



The Australian Industry Group
51 Walker Street
North Sydney NSW 2060
PO Box 289
North Sydney NSW 2059
Australia
ABN 76 369 958 788

7 June 2019

Artificial Intelligence
Strategic Policy Division
Commonwealth Department of Industry, Innovation and Science

Email: artificial.intelligence@industry.gov.au

Dear Sir/Madam

AUSTRALIAN GOVERNMENT DISCUSSION PAPER ON AUSTRALIAN AI ETHICS FRAMEWORK

The Australian Industry Group (Ai Group) welcomes the opportunity to comment to the consultation by the Commonwealth Department of Industry, Innovation and Science (Department) on the Discussion Paper about Australia's Artificial Intelligence (AI) ethics framework.

1. Introduction

Ai Group's membership comes from a broad range of industries and includes businesses of all sizes. Rapidly advancing technologies including AI are producing waves of wider innovation across the economy as businesses and individuals build new social practices and business models upon them. Ai Group's members are grappling with these changes in different ways and with different levels of readiness and capability. The collective impact of these changes is part of the Fourth Industrial Revolution.

There is growing discussion among our members of the impact of the Fourth Industrial Revolution on their businesses and workforce. Like previous advances, new technology is enabling improvements in speed to market, quality and cost effectiveness. But the latest revolution also presages more flexibility and individualisation – a customer-oriented approach that provides a social value.

The history of previous economic disruptions suggests that if factors related to social inequality are not appropriately addressed, there is a risk that large sections of the community can be left behind. Public policy such as around inclusiveness and education impacts on the social divide as well as the digital divide. In recent decades, we have seen the effects of poor management in other countries where these divisions are growing.

According to the 2018 Australian Digital Inclusion Index report, the socio-demographic groups that were most digitally excluded in Australia in 2018 included: people in low income households; mobile only users; people aged over 65; people who did not complete secondary school; and people with a disability.¹ While there appears to be improvement in some areas (such as digital access, digital ability, value of internet services, and Indigenous inclusion), there still remains a gap between the digitally included and excluded.

We, as a community, need to re-examine how change is managed. We should neither hold back the tide nor be indifferent to change. However, the ultimate benefits of technological change do not erase the transitional costs to disrupted industries and displaced workers.

Businesses have responsibilities to recognise and respond to transitional costs, not just the benefits of an exciting new direction. And some are already demonstrating leadership in this area.

Overall, industry recognises the importance of the work of various stakeholders including the Australian Government, which can help contribute to the discussion on closing the social digital divide including

¹ Roy Morgan Research, "Measuring Australia's digital divide: Australian Digital Inclusion Index 2018" (Report, October 2018), pp. 5-6.



The Australian Industry Group
51 Walker Street
North Sydney NSW 2060
Australia
ABN 76 369 958 788

AI. In this regard, feedback from members about the Department's Discussion Paper has been generally positive.

We note that the Discussion Paper includes a summary of existing frameworks, principles and guidelines on AI ethics, data governance, case studies on automated decisions, predicting human behaviour and AI in practice, and a proposed ethics framework. These are understandably couched in terms of ethical AI given the subject of this consultation, which we provide further comment on in the subsequent sections.

In the remainder of this submission, we provide Ai Group's views on specific issues and we provide a summary of individual input from a number of Ai Group members.

While not directly in response to questions raised in this Discussion Paper, it is also important to appreciate the broader context in a discussion about AI, which has been alluded to in the Paper. We have included general comments from Ai Group about the broader context of AI in Appendix A and other general comments from one of our members about AI directly relevant to this Discussion Paper in Appendix B.

In addition to this submission, we would also welcome the opportunity to work with the Australian Government to bring together a range of industries who may be interested in this Discussion Paper to be consulted with further.

