

Governance of Artificial Intelligence in Australia

Hayden Wilkinson, Australian National University

March 2019

Contents

Background	1
(1) What should be the main goals of government regulation in the area of artificial intelligence?	1
(2.b) What are the gaps in the current regulatory system for artificial intelligence in Australia?	5
(3) Would there be significant economic and/or social value for Australia in a Responsible Innovation Organisation?	6
(5) How should the business case for a Responsible Innovation Organisation be measured?	7
(6) If Australia had a Responsible Innovation Organisation:	8
(a) What should be its overarching vision and core aims?	8
(b) What powers and functions should it have?	8
(d) What internal and external expertise should it have at its disposal?	8
(e) How should it interact with other bodies with similar responsibilities?	9

Background

I am a PhD candidate in the Research School of Social Sciences at the Australian National University. I have previously worked as a Researcher at the Global Priorities Institute within the University of Oxford and am currently a Visiting Research Collaborator at Princeton University. My training is in philosophy, specifically ethics, and my research focuses on the ethical significance of how our actions and policies affect future generations.

I am also familiar with the growing academic literature on the potential benefits and risks of advanced artificial intelligence and the regulation of it. In this submission, I have tried to comment only on issues on which I have some relevant expertise.

(1) What should be the main goals of government regulation in the area of artificial intelligence?

The fundamental goal of regulating AI technologies should be to ensure that the development and deployment of those technologies result in positive outcomes. Of course, that includes mitigating the risks of negative outcomes.

Negative outcomes may involve widespread losses in human welfare, violations of rights, the intensification of economic inequality among individuals, or any combination thereof. As noted in the AHRC and WEF White Paper, we have already observed AI technologies bringing about some such negative outcomes, including:

- unnecessary burdens imposed on Centrelink customers during the ‘robodebt’ affair;
- racial discrimination in sentencing, when poorly understood algorithms were used in the US criminal justice system;¹ and
- the malicious use of automated social media accounts to encourage political discord in the US.^{2,3,4}

At the same time, the safe deployment of AI technologies promises to revolutionise medical science (and has already begun to do so)⁵, accelerate scientific research more generally⁶, improve transportation technologies⁷

¹ Angwin, J. *et al.*, "Machine Bias", *ProPublica*, 23 May 2016, <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>. Accessed 17 Mar. 2019.

² Zaman, T., "Even a few bots can shift public opinion in big ways", *The Conversation*, 5 Nov. 2018, <http://theconversation.com/even-a-few-bots-can-shift-public-opinion-in-big-ways-104377>. Accessed 17 Mar. 2019.

³ Guilbert, D. & S. Woolley, "How Twitter Bots Are Shaping the Election", *The Atlantic*, 1 Nov. 2016, <https://www.theatlantic.com/technology/archive/2016/11/election-bots/506072/>. Accessed 17 Mar. 2019.

⁴ Glenday, J., "Australia needs to be better prepared for an onslaught of 'social bots', researchers warn.", 6 Dec. 2016, <http://www.abc.net.au/news/2016-12-07/are-social-bots-a-threat-to-australian-democracy/8096120>. Accessed 17 Mar. 2019.

⁵ Detwiler, T.M., "Making New Drugs With a Dose of Artificial Intelligence", *The New York Times*, 5 Feb. 2019, <https://www.nytimes.com/2019/02/05/technology/artificial-intelligence-drug-research-deepmind.html>. Accessed 17 Mar. 2019.

⁶ Gil, Y., *et al.* "Amplify scientific discovery with artificial intelligence." *Science* 346.6206 (2014): 171-172.

⁷ Levinson J., *et al.* "Towards fully autonomous driving: Systems and algorithms". *Intelligent Vehicles Symposium (IV)*, IEEE (2011): 163–168.

, and radically increase prosperity.⁸ All of these promise to greatly improve the lives of Australian citizens, as well as humanity more widely.

To achieve positive outcomes such as these, and to avoid negative outcomes, it is essential to prevent the deployment of faulty or unsafe AI systems in contexts which may affect large numbers of people. To avoid this, regulation is needed. And it appears that public sentiment is in favour, at least in settings similar to Australia - when surveyed, 82% of US respondents (and a similar number of European respondents) agreed that artificial intelligence needs to be carefully managed.⁹

But we cannot always be certain of whether AI systems are safe. To ensure that they are, their programmed (or learned) objectives need to be well understood, and any possible negative outcomes need to be foreseen. For advanced systems, this is challenging. In part, this is due to the many unsolved technical questions which remain for such systems (see Amodei *et al.*, 2016¹⁰). And when the safety of such systems cannot be guaranteed, caution is needed.

To effectively ensure positive outcomes, some key principles have been identified by the research community. The *Asilomar AI Principles*¹¹ were developed at the 2017 Beneficial AI Conference¹² and have been endorsed both by prominent AI researchers and by industry leaders, including: the CEO of Google's DeepMind; IBM's Chief Scientist; Research Directors at Google; the Director of AI Research at Facebook; the CTO of Apple; the CEO of Toyota Research Institute; and the Research Director of OpenAI. The principles most relevant to an Australian Research Innovation Organisation are:

6) **Safety:** AI systems should be safe and secure throughout their operational lifetime, and verifiably so where applicable and feasible.

7) **Failure Transparency:** If an AI system causes harm, it should be possible to ascertain why.

8) **Judicial Transparency:** Any involvement by an autonomous system in judicial decision-making should provide a satisfactory explanation auditable by a competent human authority.

9) **Responsibility:** Designers and builders of advanced AI systems are stakeholders in the moral implications of their use, misuse, and actions, with a responsibility and opportunity to shape those implications.

10) **Value Alignment:** Highly autonomous AI systems should be designed so that their goals and behaviors can be assured to align with human values throughout their operation.

11) **Human Values:** AI systems should be designed and operated so as to be compatible with ideals of human dignity, rights, freedoms, and cultural diversity.

⁸ Brynjolfsson, E., & A. McAfee. *The second machine age: work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company, (2014).

⁹ Zhang, B., & A. Dafoe. "Artificial Intelligence: American Attitudes and Trends." Available at SSRN 3312874 (2019). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3312874. Accessed 17 Mar. 2019.

¹⁰ Amodei, D., *et al.* "Concrete problems in AI safety." *arXiv preprint arXiv:1606.06565* (2016).

¹¹ "AI Principles", *Future of Life Institute*, <https://futureoflife.org/ai-principles/>. Accessed 17 Mar. 2019.

¹² "Beneficial AI 2017", *Future of Life Institute*, <https://futureoflife.org/bai-2017/>. Accessed 17 Mar. 2019.

12) **Personal Privacy:** People should have the right to access, manage and control the data they generate, given AI systems' power to analyze and utilize that data.

13) **Liberty and Privacy:** The application of AI to personal data must not unreasonably curtail people's real or perceived liberty.

14) **Shared Benefit:** AI technologies should benefit and empower as many people as possible.

15) **Shared Prosperity:** The economic prosperity created by AI should be shared broadly, to benefit all of humanity.

16) **Human Control:** Humans should choose how and whether to delegate decisions to AI systems, to accomplish human-chosen objectives.

17) **Non-subversion:** The power conferred by control of highly advanced AI systems should respect and improve, rather than subvert, the social and civic processes on which the health of society depends.¹³

Other organisations and partnerships which endorse similar principles are OpenAI¹⁴ - a leading artificial intelligence laboratory in the US - and the Partnership on AI.¹⁵ The latter has already had success in obtaining agreement from more than 80 partner organisations, including industry leaders such as Google, Facebook, IBM, Amazon, Apple, Microsoft, Intel, Samsung, and Baidu.¹⁶

The principles listed above address concerns which are already being realised today. But, over the coming decades, AI systems are expected to be used more widely and have a far greater impact on human lives. The safety of such systems will become even more important. Indeed, according to AI experts, there is a chance that, within the 21st Century, we will see AI systems which exceed human performance on a wide variety of tasks.^{17,18} Such systems are also known as Artificial General Intelligence (AGI) (see Everitt *et al.* 2018¹⁹ for an overview of the relevant literature compiled by researchers at the ANU). Many of the same safety problems we face in the short term also apply in the long term to AGI (see Cave & ÓhÉigeartaigh 2019 as well as).²⁰ In addition, the solutions and policies needed to address short-term safety concerns also serve to address long-term concerns. The 2016 joint report from the White House and US Office of Science and Technology Policy affirms this:

"The best way to build capacity for addressing the longer-term speculative risks is to attack the less extreme risks already seen today, such as current security, privacy, and safety risks, while investing in research on longer-term capabilities and how their challenges might be managed. Additionally, as

¹³ "AI Principles", *Future of Life Institute*, <https://futureoflife.org/ai-principles/>. Accessed 17 Mar. 2019.

