

# Artificial Intelligence: governance and leadership

WHITE PAPER • 2019



Australian  
Human Rights  
Commission

WORLD  
ECONOMIC  
FORUM

COMMITTED TO  
IMPROVING THE STATE  
OF THE WORLD

© Australian Human Rights Commission and World Economic Forum 2019.

We encourage the dissemination and exchange of information presented in this publication and endorse the use of the **Australian Governments Open Access and Licensing Framework (AusGOAL)**.



All material presented in this publication is licensed under the **Creative Commons Attribution 4.0 International Licence**, with the exception of:

- photographs and images
- the Commission's logo, any branding or trademarks
- where otherwise indicated.

To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/legalcode>.

In essence, you are free to copy, communicate and adapt the publication, as long as you attribute the Australian Human Rights Commission and abide by the other licence terms.

**Please give attribution to:** © Australian Human Rights Commission and World Economic Forum 2019.

**Artificial Intelligence: governance and leadership • White paper • 2019**

ISBN 978-1-921449-99-4

For further information about copyright in this publication, please contact:

Communications Unit  
Australian Human Rights Commission  
GPO Box 5218  
SYDNEY NSW 2001  
Telephone: (02) 9284 9600  
TTY: 1800 620 241  
Email: [communications@humanrights.gov.au](mailto:communications@humanrights.gov.au).

**Design and layout:** Dancinggirl Designs

**Cover image:** Shutterstock

**Internal photography:** Shutterstock, Adobe Stock

**The Australian Human Rights Commission** is Australia's National Human Rights Institution. It is an independent statutory organisation with responsibility for leading the promotion and protection of human rights in Australia. Further information about the Commission can be found at [www.humanrights.gov.au/about-commission](http://www.humanrights.gov.au/about-commission).

**The World Economic Forum** is the International Organisation for Public-Private Cooperation. The Forum engages the foremost political, business and other leaders of society to shape global, regional and industry agendas. Further information about the World Economic Forum can be found at [www.weforum.org/about/world-economic-forum](http://www.weforum.org/about/world-economic-forum).

**Authors:** Nicholas Davis, Sophie Farthing, Edward Santow and Lisa Webber Corr.

#### **Acknowledgements:**

- The World Economic Forum Centre for the Fourth Industrial Revolution, and particularly Kay Firth-Butterfield, for her contribution.
- Australian Human Rights Commission staff Darren Dick, John Howell, Katerina Lecchi, Zoe Paleologos, Lauren Perry and Phoebe Saintilan.
- Expert Reference Group for their peer review and advice in the preparation of this Paper: Amanda Alford, The Honourable Justice Margaret Beazley AO, Professor Genevieve Bell, Peter Dunne, Dr Tobias Feakin, Dr Alan Finkel AO, Verity Firth, Peter Leonard, Brooke Massender, Sean Murphy, Myfanwy Wallwork and Professor Toby Walsh.
- Major project partners for contributing expertise and resources: the Australian Government Department of Foreign Affairs and Trade, Herbert Smith Freehills, LexisNexis, The University of Technology Sydney.
- The Australian Government Digital Transformation Agency.

# Artificial Intelligence: governance and leadership

## White paper

January 2019

Australian Human Rights Commission and World Economic Forum 2019



Australian  
Human Rights  
Commission

WORLD  
ECONOMIC  
FORUM

COMMITTED TO  
IMPROVING THE STATE  
OF THE WORLD

# Contents

<b>1</b>	<b>Foreword</b>	<b>5</b>
<b>2</b>	<b>Introduction</b>	<b>7</b>
<b>3</b>	<b>Protecting human rights in the context of AI</b>	<b>9</b>
<b>4</b>	<b>Does Australia need a Responsible Innovation Organisation?</b>	<b>11</b>
	4.1 The case for a new form of governance for AI	11
	4.2 Assessing the business case for change	12
	(a) Economic impact	12
	(b) Why does responsible AI matter?	13
	(c) How would a Responsible Innovation Organisation directly add value to Australian organisations?	13
<b>5</b>	<b>What should a Responsible Innovation Organisation look like?</b>	<b>15</b>
	5.1 Vision, objectives and guiding principles	15
	5.2 Functions and powers	16
	5.3 Partnership with government and industry	16
	5.4 Timeline: evaluation and monitoring	16
<b>6</b>	<b>Consultation questions</b>	<b>18</b>
<b>7</b>	<b>Consultation process and outcomes</b>	<b>19</b>

# 1 Foreword

The Australian Human Rights Commission and the World Economic Forum are working together to explore models of governance and leadership in artificial intelligence (AI) in Australia. This White Paper has been produced to support a consultation process on this issue.

AI can enable prediction and problem-solving approaches that save the lives of seriously ill hospital patients.<sup>1</sup> Yet AI can also be used to threaten human rights. For example, we have seen allegations of AI entrenching bias and discrimination in the United States (US) criminal justice system,<sup>2</sup> as well as in policing in Australia.<sup>3</sup>

Scandal and controversy connected to new technologies have increased public concern regarding decision-making that uses AI, data privacy, cyber security, political influence and labour market shifts.

Powerful new technological capabilities are rapidly increasing and are changing our world. AI, biotechnologies, neurotechnologies, new materials and distributed ledgers are being developed. As costs fall in data storage, processing and communication, innovative private sector players are particularly active in making these technologies more widespread.

AI and Machine Learning (ML)—thanks to vastly expanded data sets, purpose-designed chipsets and low-cost cloud computing—are enabling breakthroughs, including:

- new forms of communication
- medical diagnosis and treatment
- industrial and consumer robotics
- business analytics
- facial recognition
- natural language processing.

Our challenge as a nation is to ensure these technologies deliver what Australians need and want, rather than what they fear. There is added urgency because other countries are investing heavily in these areas.