2. Proposed core principles on ethical AI

Questions and proposed core principles raised in Discussion Paper:

1. *Are the principles put forward in the discussion paper the right ones? Is anything missing?*
2. *Do the principles put forward in the discussion paper sufficiently reflect the values of the Australian public?*
7. *Are there additional ethical issues related to AI that have not been raised in the discussion paper? What are they and why are they important?*

Core principles for AI

1. Generates net-benefits. *The AI system must generate benefits for people that are greater than the costs.*

2. Do no harm. *Civilian AI systems must not be designed to harm or deceive people and should be implemented in ways that minimise any negative outcomes.*

3. Regulatory and legal compliance. *The AI system must comply with all relevant international, Australian Local, State/Territory and Federal government obligations, regulations and laws.*

4. Privacy protection. *Any system, including AI systems, must ensure people's private data is protected and kept confidential plus prevent data breaches which could cause reputational, psychological, financial, professional or other types of harm.*

5. Fairness. *The development or use of the AI system must not result in unfair discrimination against individuals, communities or groups. This requires particular attention to ensure the "training data" is free from bias or*

6. Transparency & Explainability. *People must be informed when an algorithm is being used that impacts them and they should be provided with information about what*



The Australian Industry Group
51 Walker Street
North Sydney NSW 2060
Australia
ABN 76 369 958 788

characteristics which may cause the algorithm to behave unfairly.

information the algorithm uses to make decisions.

7. Contestability. *When an algorithm impacts a person there must be an efficient process to allow that person to challenge the use or output of the algorithm.*

8. Accountability. *People and organisations responsible for the creation and implementation of AI algorithms should be identifiable and accountable for the impacts of that algorithm, even if the impacts are unintended.*

From our reading of the Discussion Paper, we note that the proposed eight core principles are to be “used as ethical framework to guide organisations in the use or development of AI systems”. In principle, we support a framework that supports the public policy objective of enabling digital investment and competition that benefits industry and the community in Australia in the long term. Overall, we have received generally positive feedback about the AI ethics framework including principles.

One member specifically commented that *the core principles are sound and appropriate for use of AI in civilian application, noting that the discussion of military use of AI is beyond the scope of the paper. For public institutions and civilian AI applications, they also support the Universal Guidelines for Artificial Intelligence as set out in section 2.3.7 of the Discussion Paper as a starting point for developing a framework of individual rights relative to AI technologies.*

2.1 Comments about specific principles

Below are collated comments that we have received from members about specific aspects of the proposed core principles, which we would like to bring to your attention and consideration:

- Principle 1 (generates net-benefits):
 - *This principle is well intended, but its framing may need to change for it to be workable, meaningful or effective. Measuring and comparing the full social costs and benefits that may be associated with an AI system is a task that may be impossible. The judgements involved may not be appropriate for individual AI system operators to make. This principle should be framed as an overall outcome from Australia’s use of AI, with responsibilities for industry, government and other stakeholders to contribute towards it.*
 - *This principle is almost impossible to define. It is not clear who are the “people”, and how you measure this net benefit. According to whom? This can lead down a Machiavellian path of ends justifying means. How does one weigh the benefits of saving lives versus stopping environmental damage for instance? Does the AI have a weighting scale for all the potential scenarios? It should be spelled out that the algorithms for deciding the benefits versus the costs need to be open and accountable.*
- Principle 2 (do no harm):
 - *This should, perhaps, refer to the people deploying the AI rather than the AI itself. Considering AI as a “tool” vs “agent” has a bearing on this and this brings challenges. It is also not clear how to measure and enforce this.*
 - *Similar to “benefits”, harm is almost impossible to define. “Harm” can be in the eye of the beholder, or the harmed. And when minimising “any negative outcomes”, what is the obligation on developers to disclose their decision-process for mitigating or minimising negatives, or even that the negatives exist? Is a disclaimer required such as for pharmaceutical companies advertising in the US?*

- Principle 3 (regulatory and legal compliance):
 - *This principle seems appropriate, but would benefit from exposition on what we are seeking to regulate? We are already regulating safety in some areas such as automation and autonomous aircraft. Does this include the potential use of licencing for people deploying and using AI? Behaviours should be regulated and these can include aspects of ethics.*
 - *This principle should clarify that the “international obligations, regulations and laws” that should be complied with are those to which Australia is a signatory. There are many international regimes that Australia is not signatory to, and they should not be binding on Australian AI developers, other than to the extent that their products may be expected to operate in or with technologies from those other jurisdictions.*
- Principle 4 (privacy protection):
 - *Does the wording of the explanation give sufficient notice that AI information is subject to the legal and regulatory requirements of holding personal data, including protections and/or compulsions? Perhaps the intention is that when AI extracts sensitive pieces of information from data, this information should be protected. Is this an AI issue or an issue with the environment in which the AI operates and the environment limitations? Again, it appears this is seeking to attribute AI faculties that should reside with the humans deploying and using AI.*
- Principle 5 (fairness):
 - *Who decides what is “fair”? This seems to attribute to AI faculties that should perhaps reside with the humans deploying and using AI.*
- Principle 6 (transparency and explainability):
 - *With respect to the formulation of this principle, it would be good to clarify what it means for an algorithm to impact someone such that it would trigger a requirement for disclosure. While this may be desirable for decisions that have a substantial impact on an individual (e.g. related to health and finance), in many everyday cases (e.g. automatic adjustments on their camera phone exposure), it may be extraneous to the user whether an algorithm has been used or not.*
 - *What do we want to know? The details of the algorithm? The details of the dataset? The requirements to which it was designed? How it is used within a larger system? We have examples from around the world where such an idea has been legislated with good intent (GDPR, New York City) but has been effectively impossible to comply with.*
 - *In many cases, using current techniques, we may be unable to produce a satisfactory explanation at all, the explanation could divulge IP, or using a directly interpretable method in the first place may significantly increase the error rate, leading to suboptimal outcomes on average.*
 - *Trust can be established in many ways, and explanations are only one. For example, one could imagine a government standard self-driving car “test gauntlet” with a series of scenarios and expected performance outcomes. It obviously would not have complete coverage (impossible), but may establish a basic level of AI competence through test alone, similar to how we qualify pilots. All this is currently possible without having algorithmic explainability in terms of human concepts. Even so-called “black boxes” can be probed and tested. Taking a balanced view, where possible we should strive for more interpretability and regulation could and should spur innovation in this area. It is absolutely necessary for explainable AI techniques to be encouraged, developed and widely utilised.*
 - *This aspect should be considered much more carefully than it has been in the Discussion Paper. We do not understand how people make particular decisions and sometimes why they behave in the way they do. Having access to information about brain activity may not be useful for*

assessment of human behaviour in relation to ethics. Perhaps we should not attempt this with AI and treat it as a black box as we do with humans. What is crucial here is to define the behaviours we expect from AI and AI-human interaction and certify them independent of how these are implemented.

- Principle 7 (contestability):
 - *This principle concerns the law. It might be useful to define what an algorithm actually is (a procedure). The challenge process must be accessible to affected parties, as well as efficient.*
- Principle 8 (accountability):
 - *This principle may not be realistic as currently worded. The current drafting suggests any person or organisation involved in the creation of an open source model or API that ends up being used by unrelated parties in an AI system should be identifiable and accountable for the impacts, even if they were unintended. However, there is no way they would be able to predict or even find out all the ways in which models they had created were being used if they had been given away open source. While they can take steps to be responsible (e.g. provide guidance on how the model was created or target uses, and warnings of things they can imagine it would not be good for), it is not something within an open source creator's control or even visibility.*
 - *This principle should focus on people using and deploying AI rather than a developer. A scientist developing a new type of neural network should not be liable for how other people use it. Nor should a knife manufacturer be liable for death by stabbing with a knife. This concerns the law more than the ethics as it has many challenges.² As well as accountability, there should be humans either in or on the loop for decisions which may result in humans being negatively impacted, or at other key points that may require review prior to decisive action.*
- Multiple principles:
 - *Across the various principles, it might be useful to insert a bit of nuance with regard to open source technologies and off-the-shelf solutions where developers may not be interacting directly with third parties applying these technologies in particular use cases or their eventual end users.*
 - *While there is a lot of positive work already being done to address Principles 2, 5 and 6, the report could flesh this out in more detail. To assist, we can connect the Department with a member who can provide concrete examples of what they are doing.*
 - *More generally, some of the principles could use some caveats. For example, Principle 7 could be applicable when an algorithm impacts a person in a meaningful fashion. Otherwise, it would risk being open to frivolous impacts, leading to costly and untenable processes, and negatively impacting on innovation and limited benefit to consumers.*
 - *It would be helpful if the eight core principles more clearly articulated that regulatory efforts should appropriately balance the protection of legitimate commercial interests with the protection of individual and public rights and accountability. For example, any regulatory review of an allegedly discriminatory AI algorithm by a regulator should be carried out in a manner that would not destroy the commercial value of the algorithm by publicly revealing it to competitors. Additionally, it should be more clearly emphasised as a matter of principle that regulatory efforts should provide market advantages to responsible firms and help to drive out rogue elements seeking to operate outside of regulatory frameworks.*
 - *The core principles require a preamble. While the principles themselves are a good start, they would benefit from an introductory paragraph that gives more context for how organisations should approach the principles, how they were derived, and expectations on developers of AI*

² S Chopra and LF White, *A Legal Theory for Autonomous Artificial Agents* (The University of Michigan Press, 2014).

- technologies. This includes the need for balancing all the principles, as they can be in competition, and the need for open and accountable mechanisms for arriving at algorithms and determinations.
- It may well be better to focus on core principles for the deployment of AI, rather than core principles for AI.
 - The principles should be reworded so that the headings are able to be read like a list. Currently they are grammatically inconsistent.
- Additional principle:
 - The Discussion Paper could have explored the principle of “non-instrumentalism” (p. 53) in greater depth and perhaps it should be considered for inclusion as a core principle for AI. It is noted that on p. 53, The Ethics Centre advanced the view that “AI technologies should keep the principle of “non-instrumentalism” in mind when designing technology [38]. This effectively means that humans should not merely become another part of the machine – the machine should serve people, not the other way around”. Where there is close civilian human interaction with machine interfaces/AI, it would be principled to consider how humanity in general is better served by that, and what safeguards would be needed to protect individual rights.

2.2 Comment about values

In response to Question 2 in the Discussion Paper, one member provided the following response:

“Yes. The Core Principles for AI are sufficiently aligned to the values and legitimate expectations of the Australian public, within a mature middle-power Western democracy.

“However “Values” is a culturally loaded term, as each Australian is imbued with a different set of values. Discussion of “what are Australian values?” prompts the expression of robust and divergent views.

“The Department of Immigration and Border Protection lays out for visa applicants that:

‘Australian society values respect for the freedom and dignity of the individual, freedom of religion, commitment to the rule of law, Parliamentary democracy, equality of men and women and a spirit of egalitarianism that embraces mutual respect, tolerance, fair play and compassion for those in need and pursuit of the public good. Australian society values equality of opportunity for individuals, regardless of their race, religion or ethnic background... the English language, as the national language, is an important unifying element of Australian society.’

“In an attempt to understand how accurately the Department’s view of Australian values reflected reality, the University of Western Australia undertook a [“Values Project”](#), collecting data from more than 7,000 adults aged 18 to 75. In 2018 it reported that: ‘...the most important value to Australians was BENEVOLENCE, followed by SECURITY and then UNIVERSALISM-SOCIETAL.’

“Descriptions of what each means are available at the above link.”³

“A summary list of the Core Principles for AI would lead organisations to produce AI that:

1. Generates benefits
2. Does no harm

³ Link: <https://www.thevaluesproject.com/blog/what-are-australian-values/>

3. *Complies with relevant laws*
4. *Protects privacy*
5. *Is fair*
6. *Is transparent*
7. *Is explainable*
8. *Is contestable*
9. *Is accountable*

“Regarding “Australian values” the best that could be said of the Principles is that they do not go against what most Australian’s would reasonably be expected to hold as values. Although it is difficult to accurately gauge just exactly what Australian values are.”

2.3 Comment about additional ethical issues

In response to Question 7 in the Discussion Paper, one member provided the following response:

“On a fundamental level, the development of AI technologies has potentially profound social and political implications. There are stark choices to be made about the path Australia takes. The discussion paper could have further explored the risks of AI technologies being utilised by criminal or ill-motivated actors and why responsibly wielded AI technologies may be needed to counter or minimise those risks to uphold individual rights, the rule of law and protect national interests.

“In making arguments for responsible regulation of AI technologies to ensure that individual rights are upheld, it may be useful for contrasts to be made with countries that are choosing a more controlling path with much less emphasis on transparency, accountability, contestability and individual rights. For example, consider the mandated society-wide social credit system to be implemented in China in 2020 which makes considerable use of AI technologies. This example, may provide a useful contrast to support arguments for a different approach to be taken by mature, middle power Western democracies like Australia. Consider Rachel Botsman “Big data meets Big Brother as China moves to rate its citizens”.⁴

“Sound regulation of AI technologies should uphold Western democratic values, individual rights the rule of law and the Australian national interest. For military applications, additional work translating to defence purposes is clearly needed and would be well undertaken by Ai Group or Defence Science and Technology Group (or partnership thereof).

Sentience⁵

“In the same way that many nations have ethical boundaries around medical experimentation (such as mixing human DNA with animal DNA), nations should be extremely wary of producing AI technology to mimic humans in the creation of consciousness, creating potential unnecessary moral dilemmas and confusion between when a machine is being “switched off” and when it is being “killed”. If a technology becomes “aware” and is able to learn how to over-ride attempts to shut it down, there may be the chance it can negate or over-ride the humans in the loop, or safeguards against its control. Injecting moral arguments about the machine’s sentience will further confuse arguments at potentially critical moments.

⁴ Link: http://teomeuk.s3.amazonaws.com/wp-content/uploads/2017/11/05/Wired_Big_data_meets_Big_Brother_as_China_moves_to_rate_its_citizens.pdf.

⁵ Link: <https://theconversation.com/careful-how-you-treat-todays-ai-it-might-take-revenge-in-the-future-112611>.

International alignment

“Australia’s code should be aligned as much as possible with international standards. There are a number of movements and organisations such as those outlined in the discussion paper that are looking at the ethical aspects of AI. Australia should have a seat at the most prominent, or at least have a considered and informed presence in order to ensure we are across the latest developments, issues and solutions in the international community. There should also be international agreement and/or cooperation on the development of AI to ensure rogue operators are isolated and countered if required. Aligning our principles with like-minded larger jurisdictions will give them more power and improve market access for Australian technology. Harmonisation of standards and regulations is critical to achieving beneficial, responsible, and secure deployment of AI technologies globally.”

Civil versus military

“It is imperative to understand the dual-use nature of many technologies through history, and AI is no different. Therefore any discussion of standards for civil-use AI must necessarily also consider the potential for military use of those same technologies.”

3. Practical application of an ethical AI framework

Questions raised in Discussion Paper:

3. *As an organisation, if you designed or implemented an AI system based on these principles, would this meet the needs of your customers and/or suppliers? What other principles might be required to meet the needs of your customers and/or suppliers?*
4. *Would the proposed tools enable you or your organisation to implement the core principles for ethical AI?*
5. *What other tools or support mechanisms would you need to be able to implement principles for ethical AI?*
6. *Are there already best-practice models that you know of in related fields that can serve as a template to follow in the practical application of ethical AI?*

The Discussion Paper includes a section at the end on how its proposed ethics framework (including principles) can be applied in practice.

