# Review of Preparing for the Future of Artificial Intelligence"

# Review of "Preparing for the Future of Artificial Intelligence"

Alan Bundy

November 16, 2016

In October 2016, the USA's National Science and Technology Council published a report on Artificial Intelligence (AI) [4] that summarised evidence from a wide variety of sources on how they expect AI to develop, what impact it would have and what actions it recommended the USA's Government to take. It built on several previous USA Government reports, e.g., [5, 3] and consulted widely among AI experts in the USA, e.g., [1] and five workshops. A companion document has also been published: "The National Artificial Intelligence Research and Development Strategic Plan", which lays out a strategic plan for Federally-funded research and development in AI.

Overall, this is a comprehensive report, which I recommend to anyone wanting a balanced review of the state of the AI art, its potential impact and what ethical, economic and societal issues it presents. It does, of course, duck some of the more difficult issues — or rather recommend that someone else consider them in detail.

This concern about AI is, of course, triggered by the phenomenal recent successes of mainly statistical machine learning in games (Chess, Go and Jeopardy), self-driving cars, automated assistants (Apple Siri, Amazon Alexa, Google Now and Microsoft Cortana). The other main driver is the worry that Artificial General Intelligence (AGI) will exceed human intelligence and supplant us as the dominant species on the Earth — the, so called, Singularity.

The report wisely ignores concerns about the Singularity, claiming that, if it occurs at all, it's a long way in the future, and that the immediate actions should be the same whether or not it occurs. It points out that the AI successes have been in, what is calls, *narrow* AI, i.e., often super-human performance in a very narrow task, e.g., playing Go. I call such systems *idiot savants*. In contrast, progress in AGI has been disappointing, with no sign that this will improve in the foreseeable future.

So the report focuses on the prospects for narrow AI. It claims that it has already brought "major benefits to the public in fields as diverse as health care, transportation, the environment, criminal justice, and economic inclusion". The report then explores its potential in autonomous vehicles, governance, education, cyber-security and weapons. This exploration uncovers a number of important

issues and concludes with 23 recommendations. These issues and recommendations include the following:

**Advantages:** AI can increase productivity, lower costs, make products and services more widely available, and provide more accuracy and precision.

**Exploitation:** Private and public institutions need to consider how they can take advantage of this potential of AI and hire staff to enable them to do so.

**Research:** There needs to be increased investment in AI research, especially in basic research, where it is not in the immediate interests of industry to invest, so Government has to be involved.

**Education:** The report anticipates that "AI-enhanced education" can assist human teachers to make high-quality education more widely available. The NSTC Committee on Science, Technology and Mathematics Education (CoSTEM) includes this potential within its scope.

**Ethics:** The need not only for ethics to be an integral part of AI education, but also that it be "augmented with technical tools and methods for putting good intentions into practice". Security, privacy and safety, for instance, need to integrated into design and assured.

**Accountability:** The report recognises that it is hard to extract explanations from statistical machine learning programs. Such explanations are essential for many applications. For instance, medics have to take professional responsibility for their decisions. They cannot uncritically accept a machine-generated diagnosis or therapy without some explanation of the reasoning process which generated it. How this crucial drawback should be addressed is not discussed.

**Privacy:** The report recommends that the Government assembles rich collections of data to inform policy making, but mentions only as an aside that this collection must be "consistent with consumer privacy". It has nothing to say about the massive collection of personal data by the private sector, e.g., Facebook, Google. Privacy has, however, been previously addressed in [2].

**Regulation:** Incremental change of regulations is preferred if and when the inadequacies of existing regulations are revealed. Regulatory agencies need to call on AI experts to help them anticipate the changes that will be required.

**Collaboration:** The USA should collaborate with other countries over research and regulation. Government should collaborate with industry.

**Employment:** The report asserts that AI has the potential to eliminate or drive down the wages of low-skilled jobs, so measures are needed to maintain equality and spread the economic benefits broadly. How this is to

be done is hived off to a follow-on report. What is not highlighted is the mismatch between the pace of AI-led disruption and the inertia of societal readjustment. Also, that the number of jobs lost is likely to be much more than the number of new jobs created, at least initially. So, even if the effect is not long-term, we may need to go through a disruptive period of mass unemployment. Additionally, some people have asserted that it will not just be low-skilled jobs that will be affected, but also professional, knowledge-based ones [6].

**Autonomous Weapons:** The Government should develop a policy consistent with humanitarian law. Again, this very difficult issue is not discussed in detail, but hived off to a follow-on report.

**Public Debate:** The report recommends the Government initiate a public debate about these issues. One crucial ingredient of such a debate is public understand of the difference between narrow and general AI. Prior to AI, the space of intelligent systems consisted solely of general intelligence, albeit of widely different capacities across the animal kingdom. Given this background, it is natural to assume that an AI that is world-class in one area, say playing Go, would at least be able to cope in other areas, say driving a car. This is not the case for narrow AI systems, but a sensible debate, e.g., about the dangers of the 'Singularity' cannot be conducted without an understanding of this distinction. The report has nothing to say about this issue.

My criticisms above are intended to be constructive. The report is a very useful contribution to the current debate about the impact of AI. It injects a note of sanity into what is sometimes an over-excited and under-informed controversy. It would be too much to expect it also to answer all the open questions. I have tried to point to some of the unanswered open questions that must be addressed as the debate continues.

# References


[1] B. Horvitz, E. Selman. AAAI presidential panel on long-term AI futures: 2008-2009 study. Technical report, The Association for the Advancement of Artificial Intelligence, February 2009. http://www.aaai.org/Organization/presidential-panel.php.

[2] The Presidents Council of Advisors on Science and Technology. Report to the president: Big data and privacy: A technological perspective. Technical report, Executive Office of the President, May 2014.

[3] United States. Executive Office of the President. Big data: A report on algorithmic systems, opportunity, and civil rights. Technical report, The White House, May 2016.

[4] United States. Executive Office of the President and M. Holden, J.P. Smith. Preparing for the future of artificial intelligence. Technical report, National Science and Technology Council, Washington D.C. 20502, October 2016.

[5] United States. Executive Office of the President and John Podesta. Big data: Seizing opportunities, preserving values. Technical report, The White House, 2014.

[6] Daniel Susskind Richard Susskind. *The Future of the Professions: How Technology Will Transform the Work of Human Experts*. OUP Oxford, 2015.