Globally, we are witnessing a fundamental shift. Leaders in the technology industry are increasingly abandoning a long-standing hostility to government intervention. Many are starting to call for new forms of governance and regulation.

At the World Economic Forum's Annual Meeting in January 2018,<sup>4</sup> Uber Chief Executive Officer (CEO) Dara Khosrowshahi said: 'My ask of regulators would be to be harder in their ask of accountability.'

Similarly, Salesforce's Marc Benioff said that 'the point of regulators and government [is] to come in and point true north'.<sup>5</sup>

In April 2018, Facebook CEO Mark Zuckerberg said:

[T]he real question, as the Internet becomes more important in people's lives, is what is the right regulation, not whether or not there should be regulation.<sup>6</sup>

In November 2018, Apple CEO Tim Cook stated:

We have to admit the free market is not working ... and it hasn't worked here. I think it's inevitable that there will be some level of regulation.<sup>7</sup>

But what are we looking to regulate or govern better? The new, post-digital age—the so-called 'Fourth Industrial Revolution'—encompasses a variety of digital, biological and physical technologies. Many of those new technologies are powered, at least in part, by a variety of algorithmic techniques, described collectively as AI and ML.

To date, community concern has focused on the right to privacy: such as who owns, controls and exploits the personal data of individuals using AI-powered social media.

The potential impact of AI, including on other human rights, goes beyond privacy. For example, AI and related technologies could:

- bring radical changes in how we work, with predicted large-scale job creation and destruction and new ways of working
- transform decision-making that affects citizens' basic rights and interests
- increase our environmental impact
- become so important in how we live that accessibility of that technology becomes an even more important human rights issue
- have a profound impact on our democratic institutions and processes.

Adopting the right governance framework is difficult, because AI technologies are complex, are applied across all sectors of the Australian community, and have enormous capacity for social good, social harm—and often both simultaneously. However, Australian stakeholders need to consider and experiment with innovative models for ensuring that the economic gains, social influence and security impact of AI is positive for all.

Former US Secretary of State Madeleine K Albright pointed out that ‘citizens are speaking to their governments using 21st century technologies, governments are listening on 20th century technology and providing 19th century solutions’.<sup>8</sup> If our governance solutions are so far out of step with the powerful technologies we are collectively deploying, unanticipated risks are inevitable.

Protecting Australians, while powering our future economy, requires innovation that reinforces Australia’s liberal democratic values, especially human rights, fairness and inclusion. Making this vision real is a complex task. It will involve, for example, carefully crafted laws supported by an effective regulatory framework, strong incentives that apply to the public and private sectors, and policies that enable Australians to navigate an emerging AI-powered world.

The Australian Human Rights Commissioner is currently leading a project focusing on human rights and technology. An Issues Paper, that seeks to address many of these issues, was published by the Commission in July 2018. One question the Issues Paper sought feedback on is whether Australia needs a better system of governance to harness the benefits of innovation using AI and other new technologies, while effectively addressing the threats to our human rights.

The Commission and the World Economic Forum have produced this White Paper to expand on that question posed in the Commission’s Issues Paper. Based on early analysis of data received by the Commission, this White Paper starts with the hypothesis that Australia needs to match the rising levels of innovation in AI technologies with innovation in AI governance, and focuses on the *practical* challenge of exploring what that might look like. The White Paper, therefore, focuses on one key question: whether Australia needs an organisation to take a central role in promoting responsible innovation in AI and related technology.

We invite you to share your perspectives.



**Nicholas Davis**  
Head of Society and Innovation  
Member of the Executive Committee  
World Economic Forum



**Edward Santow**  
Human Rights Commissioner  
Australian Human Rights Commission

## 2 Introduction

This White Paper aims to identify how Australia can simultaneously foster innovation and protect human rights through the application of new technologies, in particular AI.

AI is a key driver of the Fourth Industrial Revolution and is expected to transform the global economy within the next decade, adding 40 per cent to the world's Gross Domestic Product (GDP) by 2030.<sup>9</sup> Today, only 9 per cent of Australia's listed companies are making sustained investments in AI, lagging behind the 20 per cent in the United States.<sup>10</sup> However, many have recognised AI's transformative potential and have begun to deploy AI in their businesses.<sup>11</sup> AlphaBeta reports that AI has opened up new markets and new opportunities in critical areas such as health care, transportation, criminal justice, the environment, and economic inclusion,<sup>12</sup> and 'will be the ultimate tool because it will help us build all possible tools'.<sup>13</sup>

The scope and pace of change generated by AI also pose unprecedented challenges, with radical disruptions to our social, governmental and economic systems.<sup>14</sup> Despite AI's potential for beneficial use, its use creates important risks to Australians, including exclusion, discrimination, privacy, skill loss, economic impacts, security of critical infrastructure, and social well-being.<sup>15</sup>



Creating an effective, efficient Australian governance regime, adapted to the range of different approaches to and applications of AI, is not a mere matter of overcoming technical challenges. Technologies can influence our environment and behaviour, with wide-ranging social impacts. A technology is never just a tool. It is imprinted with design choices that can lead to both intended and unintended consequences. This, in turn, determines how benefits are distributed or life is experienced for the different groups of people that come into contact with the technology in question.<sup>16</sup> For example, socio-economic and other characteristics can significantly affect access to technology for children as they grow, learn and socialise.

There is strong evidence to suggest that Australia's continued economic and social progress relies on AI to drive growth, improve productivity and solve a wide range of social and economic challenges. Simultaneously, AI can unleash serious harm. If that harm is not addressed, it will undermine public confidence in this technology. That, in turn, will stall AI-powered innovation and its associated benefits.

This White Paper proposes an approach to AI and associated technologies that is forward-looking and agile, while simultaneously fostering innovation and human rights. It starts with the hypothesis that Australia needs to match the rising levels of innovation in AI technologies with innovation in AI governance.