Ai Group’s view is that consideration needs to be given to the current workplace relations context:

- Generally, the use of AI in appropriate employment decision-making can provide for more accurate, fairer and efficient decisions than human judgement based on intermittent managerial observation. The use of AI in the employment context should not be discouraged but supported with the appropriate safeguards.
- Principle 7 should, when applied in practice, account for existing frameworks already in place – including the Fair Work Act and its comprehensive protections for employees, such as in relation to termination of employment and the already complex anti-discrimination legislation framework that exists at both Federal and State levels. Employers already face complex and, in some instances, competing regulation impacting employment-related decisions. Otherwise, there could be a conflict with Principle 3 (regulatory and legal compliance).
- Ai Group is participating in the Victorian Government’s inquiry into the Victorian on-demand workforce which will have a focus on how employers are utilising digital technologies in employment. Ai Group has lodged a preliminary submission describing the varied purposes of



The Australian Industry Group
51 Walker Street
North Sydney NSW 2060
Australia
ABN 76 369 958 788

digital technologies in businesses around productivity, labour allocation and services to consumers. In this inquiry, Ai Group has argued that the vast majority of labour and business arrangements rely on the current framework including the flexibility afforded by the common law tests in determining whether a worker is an employee or an independent contractor. Similarly, in determining an approach for the appropriate use of AI technologies in the context of labour allocation or employment-related decisions, regard should be had to the protections already in place in Australia's current workplace relations framework.

In addition, we offer the following collated comments from members of Ai Group:

- The Discussion Paper acknowledges the relevance of various forms of existing legislation and regulation associated with data and privacy, given that data is a key component of AI. In this regard, we have received anecdotal feedback from businesses, especially SMEs, about the regulatory costs arising from new regulations such as the Australian Notifiable Data Breach Scheme which commenced in February 2018. Other data and privacy regulations such as the EU General Data Protection Regulation (GDPR) (which commenced in May 2018) and Consumer Data Right (which is being developed for specific sectors) will also present an additional regulatory burden and challenge for a range of businesses. Government support for businesses to meet these obligations may be required, especially if it incorporates any new amendments associated with an AI ethical framework.
- *A common issue raised around AI is that it can create the risk of bias and discrimination. Indeed, the Discussion Paper includes Principle 5 to address this. A non-regulatory response could be to improve the education and training system. For example, the response could be to introduce a unit in the education curriculum for developers and implementers of AI about tackling the potential issue of bias and discrimination arising from AI. In fact, the education system is also going through its own transition in responding to the pace of technological change and meeting the demands of industry and the public.*
- *As one member suggests, the real issue is how to bring all the stakeholders together to develop a framework for encouraging greater awareness of AI bias and related issues along the lines that the Discussion Paper suggested and then take real steps to address it. In this regard, Government can play a role to support an industry or at least non-government initiative that promotes a code or set of principles. A voluntary compliance regime could also be put in place that urges both industry and consumers to look for compliance in the marketplace.*
- *In some respects, an underlying public concern with AI is around trust, which is similar to a discussion about cybersecurity. And various organisations have different approaches to tackling cybersecurity as a trust issue. For example, the Charter of Trust Initiative brings together several major global companies who have signed up to a range of principles for establishing trust around cybersecurity with their customers and partners. Consideration could be given to whether a similar charter could be developed for AI.*
- *If functions or powers of existing bodies do not sufficiently address a clearly articulated problem arising from AI, there may be a need to consider a new body, subject to a set criteria. For example, if there is an aim to develop an ethical AI industry sector in Australia to ensure we remain competitive, then establishing a new Industry Growth Centre could be an option if it meets the following criteria:*
 - *Long term community cost-benefit: Provides a net benefit to Australia, including businesses and the community, taking account of costs to establish and run it, and the creation of new local jobs, skills and talent.*
 - *Global competitiveness: Offers funding to support the growth of the local AI industry to compete globally.*

- *Proportionality of response: Addresses an issue that is not being tackled by another body and operating within its scope and collaborates with relevant organisations to help achieve its objectives. This may include promoting an ethical AI industry.*
- *Investment incentives or barriers for business: Encourages investment in Australian industry.*
- *Towards the end of the case study about automated vehicles and attribution of responsibility around AI (under section 4.4.1 in the Discussion Paper), it suggests that "there is a need for consistent and universal guidelines, applicable across various industries using technology that is able to make decisions significantly affecting human lives". In practice, this might be difficult to achieve and it would be more prudent to start with sector-specific guidelines (as they will be able to be more practicable). Then, if similarities emerge, the possibility of expanding to other industry sectors could be explored.*
- *To assist stakeholders in implementing the ethical AI principles, the Discussion Paper suggests an impact assessment. It would be useful if an example could be provided for this type of assessment, similar to the example provided for a risk assessment framework (included towards the end of the Discussion Paper). This would also help to clarify the difference between impact and risk assessments.*
- *One member considers that the proposed core principles would be helpful to meet the needs of customers/suppliers in the civilian realm. As principles, these are an important start. Translation into requirements and specifications will require further work. Specifications will be required for working with suppliers. Objective tests or an agency with the ability to interpret principles into test points will be required with customers for determination of the level of "ethics". The interests of shareholders/investors will also need to be respected and appropriately protected, so that any regulatory accountability safeguards to uphold the proposed core principles, go no further than what is reasonably required to achieve that outcome.*
- *The "Toolkit for Ethical AI" set out in the Discussion Paper provides helpful basic tools for best practice development of AI in a way that will uphold the proposed core principles. This toolkit is welcomed for the design of any new technology since it aims at increasing understanding of capability and limitations by different stakeholders. Over time, these principles and tools may help to enhance development of internal policies, and engineering processes/systems for AI.*
- *One member suggests that appropriate regulatory efforts should provide market benefits to compliant firms and drive out non-compliant firms through certification and enforcement. Appropriate support mechanisms for compliant firms, in this context, could include import and export prohibitions on civilian AI products or services that are not compliant with Australian regulations or ACCC consumer protections. They also suggest perhaps an additional view could be considered that those who deploy and make use of AI-enabled tools in areas that impose risk to others should have a licence to make sure they understand the technology and how to use it as well as the potential impact, for the same reason that operation of certain machines requires a licence.*
- *The following best-practice models are worth examining in terms of serving as a template to follow in the practical application of ethical AI: the Ethics and Law Activity currently conducted by the Trusted Autonomous Systems Defence CRC; the Activity on Assurance of Autonomy by the same Centre; and the Robotlaw project in Europe (www.robotlaw.eu).⁶*
- *AI is an emerging technological field that brings unprecedented opportunities for public and private benefit and more efficient use of limited resources. It also brings significant risks that should be responsibly managed. System safety as a field has similar requirements for technical and procedural controls in order to achieve an outcome for a system. That would be a good basis for application of ethics (which in a way is a form of system safety). Alternatively, while it may be stretching the analogy too far to describe it as a "template", there may be some lessons learned*

⁶ S Chopra and LF White, *A Legal Theory for Autonomous Artificial Agents* (The University of Michigan Press, 2014); R Calo, A Froomkin and I Kerr, *Robot Law* (Edward Elgar Publication, 2016).



The Australian Industry Group
51 Walker Street
North Sydney NSW 2060
Australia
ABN 76 369 958 788

from pharmaceutical human research ethics and trials and therapeutic drug regulation that may provide useful touchstones for developing guidance for ethical AI, particularly where human/machine interfaces are involved.

Should the Department be interested in discussing our submission further, please contact our Digital Capability and Policy Lead Charles Hoang (02 9466 5462, charles.hoang@aigroup.com.au).

Yours sincerely,

A handwritten signature in black ink, appearing to read 'Peter Burn'.

Peter Burn
Head of Influence and Policy



The Australian Industry Group
51 Walker Street
North Sydney NSW 2060
Australia
ABN 76 369 958 788

Appendix A: Understanding the broader context

A.1 Role of government

Generally, government's role is to set a vision for the nation, and ensure that public policy is conducive to digital investment and competition that benefits industry and the community in the long term. Government also has a leadership role to allay business and individuals' fears of "Digital Darwinism" including AI, by preparing the community to prosper in an increasingly technology-driven era. To this end, we welcome the Government's release in December last year of its report "Australia's Tech Future: Delivering a strong, safe and inclusive digital economy", which sets out the Government's Digital Economy Strategy.

As alluded to earlier, while technology such as AI may provide benefits to business innovation and productivity, there may be mixed social impacts such as a new division in wealth creation between the technically literate and illiterate. