artificial Intelligence (AI) / Machine Learning (ML) have led to discussions about singularities: the economic singularity when AI displaces human jobs and the technological singularity when AI surpasses human capabilities in general intelligence. We seek to clarify some issues in the discussion.

Keywords: Artificial Intelligence, Machine Learning, Singularities

1. A Tale of Two Singularities

The *Technological Singularity* refers to the argument first enunciated by the British mathematician I.J. Good, and later popularized by techno-visionaries like Vernor Vinge and Ray Kurzweil, that artificial intelligence (AI) systems get better exponentially and so they would eventually reach human capabilities, at which point there is an ``intelligence explosion'' leading to *superintelligence*. If this is a possibility, then there is a risk, the argument goes, that such a superintelligence could become an existential threat to humanity [1,2]. The *Economic Singularity*, on the other hand, refers to the possibility of AI systems reaching or exceeding human level performance levels at *specific* tasks which could lead to the automating away of a large number of jobs, perhaps eventually eliminating all jobs [3,4].

2. Technological Singularity

In a recent viewpoint piece [5], we argued that the Economic Singularity is a real and imminent danger of AI to society whereas the Technological Singularity is at this point in time, only a very distant logical possibility. Häggström took issue with us on his blog[1] and asks us to consider two hypotheses:

- (H1) Achieving superintelligence is hard - not attainable (other than possibly by extreme luck) by human technological progress by the year 2100,
- (H2) Achieving superintelligence is relatively easy - within reach of human technological progress, if allowed to continue unhampered, by the year 2100.

He claims that "it is not a priori obvious which of hypotheses (H1) and (H2) is more plausible than the other, and as far as burden of proof is concerned, I think the reasonable thing is to treat them symmetrically". It is true that (H1) and (H2) are symmetrical, but not in the way Häggström suggests, namely, that one can assign a prior belief of 50% to both! They are symmetrical in the sense that both are so vague that one cannot assign any meaningful probabilities to them. The risk of super intelligent artificial agents cannot be quantified precisely because the phenomenon is not clearly specified or defined and hence it is not possible to argue about this risk in an evidentially grounded way. There have been some surveys quoted by both Bostrom and Häggström where people have been asked

[1] http://haggstrom.blogspot.se/2017/02/vulgopopperianism.html

their opinion on (H1) versus (H2). What people answer in such surveys is not the result of a careful weighing of evidence because there is simply no meaningful evidence to weigh! Rather, these answers are *gutestimates* i.e. gut feelings and nothing more. To the extent that one takes such ``gutestimates'' seriously, not all of them carry equal weight -- surely a researcher who works actively in AI should be given more weight than an armchair philosopher. Unfortunately, the surveys quoted by Bostrom and do not distinguish between the two kinds of respondents.

So, while it is impossible to attach any meaningful probabilities to (H1) or (H2), one could ask for evidence for (H1) or (H2). The situation here is not symmetrical. Evidence for (H2) would be scientific and technological advances and the current state-of-the-art. Evidence for (H1) could be arguments about fundamental logical or technological limits. Argument of the latter sort have been made on the basis of theoretical results on computational complexity. We do not find these persuasive - computational complexity is an intellectually rich area but it is very far from being directly relevant to practice. Indeed, by similar considerations, many of the recent achievements such as the thousands of miles clocked up by self-driving cars should also have been impossible, since closely related problems are provably intractable in that theory! While we do not know of any convincing fundamental limits for (H1), we [5] did argue that the current state of AI science and technology is very far removed from that required for (H2).

We should clarify that we do believe there are dangers of AI safety to worry about e.g. in the context of self-driving cars or other autonomous systems currently under development. Recent position papers ground concerns about safety in real machine-learning research, and have initiated discussions of practical ways for engineering AI systems that operate safely and reliably. We believe this is a much more fruitful approach to AI safety than worrying about ``superintelligence'' -- here we may draw historical lessons from Francois Jacob (*The Possible and the Actual*, 1982): ``The beginning of modern science can be dated from the time when general questions were replaced by more modest questions ... While asking very general questions lead to very limited answers, asking limited questions turned out to provide more and more general answers.''

3. Economic Singularity

Turning now to the Economic Singularity, three of the most common arguments raised against it are: Technology has always displaced workers from traditional jobs into other sectors, there are some tasks that humans can do that computers could never do, there will be new jobs created by new AI technologies. Let us consider each in turn.

