



This is a repository copy of *The AI singularity and runaway human intelligence*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/107509/>

Version: Accepted Version

Proceedings Paper:

Prescott, T.J. (2013) The AI singularity and runaway human intelligence. In: Biomimetic and Biohybrid Systems. Second International Conference, Living Machines 2013, July 29 – August 2, London, UK. Lecture Notes in Computer Science, 8064 L . Springer Berlin Heidelberg , pp. 438-440. ISBN 9783642398018

<https://doi.org/10.1007/978-3-642-39802-5-59>

The final publication is available at Springer via
http://dx.doi.org/10.1007/978-3-642-39802-5_59

Reuse

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

The AI Singularity and Runaway Human Intelligence

Tony J. Prescott^{1*}

¹Sheffield Centre for Robotics, University of Sheffield, Sheffield, S10 2TN.
t.j.prescott@sheffield.ac.uk

Abstract. There is increasing discussion of the possibility of AI being developed to a point where it reaches a “singularity” beyond which it will continue to improve in a runaway fashion without human help. Worst-case scenarios suppose that, in the future, *homo sapiens* might even be replaced by intelligent machines as the dominant “species” on our planet. This paper argues that the standard argument for the AI singularity is based on an inappropriate comparison of advanced AI to average human intelligence, arguing instead that progress in AI should be measured against the collective intelligence of the global community of human minds brought together and enhanced by smart technologies that include AI. By this argument, AI as a separate entity, is unlikely to surpass “runaway” human (or, perhaps, posthuman) intelligence whose continued advance, fueled by scientific and cultural feedback, shows no sign of abating. An alternative scenario is proposed that human collective intelligence will take an increasingly biohybrid form as we move towards a greater, deeper and more seamless integration with our technology.

Key-words. Societal impact of technology, AI singularity, collective intelligence, human-machine symbiosis, biohybrid.

1 Introduction

Most approaches to the prospects of an AI singularity follow in the path of Good [1] who, writing in 1965, defined an “ultra-intelligent machine” as a device that “can far surpass all the intellectual activities of any man however clever”. As set out with great clarity by Chalmers [2], concerns about the AI singularity focus on the prospect of a positive feedback whereby future machines, initially developed to just exceed human intellect (AI+), rapidly bootstrap themselves to a level of intelligence far greater than that of any person (AI++). Science fiction scenarios that have explored this idea are usually dystopian, with meagre and unimprovable human intelligence fighting to maintain a foothold in a world dominated by ever-strengthening AI. This pessimism also extends to much academic writing on the topic, for instance, in a recent review Muehlhauser and Salamon suggested that “the default outcome from advanced AI is human extinction” [3].

2 Runaway Human Intelligence

But this standard scenario starts from a particular assumption about how we measure human intelligence. The metric chosen for defining AI+ is typically the intelligence of an “average human” (e.g. Chalmers), which we could call HI. If we think of HI as “raw brain power”, which seems to be how many writers conceive of it, then this is, indeed, a more-or-less stationary quantity which, having evolved to its current capacity around 100,000 years ago has changed relatively little since. Given the slow pace of natural selection there is little prospect for further improvement any time soon. It seems plausible, then, that AI could surpass HI in many of its major aspects in the foreseeable future; in some domains it is already, unarguably, ahead.

An alternative comparison can be made, however, that casts the prospect of the singularity in a quite different light. Specifically, we can compare AI, not with individual human intelligence, but with the collective intelligence of the world human population. Furthermore, there is no obvious reason to consider human intelligence stripped of intelligence-enhancing artifacts. Since at least the Upper Paleolithic (10-50,000 years ago), humans have used external symbol systems to store and communicate knowledge and to boost their individual and collective reasoning capacity (see e.g. [4, 5, 6]). Indeed, computers, the internet, and AI itself, are simply the latest inventions in a set of technologies whose prior members include red ochre, papyrus, the abacus, the slide rule, the typewriter, and the telephone. By inventing and exploiting these intelligence-boosting and knowledge preserving technologies, humanity has precipitated an exponential increase in our shared knowledge and in our ability to apply these insights to control our environment according to our goals. This is “runaway intelligence” at the societal level, fueled by its own positive feedback, as cultural and scientific development has led to a larger, more long-lived and better-educated world population. The argument then, is that human intelligence is not constituted, or best described, at the level of the single mind but in terms of the species. We are as intelligent as the culture to which we belong, able to contribute the raw processing power of own brains to an enhanced collective intelligence (ECI) or—as telecommunications increasingly creates a single world community—to what Heylighen has called the “Global Brain” [7]. Being part of this cultural network, in turn, has a transforming effect on what our individual brains can do. Born with an immature and highly plastic nervous system we spend nearly two decades tuning our brains to take advantage of the intelligence-boosting tools that culture has to offer. In the long run this species-level technologically-enhanced intelligence has no obvious ceiling, we can continue to create technologies that complement our natural intelligence, allow us to communicate faster, and make us collectively smarter. If the prospects for the singularity are considered by comparing future AI with this ECI then the notion that humanity will be outstripped and left behind looks much less plausible. An advance in AI to A+ is after all, also an advance for the culture that generated that AI, so AI+ implies ECI+, AI++ implies ECI++, and so on.

3 A Symbiotic Biohybrid Future?

One question we might still ask is how likely it is that humanity will cease to exploit advances in AI that have the potential to boost collective intelligence. A possible threat here, is of a split emerging between AI and ECI, with a sneaky and malevolent version of AI attempting to conceal its advances, biding its time until it is ready to eliminate all of the unnecessary humans—back to the extinction story again. But this scenario, popular in some recent books and films (for example, both the Terminator quadrilogy and Wilson’s *Robopocalypse*), smacks of anthropomorphism, that is, the assumption that AI systems will necessarily share some of humanity’s worst instincts for tribalism. This also underestimates the likely contribution of biological intelligence to the future human-machine collective—there are many things that our brains and bodies do exceptionally well that will be hard for machines to master, and where there will be little economic incentive to improve them in order to do so; symbiotic systems are successful by virtue of their complementarity. The more plausible scenario, then, is that ECI will continue its runaway path but with an increasingly bio-hybrid form due to greater and deeper integration between humans and our intelligence-enhancing technologies. What is good for AI, will then also be good for us.

References

1. Good, I.J. Speculations concerning the first ultraintelligent machine, in Alt, F. & Rubinoff, M. (eds.) *Advances in Computers*, vol 6, New York: Academic Press (1965)
2. Chalmers, D. The Singularity: A philosophical analysis. *Journal of Consciousness Studies*, 17, 9(10), 7–65 (2010)
3. Muehlhauser, L., and Salamon, A. Intelligence explosion: evidence and import. In *Singularity hypotheses: A scientific and philosophical assessment*, ed. Amnon Eden, Johnny Søraker, James H. Moor, and Eric Steinhart. Berlin: Springer (2012)
4. Clark, A. *Being There*. Cambridge, MA: MIT Press (1996)
5. Deacon, T. *The Symbolic Species: The Co-evolution of Language and the Brain*. W. W. Norton & Co., New York (1997)
6. Mithen, S. *The Prehistory of the Mind: A Search for Origins of Art, Religion, and Science*. Thames and Hudson, Ltd., London (1996)
7. Heylighen F. “The Global Brain as a New Utopia”, in: R. Maresch & F. Rötzer (eds.) *Zukunftsfiguren*, Suhrkamp, Frankfurt (2002)