¹⁴ "OpenAI Charter", *OpenAI*, 9 Apr. 2018, <https://openai.com/charter/>. Accessed 17 Mar. 2019.

¹⁵ "Tenets - The Partnership on AI", *The Partnership on AI*, <https://www.partnershiponai.org/tenets/>. Accessed 17 Mar. 2019.

¹⁶ "Partners", *The Partnership on AI*, 7 Jun. 2018, <https://www.partnershiponai.org/partners/>. Accessed 17 Mar. 2019.

¹⁷ Grace, K., *et al.* "When will AI exceed human performance? Evidence from AI experts." *Journal of Artificial Intelligence Research* 62 (2018): 729-754.

¹⁸ Müller, V.C., & N. Bostrom. "Future progress in artificial intelligence: A survey of expert opinion." *Fundamental issues of artificial intelligence*. Springer, Cham, 2016. 555-572.

¹⁹ Everitt, T., G. Lea, & M. Hutter. "AGI safety literature review." *arXiv preprint arXiv:1805.01109* (2018).

²⁰ Cave, S., & S. ÓhÉigeartaigh. "Bridging near-and long-term concerns about AI." *Nature Machine Intelligence* 1.1 (2019): 5.

research and applications in the field continue to mature, practitioners of AI in government and business should approach advances with appropriate consideration of the long-term societal and ethical questions – in addition to just the technical questions – that such advances portend.”²¹

The implementation of solutions to immediate AI safety problems will hence become all the more important in the coming decades. Effective regulation in this area will become even more crucial in order to safeguard the wellbeing of Australians.

In addition, to address longer-term concerns specifically, the Asilomar Principles also include the following:

- 19) **Capability Caution:** There being no consensus, we should avoid strong assumptions regarding upper limits on future AI capabilities.
- 20) **Importance:** Advanced AI could represent a profound change in the history of life on Earth, and should be planned for and managed with commensurate care and resources.
- 21) **Risks:** Risks posed by AI systems, especially catastrophic or existential risks, must be subject to planning and mitigation efforts commensurate with their expected impact.
- 22) **Recursive Self-Improvement:** AI systems designed to recursively self-improve or self-replicate in a manner that could lead to rapidly increasing quality or quantity must be subject to strict safety and control measures.²²

I recommend that the goals of AI regulation in Australia should be, in addition to other priorities, to ensure that these principles are satisfied in the deployment of AI systems in Australia and, ideally, by Australian firms elsewhere.

But regulating artificial intelligence will be no easy task. Due to the nature of AI, there are several key considerations which may hinder regulatory efforts. For a new regulatory system to be effective, its goals need to incorporate the following considerations as well as strategies to overcome them.

1. **Technical developments are being made rapidly**, and new AI capabilities can be deployed by major firms even more rapidly. In this area, regulation will very quickly become outdated. It is crucial, therefore, that regulatory bodies respond quickly and appropriately to new developments. This will mean that they will need to be highly flexible and have the discretion to quickly respond to emerging issues. It also means that regulatory bodies must be aware of developments in real time, and hence must consistently monitor new advances. Ideally, such bodies should also endeavour to analyse and predict what impacts future advances in AI are likely to have. Achieving this will require close collaboration with the technical research community, especially those parts of the community working on cutting-edge techniques like reinforcement learning. It may also require the recruitment of in-house technical experts.

²¹"Preparing for the Future of Artificial Intelligence (Executive Office of the President, National Science and Technology Council, Office of Science and Technology Policy)", 12 Oct. 2016, https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf. Accessed 17 Mar. 2019.

²²"AI Principles", *Future of Life Institute*, <https://futureoflife.org/ai-principles/>. Accessed 17 Mar. 2019.