An early example of warnings about the effect of technology on jobs is the Luddite movement of the early 19th century, in which a group of English textile artisans protested the automation of textile production by seeking to destroy some of the machines. In his widely discussed Depression--era essay ``Economic Possibilities for our Grandchildren'' (1930), John Maynard Keynes foresaw that in a century's time, ``we are being afflicted with a new disease … *technological unemployment*.'' Keynes was sanguine about the long run, opining that ``this is only a temporary phase of maladjustment,'' A much more urgent warning came from Norbert Wiener in his classic *Human Use of Human Beings* (1954): ``It is perfectly clear that [automation] will produce an unemployment situation, in comparison with which…the depression of the [nineteen] thirties will seem a pleasant joke. This depression will ruin many industries -- possibly even the industries which have taken advantage of the new potentialities''. These early warnings were perhaps a bit ahead of their time, but very prescient today: in their report *The Future of Employment*, Frey and Osborne from Oxford Martin School state that ``According to our estimate, 47 percent of total US employment is in the high risk category.''

Polanyi's paradox is named after Michael Polanyi, the economist, philosopher, chemist and younger brother of the more famous Karl Polanyi who observed in 1966, ``We know more than we can tell''. Polanyi emphasised that we as humans employ a lot of *tacit knowledge,* the kind of knowledge that is

difficult to transfer to another person by means of writing it down or verbalizing it. This is used as an argument against AI -- there are some tasks humans can do which simply cannot be coded into AI systems. Polanyi's paradox argues against the possibility of AI systems that work on the basis of explicit hard coded rules, such as the so--called *expert systems* of the 80s and 90s. However, the argument loses its force almost entirely in the wake of today's generation of AI systems that are based on machine learning technologies which learn automatically from data without the need for explicit hard coded rules. Autor [6] discusses this, but somewhat confusingly refers to it as ``an atheoretical brute force technique.''. In fact, machine learning is also based on theoretical principles - those that arise out of statistical learning theory - and combines these with advances in algorithmic techniques, computing hardware and engineering of highly scalable systems. In certain narrowly defined tasks such as image recognition, these techniques have now caught up or surpassed human performance [7]. While these advances used *supervised learning* based on training data as noted by Autor, another area of machine learning has shown even more impressive capabilities -- the ability to learn by interaction with the environment via *reinforcement learning*. This was vividly brought into the public consciousness via two *tour-de-force* demonstrations by Google DeepMind. In the first demonstration [8], the computer was able to learn to play a number of Atari games simply by receiving an image of the board and the reward points for each move. In the second demonstration, the computer learnt to play the game of Go [9]. It learnt partly by using an archive of past games between humans and partly by playing millions of games against itself and reached a level of performance that handily defeated one of the world's top human players. *No rules for playing the games were hard coded in the system -- it discovered the strategies on its own*. These results vindicate the conclusions of Brynjolfsson and McAfee and Ford that AI/ML is a *general purpose technology* that will permeate all domains. While routine and repetitive tasks are the easiest targets, even tasks that seemingly require more complex cognitive skills (such as playing Atari games or Go) can be automated. Hence no task will be immune to automation.

Autor [6] has studied the effect of technology on employment. On the one hand, he argues that while automation does indeed substitute for labour, it also complements labor, raises output in ways that lead to higher demand for labour. *Cognitive computing* is a vision which harnesses AI to create digital assistants that can work in synergy with humans to solve complex tasks that would be beyond the abilities of either alone. An example is a doctor working together with a digital assistant to make a complex diagnosis in a cancer – he can leverage a vast store of data and information from past cases and make a more evidence based judgement. As Bundy [10] observes, "the productivity of humans will be, thereby, dramatically increased."

References

1. N. Bostrom, *Superintelligence*: *Paths, Dangers, Strategies*, Oxford University Press, UK 2014.
2. O. Häggström, *Here Be Dragons*, Oxford University Press, UK 2016.
3. E. Brynjolfsson and A. McAfee, *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*, W.W. Norton and Co. 2016.
4. M. Ford, *Rise of the Robots: Technology and the Threat of a Jobless Future*, Basic Books 2015.
5. D. Dubhashi and S. Lappin, ``AI Dangers: Imagined and Real'', *Comm. ACM* ,60:2, 2017, pp. 43-45.
6. D. Autor, ``Why Are There Still So Many Jobs? The History and Future of Workplace Automation'', *Journal of Economic Perspectives*, 29:3 2015, pp. 3-30.
7. Y. LeCun, Y. Bengio and G. Hinton, ``Deep Learning'', *Nature*, 521, 2015, pp. 436—444.
8. V. Mnih et al, ``Human level control through deep reinforcement learning'', *Nature*, 518, 2015, pp. 529-533.
9. D. Silver et al, ``Mastering the game of Go with deep neural networks and tree search'', *Nature*, 529,2016, pp. 484—489.
10. A. Bundy, ``Smart Machines are not a threat to Humanity'', *Comm. ACM*, 60:2, 2017 pp. 40-42.