The governance of AI could be achieved in multiple ways, including by assigning a leadership role in this area to a new or existing organisation, which this White Paper calls the 'Responsible Innovation Organisation'.

Such an organisation could combine capacity building, expert advice, governance, leading practices and innovative interventions that foster the benefits of AI while mitigating risks. In testing these hypotheses, it is then necessary to consider what would be the Responsible Innovation Organisation's key features — its aims, functions, powers, structure and so on.

Accordingly, the questions in this White Paper invite stakeholders to comment on:

- the nature and scope of the challenge for human rights protection posed by the rise of AI
- whether Australia needs a new or existing organisation to lead in the promotion of responsible innovation in AI
- if so, what might be the aims, functions and roles of such an organisation.





## 3 Protecting human rights in the context of AI

International and Australian human rights law requires that individuals be treated without discrimination.<sup>17</sup> Governments must uphold human rights, while businesses have a responsibility to respect human rights in all their operations.<sup>18</sup>

The most effective way of complying with these obligations in the context of AI is to ensure that it is designed and used responsibly, by protecting privacy, fairness, equality and other human rights.

This is, of course, not solely an Australian challenge. Public and private sector organisations globally are exploring ways to understand and manage the impact of bias in AI and ML. The interconnected nature of the global economy, combined with the fact that such explorations in innovative governance are at an early stage in all jurisdictions, means that Australian institutions have an opportunity to lead in developing new structures, policies and relationships that can help address these important issues on behalf of all Australians.

Accordingly, a responsible innovation framework could accommodate imperatives that sometimes sit in tension, by anticipating and addressing the potential harms of AI, so that it can be deployed in a way that is safe and beneficial for Australia.

### **Case study: artificial intelligence and the risk of discrimination**

Bias and discrimination in technology have entered the public consciousness along with our increasing reliance on and understanding of AI and ML. AI systems can discriminate and operate unfairly for many reasons.

For example:

- AI is designed by human beings who possess inherent biases and is often trained with data that reflects the imperfect world that we live in.<sup>19</sup>
- Training AI systems with data that is not representative, or using data that reflects bias or prejudice (for example, sexism or racism), can lead to an AI-supported decision that is unfair, unjust, unlawful or otherwise wrong.<sup>20</sup>
- AI's algorithms can include discriminatory variables (for example, including a variable for private school attendance in a loan application algorithm) that results in further discrimination.<sup>21</sup>
- Where users do not understand AI's limitations, especially if they assume AI's predictions to be more accurate and precise (and thus more authoritative) than those made by people, this can result in unfairness.<sup>22</sup>
- AI can be deployed in an inappropriate context (for example, deploying a model in a different cultural context from that in which it was originally trained).<sup>23</sup>
- Personal data is the 'fuel' for AI.<sup>24</sup> It can be at risk when deployed in ML models, as hackers can often threaten individual privacy by reverse-engineering algorithms, which could allow access to the personal data the algorithm is trained on.<sup>25</sup>

There are three primary reasons for Australia to be concerned about bias and discrimination in AI systems:

1. Automated decision-making systems will be applied more often by both the private and public sectors and at a greater scale across a wide variety of essential services, from decisions in health care to financial services. Discrimination in these decisions is both more likely and of greater consequence for groups that are already vulnerable.<sup>26</sup>
2. It is difficult to know the decision-making process adopted in an AI system, because ML tends to involve opaque proprietary algorithms.<sup>27</sup> Without understanding this process, it is hard to discern whether, when or how such systems are discriminating against a group or individual. This fundamentally challenges the concept of procedural fairness in administrative decision-making.
3. Hasty implementation of AI puts at risk its benefits by undermining public trust in new technologies. Public trust in Australian businesses, the government, media and civil society has fallen rapidly in the last decade to record lows.<sup>28</sup> If this trust is further eroded by the emergence of widespread discrimination through the deployment of AI systems, it may slow adoption in ways that prevent Australians from harnessing the many positive impacts of AI and ML.

The first set of consultation questions focuses on understanding your sense of the challenge itself and the general approach that you feel the Government should take in this area.

## Consultation questions

1. What should be the main goals of government regulation in the area of artificial intelligence?
2. Considering how artificial intelligence is currently regulated and influenced in Australia:
  - (a) What existing bodies play an important role in this area?
  - (b) What are the gaps in the current regulatory system?



# 4 Does Australia need a Responsible Innovation Organisation?

## 4.1 The case for a new form of governance for AI

This White Paper is published alongside the Issues Paper for the Commission's Human Rights and Technology Project (the Commission's Project). While the Commission's Project is a separate consultation process, a number of submissions to the Commission's Project have raised issues related to governance that are relevant to the issues considered in this White Paper. Some stakeholders, for example, commented on the need for whole-of-government responses to protecting human rights given rapid technological advances, emerging gaps between portfolio responsibilities and legislative instruments and the likelihood that reform may be required across the range of government portfolios.<sup>29</sup>

Stakeholders also commented on the potential role of co- and self-regulation, as well as enforceable sanctions and compliance monitoring, within a framework that incorporates Australian cultural, legal and social norms.<sup>30</sup>

Submissions to the Commission's Project have consistently expressed support for the concept of a new type of regulatory body to govern and promote responsible innovation.

One stakeholder, the University of Technology Sydney (UTS), made a central recommendation for the Australian Government to establish a 'Technology Assessment Office' (TAO) in order to protect Australia's interests and ensure technology develops to promote the interests of all Australians in a positive and human rights compliant way. UTS recommended the TAO perform a range of functions to help shape the future of technological innovation and design in Australia, including research, collaboration with relevant government bodies and other organisations and to conduct social outreach and education in both the public and private sectors.<sup>31</sup>

This is also consistent with recent preliminary recommendations from the Australian Competition and Consumer Commission (ACCC) for greater regulatory oversight of digital platforms.