2. **AI technologies often have multiple potential uses**, some of which have serious security implications. For instance, the text generator GPT-2, developed by OpenAI in 2019²³, can quickly generate convincing human-sounding text to match any given writing prompt. If misused, it could generate large quantities of ‘fake news’ content while requiring only minimal human labour. (Fortunately, OpenAI opted not to release the tool to the public.²⁴) To ensure that similarly risky AI systems are not made widely available, regulators need to effectively identify potential malicious uses of such systems. As above, this too may require significant in-house expertise and/or the input of external experts who fully understand the capabilities of new systems. This is a key recommendation of the 2018 report *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*.²⁵ (See the full report for a comprehensive discussion of dual-use security implications and policy recommendations.)
3. **New AI technologies are often easily deployed across national borders**. A dangerous new AI tool can be developed and deployed elsewhere in the world in a way which affects Australians - for instance, any AI tool applied by social media websites. Alternatively, dangerous algorithms which are developed internationally can then be downloaded in Australia, to then be used domestically. In particular, this raises security concerns, whether for cybersecurity, physical security, or political security (see Brundage *et al.* 2018: 23-49). For instance, the software needed to run a lethal autonomous weapons system could, in future, be developed in an AI laboratory elsewhere and then be obtained and deployed by bad actors within Australia (as long as they have access to the necessary physical materials). Or, in an example unrelated to security, a foreign firm can develop a loan evaluation algorithm which implicitly discriminates based on race, and this algorithm can then be obtained and deployed by an Australian loan provider. To curtail the development of risky new AI systems such as these, Australian regulators must coordinate with their international counterparts to ensure consistent monitoring and enforcement across borders. The Australian government must also support and actively participate in the setup of new international bodies to serve these purposes.

(2.b) What are the gaps in the current regulatory system for artificial intelligence in Australia?

To my knowledge, the current regulatory system neglects all three of the considerations raised above (and the corresponding strategies needed to address them).

First, legislation is far too slow to address the constantly advancing field of AI. Non-legislative regulatory bodies may be faster, but I am not aware of any existing bodies which address unsafe AI systems in general. I

²³ Radford, A., *et al.*, “Language models are unsupervised multitask learners”, *OpenAI*, https://d4mucfpksywv.cloudfront.net/better-language-models/language_models_are_unsupervised_multitask_learners.pdf. Accessed 17 Mar. 2019.

²⁴ “OpenAI says its text-generating algorithm GPT-2 is too dangerous to release to the public”, *Slate*, 22 Feb. 2019, <https://slate.com/technology/2019/02/openai-gpt2-text-generating-algorithm-ai-dangerous.html>. Accessed 17 Mar. 2019.

²⁵ Brundage, M., *et al.* “The malicious use of artificial intelligence: forecasting, prevention, and mitigation.” *arXiv preprint arXiv:1802.07228* (2018).

am aware, however, that leading technical researchers are rarely if ever consulted on new developments or regulatory measures. This situation could be greatly improved. Without expert technical advice, I cannot envisage Australian regulatory efforts being able to keep pace with technical developments, being able to identify the possible malicious uses of new AI systems, or being able to effectively regulate them.

Second, as far as I am aware, Australian regulators do not coordinate with regulators in other countries to prevent the development of risky technologies. Regulation of AI seems not to be a priority of Australian diplomatic missions. This is despite the potential impact that internationally-developed AI systems may have on the lives of Australians. This is particularly worrying as developments in the field of AI have so far been concentrated within specific firms, most notably: Google's DeepMind laboratory (based in the UK)²⁶; OpenAI (based in the US)²⁹; Facebook (US)³⁰; and IBM (US)³¹. Chinese firms, such as Baidu and Tencent, are also likely to contribute to future AI developments (see Ding 2018).³² No commercial labs in Australia have made contributions comparable to these, or are likely to in the next decade. Thus, one of the most vital roles that Australia can play in this field is likely to be in advocating for effective regulation in other countries.

(3) Would there be significant economic and/or social value for Australia in a Responsible Innovation Organisation?

One factor which will limit the impact of a Responsible Innovation Organisation in Australia is that most technical developments in AI are taking place overseas. (See (2.b) above.) But this does not mean that there won't still be considerable value in such an organisation if it has the right goals and implementation.

It is clear that many of the impacts of more widespread use of AI systems will be felt in Australia (see (1) above). For example, the use of faulty algorithms by the Department of Human Services has already had adverse effects on Centrelink customers; discriminatory algorithms could easily have been implemented in the Australian court system; social media bots already pose a threat to Australian electoral processes.³³ In addition to these problems, dual-use AI technologies pose a considerable security risk to government, to

²⁶ "AlphaZero: Shedding new light on the grand games of chess, shogi" 6 Dec. 2018, <https://deepmind.com/blog/alphazero-shedding-new-light-grand-games-chess-shogi-and-go/>. Accessed 17 Mar. 2019.

²⁷ "AlphaFold: Using AI for scientific discovery | DeepMind." 2 Dec. 2018, <https://deepmind.com/blog/alphafold/>. Accessed 17 Mar. 2019.

²⁸ "AlphaStar: Mastering the Real-Time Strategy Game StarCraft II" 24 Jan. 2019, <https://deepmind.com/blog/alphastar-mastering-real-time-strategy-game-starcraft-ii/>. Accessed 17 Mar. 2019.

²⁹ "Better Language Models and Their Implications - OpenAI." 14 Feb. 2019, <https://openai.com/blog/better-language-models/>. Accessed 17 Mar. 2019.

³⁰ "Facebook AI Creates Its Own Language In Creepy Preview Of ... - Forbes." 31 Jul. 2017, <https://www.forbes.com/sites/tonybradley/2017/07/31/facebook-ai-creates-its-own-language-in-creepy-preview-of-our-potential-future/>. Accessed 17 Mar. 2019.

³¹ "IBM's Watson Is Everywhere—But What Is it? - MIT Technology Review." 27 Oct. 2016, <https://www.technologyreview.com/s/602744/ibms-watson-is-everywhere-but-what-is-it/>. Accessed 17 Mar. 2019.

³² Ding, J., "Deciphering China's AI dream", *Centre for the Governance of AI, University of Oxford*, (2018) https://www.fhi.ox.ac.uk/wp-content/uploads/Deciphering_Chinas_AI-Dream.pdf. Accessed 17 Mar. 2019.

³³ Glenday, J., "Australia needs to be better prepared for an onslaught of 'social bots', researchers warn.", 6 Dec. 2016, <http://www.abc.net.au/news/2016-12-07/are-social-bots-a-threat-to-australian-democracy/8096120>. Accessed 17 Mar. 2019.

firms, and to individuals (see Brundage *et al.* 2018).³⁴ Even just in cases such as these, there would be significant economic and social value in mitigating these risks.

(5) How should the business case for a Responsible Innovation Organisation be measured?

I can speak only in general terms here. Fundamentally, the business case for such an RIO should be measured by:

a) the magnitude of the benefits and harms it will bring, both for present and (crucially) for future generations.

And, as with any effort to manage probabilistic benefits and harms, this measure must be appropriately sensitive to:

b) the probability of those benefits and harms; and

c) the probability that the RIO will succeed in bringing about those benefits or preventing those harms.

This means that a high probability of some harm is more significant than a low probability of an equivalent harm. So too, it is better that the RIO have a high probability of preventing some harm than a low probability of preventing an equivalent harm. Most crucially, it means that, if a harm is great enough, even a low probability of that harm is worth responding to. For instance, there may be only a low probability of a breakthrough AI system in the next decade posing a major threat to cybersecurity, financial markets, or electoral fidelity. But the disastrous effects which it would bring should still make that risk a high priority. It ought to be a higher priority than many risks of with smaller effects, even if they are more likely to occur.

Given this, the case for an RIO should be measured not just by the benefits it will definitely bring, but also by how well it mitigates these low-probability risks of disaster. There is a considerable amount of technical research on the risks of unsafe AI systems, which must inform the RIO's priorities and regulatory measures (see Amodei *et al.*, 2016³⁵). Given this, when measuring the business case for an RIO, it is important to involve technical experts who can assess whether it will be likely to address the most significant risks as is needed.

³⁴ Brundage, M., *et al.* "The malicious use of artificial intelligence: forecasting, prevention, and mitigation." [arXiv preprint arXiv:1802.07228](https://arxiv.org/abs/1802.07228) (2018).

³⁵ Amodei, D., *et al.*, "Concrete problems in AI safety." [arXiv preprint arXiv:1606.06565](https://arxiv.org/abs/1606.06565) (2016).

(6) If Australia had a Responsible Innovation Organisation:

(a) What should be its overarching vision and core aims?

The RIO's overarching aims should be in line with the best current thinking of AI experts and industry leaders (see (1) above). They should match, or at least include, many of the relevant Asilomar Principles.

(b) What powers and functions should it have?

The powers and functions listed on page 16 of the White Paper are all laudable. If these are used in pursuit of the principles listed above, I expect that the RIO will have a considerable positive impact in the decades to come.

I recommend three additional functions.

The first: **to encourage and support research** within Australia to improve, ensure, and verify the safety of AI systems. As mentioned above (see (2.b)), most research developments in AI capabilities are taking place overseas at major commercial labs. But research into the *safety* of AI systems can occur at any university. Amodei *et al.* (2016)³⁶ describes various concrete problems, if we could find solutions for which, we could greatly improve the safety of advanced AI systems. Solutions to some of these problems would also allow us to determine the safety of some advanced systems before allowing them to be deployed within Australia.

The second: **to coordinate with regulators in other countries** and to advocate for improved regulation of AI worldwide. Given the fact that most developments on the capabilities front are occurring overseas (see above), this may be one of the most high-impact avenues open to a government body seeking to ensure that Australians are benefited rather than harmed by new AI technologies.

The third: **to monitor research developments**, especially on potentially dangerous dual-use technologies, with the assistance of technical experts. As mentioned above (in (1)), this will be crucial to developing effective standards for AI governance domestically. A reliable scheme for monitoring new AI developments, and making that information widely available, would also provide a global public good and assist in encouraging international action on AI regulation.

(d) What internal and external expertise should it have at its disposal?

Ideally, an RIO should have subject-area experts available internally to monitor new research developments and, crucially, to assist in producing new regulatory standards (as mentioned under (1) and (2.b) above). This includes not only technical experts but also experts in the social sciences to accurately predict the impacts of new AI technologies (and the impacts of regulatory proposals).

Whether or not it employs internal experts, an RIO must also make use of the researchers who are at the forefront of current developments. There will still be valuable insights to be gained from consulting

³⁶ *ibid.*

technical researchers who themselves work on advancing AI capabilities. There is also a great deal of value in consulting researchers at the forefront of technical AI safety, as well those researchers who are contributing to the growing field of AI governance (such as the recently established Center for the Governance of AI³⁷ at the University of Oxford). These are highly specialised areas in which generalist policy researchers will often lack expertise.

(e) How should it interact with other bodies with similar responsibilities?

To best ensure that new AI technologies benefit Australians rather than harm their interests, an RIO should have access to the best available knowledge on the issue. In the last decade, specific research institutes have been established to intensively research this issue, and have made considerable progress. These include: the Future of Humanity Institute at the University of Oxford³⁸; the Center for the Governance of AI, also at the University of Oxford³⁹; the Centre for the Study of Existential Risk at the University of Cambridge (headed by Australian philosopher Huw Price)⁴⁰; and the Center for Human-Compatible AI at the University of California, Berkeley.⁴¹ The RIO should not have to reinvent the wheel, so it would be wise to make use of the current state of the art in this area, as represented by these, the most well-established research bodies working on these problems. Where possible, it would be beneficial to directly seek advice from such bodies. The Future of Humanity Institute and Centre for the Study of Existential Risk, in particular, have been closely involved in existing regulatory efforts in the United Kingdom and internationally.^{42,43,44,45}

I should also briefly reiterate from above that an RIO, or other Australian government body, should actively seek to establish multilateral partnerships with regulatory bodies in other nations. Ideally, the regulation of unsafe AI technologies should be a high-priority issue within Australia's diplomatic efforts. When some AI technologies can easily be transferred over the internet, effective regulation within Australia may well depend on the success of international coordination.

³⁷ "Center for the Governance of AI" <https://www.fhi.ox.ac.uk/governance-ai-program/>. Accessed 17 Mar. 2019.

³⁸ "Contact - Future of Humanity Institute." <https://www.fhi.ox.ac.uk/contact/>. Accessed 17 Mar. 2019.

³⁹ "Center for the Governance of AI" <https://www.fhi.ox.ac.uk/governance-ai-program/>. Accessed 17 Mar. 2019.

⁴⁰ "Contact - The Centre for the Study of Existential Risk." <https://www.cser.ac.uk/contact/>. Accessed 17 Mar. 2019.

⁴¹ "Center for Human-Compatible AI." <https://humancompatible.ai/>. Accessed 17 Mar. 2019.

⁴² "FHI researchers cited in UK Parliamentary "Robotics and artificial" 20 Oct. 2016, <https://www.fhi.ox.ac.uk/fhi-parliamentary/>. Accessed 17 Mar. 2019.

⁴³ "CSER news Advice to UN High-level Panel on Digital Cooperation" 26 Feb. 2019, <https://www.cser.ac.uk/news/advice-un-high-level-panel-digital-cooperation/>. Accessed 17 Mar. 2019.

⁴⁴ "CSER news Advice to EU High-Level Expert Group on Artificial" 26 Feb. 2019, <https://www.cser.ac.uk/news/advice-eu-high-level-expert-group-artificial-intel/>. Accessed 17 Mar. 2019.

⁴⁵ "Advice to the US Dept of Commerce." 19 Feb. 2019, <https://www.cser.ac.uk/news/advice-us-dept-commerce/>. Accessed 17 Mar. 2019.