Throughout 2018 the ACCC has undertaken an inquiry into the impact of online search engines, social media and digital content aggregators (digital platforms) on competition in the media and advertising services market. The preliminary report,<sup>32</sup> released on 10 December 2018, acknowledges this is a critical time 'in the development of digital platforms and their impact on society'. The report notes, amongst other things, concerns with the lack of transparency in digital platform operations (including in their algorithms) as well as with consumer awareness and understanding of the extensive amount of information collected about them.

The ACCC considers that a regulatory authority could monitor, investigate and report on discriminatory and anti-competitive conduct and provide assurances to businesses, consumers and governments on the performance and impact of key algorithms and policies. Proposed powers would include complaints investigations, referrals to other government agencies and also to publish reports and make recommendations. Further, the ACCC suggests that the establishment of a digital platforms ombudsman could—without duplicating other regulatory functions—deal with complaints from consumers, advertisers, media companies and other business users of digital platforms. Terms of Reference would include remedies the proposed ombudsman could recommend or implement. Other preliminary recommendations include the development of an enforceable Code of Practice for digital platforms and amendments to the *Privacy Act 1988* regarding the use and collection of personal information. The ACCC Final Report is due in mid-2019.

The concept of a new regulatory and governance body also reflects similar ideas and processes that are being trialled in other countries.

The Government of the United Kingdom (UK), for example, announced in November 2017 that it would establish a new Centre for Data Ethics and Innovation ‘to enable and ensure safe, ethical and ground-breaking innovation in AI and data driven technologies’.<sup>33</sup> Consultation was conducted in mid-2018 on the anticipated activities and work of the Centre, with six possible areas proposed to strengthen governance of uses of data and AI (targeting fairness, transparency, liability, data access and intellectual property and ownership).<sup>34</sup> The Centre has commenced operations, and is one of three new organisations being established by the UK Government to guide policy, develop opportunities and harness the potential for AI.<sup>35</sup>

## 4.2 Assessing the business case for change

### (a) Economic impact

According to analysis by consultancy AlphaBeta, automation represents a more than two trillion dollar opportunity for the Australian economy, with the potential to add 2.2 trillion dollars to cumulative Australian GDP between 2017 and 2030.<sup>36</sup> This economic calculus is based on the ability to prepare future workforces and transition current workers to jobs and sectors where productivity is higher.

Meanwhile, Accenture analysed 12 developed economies and found that AI has the potential to double their annual economic growth rates by 2035 by acting as a new factor of production alongside capital and labour.<sup>37</sup>

Such analysis suggests there is a strong comparative business case for accelerating Australia’s deployment of AI across the economy. It also aligns with Australian Government policies that promote an open, innovative economy in an interconnected world. For example, the Australian Government provided \$29.9m in the 2018–19 Budget (over four years) to strengthen Australia’s capability in AI and ML, supporting economic growth and the productivity of Australian businesses.<sup>38</sup>

AI is expected to have the greatest direct impact on sectors and companies whose business models and data infrastructure allow the immediate adoption of emerging ML techniques. For example, Accenture research shows that information and communication, manufacturing and financial services are the three sectors that will see the highest annual Gross Value Add growth rates in an AI scenario, comprising 4.8 per cent, 4.4 per cent and 4.3 per cent respectively by 2035.<sup>39</sup> Another potential source of Australian competitiveness is growing and hosting future leading companies for whom AI is at the heart of their business model.

However, Australia is not a leading nation in the implementation of automation and AI. Australia currently lags global leaders across the G20 in the adoption of automation: 50 per cent fewer Australian firms are actively investing in automation compared to firms in comparable economies.<sup>40</sup> To remedy this, the government and private sector would need to work together to build a more dynamic innovation ecosystem, specifically in regard to developing and implementing automation technologies.

Accelerating the deployment of AI across Australia would require organisations of all sizes—including a rising proportion of the small and medium-sized enterprises that contribute more than half of the country’s GDP—to explore new data-driven processes and business models that would benefit from ML.

One potential goal of the Responsible Innovation Organisation could therefore be to provide resources and information that would lower the barrier for Australian organisations to designing and deploying AI systems, while ensuring that such systems are protective of human rights.

## (b) Why does responsible AI matter?

The business case for a Responsible Innovation Organisation would likely rest on the ability to accelerate the adoption of technologies to maximise the economic and social benefits.

To realise the social benefits of AI, jurisdictions must decide the critical objectives, and work to minimise trade-offs between them. These may include, for example, ensuring that automation is implemented in an equitable way, addressing benefits, externalities and quality of life.

The business case should also take into account the costs of failing to act. The loss of social benefits can undermine the realisation of economic benefits. For example, Larry Fink's 2018 investor letter to Blackrock clients highlighted the importance of social responsibility and purpose — a potent signal given the more than six trillion US dollars that Blackrock manages.<sup>41</sup>

Beyond the opportunity costs, it is important to assess the costs to businesses and citizens when innovation has directly negative impacts. Historically, these are most visible when new technologies threaten health, safety or the environment. The increasing power of technologies such as AI, and their ability to be used across a wide range of industries and applications, makes it important to consider other, indirect negative impacts.

For example, understanding how individuals' privacy, autonomy and access to products and services are affected by the introduction of AI systems is critical to keeping both the economic and social costs low. In the public sector, the recent Centrelink 'robodebt' controversy highlights that the adoption of automated systems with efficiency as a primary goal can have unintended consequences, including greater consumer burdens, unjust outcomes, loss of trust, and cost of ex-post interventions.<sup>42</sup>

## (c) How would a Responsible Innovation Organisation directly add value to Australian organisations?

There are at least three ways in which the link between governance and economic impact may be considered.

First, by providing practical frameworks and other resources for assessing the potential discriminatory effects of AI systems, it could reduce the risk of the failed deployment of systems due to the emergence of unfair bias, which could have been anticipated and addressed with reasonable measures. In addition to the impact on affected individuals, liabilities linked to discrimination include brand damage, compensation for loss and significant investments in management time and focus.

Second, by creating a common benchmark for the design and deployment of AI systems across Australia, it might reward companies and entrepreneurs that provide automated services fairly and effectively. This creates benefits that flow not only to Australian citizens and users, but also to local innovators able to provide the services.

Third, raising the quality of AI deployment in Australia could create competitive benefits for local organisations intending to offer their products and services in other jurisdictions.

## Consultation questions

3. Would there be significant economic and/or social value for Australia in establishing a Responsible Innovation Organisation?
4. Under what circumstances would a Responsible Innovation Organisation add value to your organisation directly?
5. How should the business case for a Responsible Innovation Organisation be measured?



# 5 What should a Responsible Innovation Organisation look like?

Part 5 of this White Paper considers the key features of an organisation that may be given a leadership role in respect of AI in Australia, which this White Paper refers to as the Responsible Innovation Organisation. This White Paper sketches some possible components with a view to fostering discussion. It does not endorse any specific approach.

## 5.1 Vision, objectives and guiding principles

A Responsible Innovation Organisation would be unlike traditional oversight or compliance bodies in the Australian regulatory network.

The organisation will have to establish a normative framework for the development and deployment of AI. This approach would likely draw on international human rights law, such as the rights to a fair trial, privacy and non-discrimination. Using the human rights framework would enable the body to draw on established norms that have been accepted as central in any liberal democracy, such as the centrality of human dignity and personal autonomy, the importance of fairness and the need to carefully balance legitimate interests, as well as transparency, due process and accountability.

To support its proposal for a centralised body to govern the ethical use of data, the British Academy and Royal Society, for example, recommended that a body be established with the overarching objective of promoting 'human flourishing' to 'guide the development of systems of data governance'. A number of principles were then outlined to support this overarching principle, including the need to protect individual and collective rights and freedoms.<sup>43</sup>

It would be important that the Organisation's approach and governance structure are inclusive. Special attention should be made to include those who are particularly affected by new technologies, and most susceptible to the threats associated with them. This would include children and young people; people with disability; older people; people from culturally and linguistically diverse backgrounds and others.

The scope of the Organisation's work would also need to be considered. The Organisation could focus on the use of AI in government as well as the private sector, particularly where relying on AI to make a decision or determination has a consequence for the protection of an individual's human rights. The Organisation could provide an authoritative voice across government and industry.

The Organisation could also, for example, examine the deployment of AI, including its foundation: big data sets. This would involve considering ownership of big data and related concerns about the concentration of ownership among a small number of private sector entities. If we are to avoid giving 'strategic advantage' to those with data and computing resources, the collection of data and ownership of datasets will need to be democratic.<sup>44</sup>

## 5.2 Functions and powers

If a Responsible Innovation Organisation were created, there are a number of powers and functions it could possess, depending on its core objectives.

A Responsible Innovation Organisation could have a range of both coercive and non-coercive powers, such as:

- powers of inquiry and the ability to secure evidence, akin to the power of the Australian Competition and Consumer Commission (ACCC) to obtain information, documents and evidence<sup>45</sup>
- the ability to develop standards, codes of practice or regulations for data management and governance, with the power to monitor and evaluate compliance or power to impose fines for regulatory breaches, similar to the power of the Australian Privacy Commissioner to apply to a federal court for an order that an entity be fined for privacy breaches,<sup>46</sup> or the power of the UK Information Commissioner's Office to fine a data controller for breaching the right to privacy or commence criminal prosecution<sup>47</sup>
- a certification scheme for the human rights compliant development of AI, identifying AI products that have reached certain standards, as recently proposed by the Chief Scientist of Australia, Dr Alan Finkel<sup>48</sup>
- the ability to receive and adjudicate complaints from individuals adversely affected by a determination or decision using AI or assisted by the use of AI, including the powers to order remedies and redress
- a function to advise government on law and policy development regarding data governance and government procurement of AI systems
- evaluating data sets, promoting open data standards and collating data for model training and testing to minimise bias and discrimination
- building a repository of leading practices regarding inclusive stakeholder engagement
- writing and publishing professional codes of ethics for industry, drawing on pre-existing codes of practice and ethical standards in other jurisdictions and in the international sphere, such as those developed by the Institute for Electrical and Electronics Engineers (IEEE).

## 5.3 Partnership with government and industry

Responsible innovation in AI may require a combination of guidelines, regulations, and education with active participation from governments, businesses and industries, academia, civil society, and other interested stakeholders.<sup>49</sup> Each of these stakeholder groups has respective strengths and roles to play in helping create ethical AI.

For example, government can help set the agenda for public debate, monitor the safety and fairness of AI applications, and create or adapt regulatory frameworks to protect the public.<sup>50</sup> Industry can help create best practices for key aspects of the development of AI systems, such as the nature of the data used to train AI systems, the analytical techniques deployed, and how the results of AI are explained to the people using the AI systems.<sup>51</sup> Academia and civil society can help inform and educate the public, as well as ensure that proper mechanisms are put in place to safeguard the various and overlapping interests impacted by the use of AI.

A Responsible Innovation Organisation could facilitate this multi-stakeholder effort, particularly between government and industry. Such an Organisation could help establish a new governance model that addresses the different stakeholders' interlinked dynamics, the transnational and societal scope of their impact, and the political dimensions of AI technologies.<sup>52</sup> It could also bridge the gaps between the government and industry players so that the private sector's insight and influence can be effectively harnessed, while ensuring that citizens are protected from harm.<sup>53</sup>

## 5.4 Timeline: evaluation and monitoring

The long-term success of any Responsible Innovation Organisation would rest on its ability to interact constructively and productively with industry, civil society and government, and to bring these key stakeholders into a joined-up conversation. Reporting, evaluation and monitoring will be integral to the Organisation's operations from the outset.

At a minimum, the Organisation could provide annual public reports through the parliamentary reporting system. Good practice in evaluation is that it be periodic, strategic and instructive, measured against clear goals. Given the pace of technological development, it may be beneficial for the Organisation to report more frequently in order to update key stakeholders on its progress and respond to feedback in a timely way.



In addition, if a Responsible Innovation Organisation were established as a pilot, there may be merit in providing a review in the period following establishment. Reporting at the 12- and 18-month marks, for example, may allow for a robust assessment of whether the Organisation is achieving its objectives.

Finally, how the Organisation interacts with government and parliament should also be considered as part of the monitoring and evaluation framework. Government could, for example, be required to respond to any evaluation or monitoring reports, to ensure the proposed body is appropriately equipped to fulfil its mandate.

## Consultation questions

6. If Australia had a Responsible Innovation Organisation:

- (a) What should be its overarching vision and core aims?
- (b) What powers and functions should it have?
- (c) How should it be structured?
- (d) What internal and external expertise should it have at its disposal?
- (e) How should it interact with other bodies with similar responsibilities?
- (f) How should its activities be resourced? Would it be jointly funded by government and industry? How would its independence be secured?
- (g) How should it be evaluated and monitored? How should it report its activities?



## 6 Consultation questions

For ease of reference, the questions posed in this White Paper are listed below:

1. What should be the main goals of government regulation in the area of artificial intelligence?
2. Considering how artificial intelligence is currently regulated and influenced in Australia:
  - (a) What existing bodies play an important role in this area?
  - (b) What are the gaps in the current regulatory system?
3. Would there be significant economic and/or social value for Australia in establishing a Responsible Innovation Organisation?
4. Under what circumstances would a Responsible Innovation Organisation add value to your organisation directly?
5. How should the business case for a Responsible Innovation Organisation be measured?
6. If Australia had a Responsible Innovation Organisation:
  - (a) What should be its overarching vision and core aims?
  - (b) What powers and functions should it have?
  - (c) How should it be structured?
  - (d) What internal and external expertise should it have at its disposal?
  - (e) How should it interact with other bodies with similar responsibilities?
  - (f) How should its activities be resourced? Would it be jointly funded by government and industry? How would its independence be secured?
  - (g) How should it be evaluated and monitored? How should it report its activities?

## 7 Consultation process and outcomes

This White Paper supports a collaborative consultation that has been initiated by the Commission and World Economic Forum.

There are two ways to contribute to the White Paper consultation process. First, written submissions to the consultation questions in the White Paper can be emailed to [tech@humanrights.gov.au](mailto:tech@humanrights.gov.au). Submissions are due by 5pm on 8 March 2019.

This consultation will form part of the inquiry for the Commission's human rights and technology project. Details of the Commission's Project and the consultation, including public submissions received, are available at <https://tech.humanrights.gov.au>.

In addition, the Commission and the World Economic Forum plan to consult with relevant experts on the issues raised in this White Paper.

The White Paper consultation process also aims to benefit from other collaborative approaches taking place globally. For example, the World Economic Forum and the UK Government have agreed to work together to identify best practices for the creation and procurement of AI used by the government.<sup>54</sup>

If the consultation process reveals strong support for a Responsible Innovation Organisation, the Commission and the World Economic Forum may consider further steps, which could include a blueprint for action. Drawing on material gleaned from the consultation on the White Paper, this blueprint could outline the core elements of the Responsible Innovation Organisation and how it could fit within Australia's governance and regulatory framework. The blueprint could also consider how an Australian Responsible Innovation Organisation could fit into the international network of AI governance being supported by the World Economic Forum.



- 1 For example: Robert Pearl, 'Artificial Intelligence in Healthcare: Separating Reality from Hype,' *Forbes* (online), 13 March 2018 <[www.forbes.com/sites/sobertpearl/2018/03/13/artificial-intelligence-in-healthcare/#c2d77d21d750](http://www.forbes.com/sites/sobertpearl/2018/03/13/artificial-intelligence-in-healthcare/#c2d77d21d750)>.
- 2 Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, 'Machine Bias: There's Software Used Across The Country to Predict Future Criminals. And It's Biased against Blacks,' *ProPublica* (online) 23 May 2016 <[www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing](http://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing)>; Osonde Osoba and William Wesler IV, *An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence* (RAND Corporation, 2017).
- 3 Law and Safety Committee, Parliament of New South Wales, *The Adequacy of Youth Diversionary Programs in New South Wales* (Report 2/56, September 2018), 2.108-2.110; Vicki Sentas and Camilla Pandolfini, *Policing Young People in NSW: A Study of the Suspect Targeting Management Plan* (Youth Justice Coalition NSW, 2017), September 2018.
- 4 Stuart Lauchlan, 'World Economic Forum 2018 – Why Trust Has to Be Valued Higher than Growth in the 4IR,' on *Diginomica Government* (23 January 2018) <[www.government.diginomica.com/2018/01/23/world-economic-forum-2018-trust-valued-higher-growth-4ir/](http://www.government.diginomica.com/2018/01/23/world-economic-forum-2018-trust-valued-higher-growth-4ir/)>.
- 5 Ibid.
- 6 Arjun Kharpal, 'Mark Zuckerberg's Testimony: Here Are the Key Points You Need to Know', *CNBC* (online), 11 April 2018 [www.cnbc.com/2018/04/11/facebook-ceo-mark-zuckerberg-testimony-key-points.html](http://www.cnbc.com/2018/04/11/facebook-ceo-mark-zuckerberg-testimony-key-points.html)>.
- 7 Brian Fung, 'We Have to Admit When the Free Market Is Not Working': Apple Chief's Privacy Call', *Sydney Morning Herald* (online), 20 November 2018 <<https://www.smh.com.au/business/companies/we-have-to-admit-when-the-free-market-is-not-working-apple-chief-s-privacy-call-20181120-p50h2n.html>>.
- 8 Digital Forensic Research Lab, 'We Need 21st Century Responses: Secretary Albright Speaks at #DisinfoWeek' on *Medium* (30 June 2017) <[www.medium.com/dfrlab/we-need-21st-century-responses-6b7eed6750a4](http://www.medium.com/dfrlab/we-need-21st-century-responses-6b7eed6750a4)>.
- 9 University of Adelaide, *Australian Institute for Machine Learning: Catching the Wave of the Next Industrial Revolution* (2017) <[www.adelaide.edu.au/aiml/docs/australian-institute-for-machine-learning-brochure.pdf](http://www.adelaide.edu.au/aiml/docs/australian-institute-for-machine-learning-brochure.pdf)>.
- 10 Ibid 3.
- 11 Chris Griffith, 'Australian Bosses Embrace Artificial Intelligence,' *The Australian* (online), 30 January 2018 <[www.theaustralian.com.au/business/technology/australian-bosses-embrace-artificial-intelligence/news-story/336a20c3a1df43d21947d57be780e7d1](http://www.theaustralian.com.au/business/technology/australian-bosses-embrace-artificial-intelligence/news-story/336a20c3a1df43d21947d57be780e7d1)>.
- 12 Executive Office of the President National Science and Technology Council Committee on Technology, *Preparing for the Future of Artificial Intelligence* (2017) <[www.obamawhitehouse.archives.gov/sites/default/files/whitehouse\\_files/microsites/ostp/nstc/preparing\\_for\\_the\\_future\\_of\\_ai.pdf](http://www.obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/nstc/preparing_for_the_future_of_ai.pdf)>.
- 13 Microsoft, *The Future Computed: Artificial Intelligence and Its Role in Society* (Microsoft, 2018).
- 14 Australian Human Rights Commission, *Human Rights and Technology* (7 February 2018) <<https://www.humanrights.gov.au/our-work/rights-and-freedoms/projects/human-rights-and-technology>>.
- 15 IEEE, *Ethically Aligned Design: A Vision for Prioritising Human Well-Being with Autonomous and Intelligent Systems* (IEEE, 2nd ed, 2017) 6.
- 16 Nicholas Davis, World Economic Forum, Ch. 7: The Future Relationship Between Technology and Inequality, in *How Unequal? Insights on Inequality* (Committee for Economic Development of Australia, April 2018).
- 17 Australian Human Rights Commission, *Legislation*, <[www.humanrights.gov.au/our-work/legal/legislation](http://www.humanrights.gov.au/our-work/legal/legislation)>.
- 18 Australian Human Rights Commission, *Business and Human Rights*, <<https://www.humanrights.gov.au/employers/business-and-human-rights>>.
- 19 Microsoft, *The Future Computed: Artificial Intelligence and Its Role in Society* (Microsoft, 2018) 58.
- 20 Ibid 58-59.
- 21 Nicholas Davis, World Economic Forum, Ch. 7: The Future Relationship Between Technology and Inequality, in *How Unequal? Insights on Inequality* (Committee for Economic Development of Australia, April 2018).
- 22 Microsoft, *The Future Computed: Artificial Intelligence and Its Role in Society* (Microsoft, 2018) 59.
- 23 Nicholas Davis, World Economic Forum, Ch. 7: The Future Relationship Between Technology and Inequality, in *How Unequal? Insights on Inequality* (Committee for Economic Development of Australia, April 2018).
- 24 Christian Ehl, 'Data – the Fuel for Artificial Intelligence' on *Medium* (14 January 2018) <<https://medium.com/@cehl/data-the-fuel-for-artificial-intelligence-ed90bf141372>>; see also Bernard Marr, 'Why AI Would Be Nothing Without Big Data,' *Forbes* (online), 9 June 2017 <[www.forbes.com/sites/bernardmarr/2017/06/09/why-ai-would-be-nothing-without-big-data/#6fe8a7a94f6d](http://www.forbes.com/sites/bernardmarr/2017/06/09/why-ai-would-be-nothing-without-big-data/#6fe8a7a94f6d)>.
- 25 Justin Sherman, 'AI Innovation: Security and Privacy Challenges' on *Medium* (22 January 2018) <[www.medium.com/swlh/ai-innovation-security-and-privacy-challenges-84c0200b1bae](http://www.medium.com/swlh/ai-innovation-security-and-privacy-challenges-84c0200b1bae)>.
- 26 Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (St. Martin's Press, 2018).
- 27 World Economic Forum, *How to Prevent Discriminatory Outcomes in Machine Learning* (2018) <[www3.weforum.org/docs/wef\\_40065\\_white\\_paper\\_how\\_to\\_prevent\\_discriminatory\\_outcomes\\_in\\_machine\\_learning.pdf](http://www3.weforum.org/docs/wef_40065_white_paper_how_to_prevent_discriminatory_outcomes_in_machine_learning.pdf)>.
- 28 Edelman, *Edelman Trust Barometer 2018*, <[www.slideshare.net/edelmaninsights/2018-edelman-trust-barometer-australia-results](http://www.slideshare.net/edelmaninsights/2018-edelman-trust-barometer-australia-results)>.
- 29 See, for example, Consumer Policy Research Centre, Submission 42 to the Australian Human Rights Commission, Human Rights and Technology Issues Paper (2 October 2018).
- 30 See, for example, Nicolas Suzor, Kim Weatherall, Angela Daly, Ariadne Vromen, Monique Mann, Submission 91 to the Australian Human Rights and Technology Issues Paper (October 2018).
- 31 University of Technology, Submission 103 to the Australian Human Rights and Technology Issues Paper (October 2018), 95.
- 32 Australian Competition and Consumer Commission Digital Platforms Inquiry Preliminary Report (10 December 2018) <<https://www.accc.gov.au/system/files/accc%20digital%20platforms%20inquiry%20-%20preliminary%20report.pdf>>.
- 33 HM Treasury, UK Government, *Autumn Budget 2018* (2017) [5.3].
- 34 UK Government, *Consultation on the Centre for Data and Innovation* (13 June 2018) <<https://www.gov.uk/government/consultations/consultation-on-the-centre-for-data-ethics-and-innovation>>.
- 35 UK Government, *World-Leading Expert Demis Hassabis to Advise New Government Office for Artificial Intelligence* (26 June 2018) <[www.gov.uk/government/news/world-leading-expert-demis-hassabis-to-advise-new-government-office-for-artificial-intelligence](http://www.gov.uk/government/news/world-leading-expert-demis-hassabis-to-advise-new-government-office-for-artificial-intelligence)>. The other two organisations are the AI Council and the Office for AI.
- 36 AlphaBeta, *The Automation Advantage* (2017) <[www.alphabeta.com/wp-content/uploads/2017/08/the-automation-advantage.pdf](http://www.alphabeta.com/wp-content/uploads/2017/08/the-automation-advantage.pdf)>.
- 37 Mark Purdy and Paul Daugherty, *Why Artificial Intelligence Is the Future of Growth* (Accenture, 2016).
- 38 The Commonwealth of Australia, 'Budget Paper No. 2 – Budget Measures' *Budget 2018-19* (8 May 2018) <[www.budget.gov.au/2018-19/content/bp2/download/bp2\\_combined.pdf](http://www.budget.gov.au/2018-19/content/bp2/download/bp2_combined.pdf)>.
- 39 Mark Purdy and Paul Daugherty, *Why Artificial Intelligence Is the Future of Growth* (Accenture, 2016).
- 40 AlphaBeta, *The Automation Advantage* (2017) <<http://www.alphabeta.com/wp-content/uploads/2017/08/the-automation-advantage.pdf>>.
- 41 Larry Fink, *A Sense of Purpose* [www.blackrock.com/corporate/investor-relations/larry-fink-ceo-letter](http://www.blackrock.com/corporate/investor-relations/larry-fink-ceo-letter)>.
- 42 Senate Community Affairs References Committee, *Design, Scope, Cost-Benefit Analysis, Contracts Awarded and Implementation Associated with the Better Management of the Social Welfare System Initiative* Parliament of Australia, <[www.aph.gov.au/parliamentary\\_business/committees/senate/community\\_affairs/socialwelfaresystem](http://www.aph.gov.au/parliamentary_business/committees/senate/community_affairs/socialwelfaresystem)>.
- 43 British Academy for the Humanities and Social Sciences and the Royal Society, *Data Management and Use: Governance in the 21st Century*, <[www.royalsociety.org/topics-policy/projects/data-governance/](http://www.royalsociety.org/topics-policy/projects/data-governance/)>.
- 44 AI Now, *The AI Now Report: The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term* (2016) <[https://ainowinstitute.org/ai\\_now\\_2016\\_report.pdf](https://ainowinstitute.org/ai_now_2016_report.pdf)>.
- 45 The ACCC has compulsory information-gathering powers under s 155 of the *Competition and Consumer Act 2010* (Cth).

- 
- 46 Section 80W of the *Privacy Act 1988* (Cth) empowers the Privacy Commissioner to apply to the Federal Court or the Federal Circuit Court for an order that an entity pay the Commonwealth a pecuniary penalty where that entity has contravened a civil penalty provision, such as a serious or repeated interference with privacy.
- 47 See Information Commissioner's Office, *Taking Action-Data Protection*, <<https://ico.org.uk/about-the-ico/what-we-do/taking-action-data-protection/>>.
- 48 See Yolandra Redrup, 'Chief Scientist Alan Finkel Calls for Artificial Intelligence Regulation,' *The Australian Financial Review* (online), 14 November 2017 <<https://www.afr.com/technology/chief-scientist-alan-finkel-calls-for-artificial-intelligence-regulation-20171114-gzkt17>>.
- 49 Executive Office of the President National Science and Technology Council Committee on Technology, *Preparing for the Future of Artificial Intelligence* (2016) <[https://obamawhitehouse.archives.gov/sites/default/files/whitehouse\\_files/microsites/ostp/nstc/preparing\\_for\\_the\\_future\\_of\\_ai.pdf](https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/nstc/preparing_for_the_future_of_ai.pdf)>. ("AI can be a major driver of economic growth and social progress, if industry, civil society, government, and the public work together to support development of the technology, with thoughtful attention to its potential and to managing its risks."); see also Microsoft, *The Future Computed: Artificial Intelligence and Its Role in Society* (Microsoft, 2018) 17.
- 50 Executive Office of the President National Science and Technology Council Committee on Technology, *Preparing for the Future of Artificial Intelligence* (2016) 39.
- 51 Ibid.
- 52 Nima Elmi and Nicholas Davis, *How Governance Is Changing in the 4IR* (18 January 2018) <[www.weforum.org/agenda/2018/01/agile-governance-changing-4ir-public-private-emerging-technologies](http://www.weforum.org/agenda/2018/01/agile-governance-changing-4ir-public-private-emerging-technologies)>.
- 53 Ibid.
- 54 Prime Minister's Office and The Rt Hon Theresa May MP, *PM's Speech at Davos 2018: 25 January* (25 January 2018) <<https://www.gov.uk/government/speeches/pms-speech-at-davos-2018-25-january>>.

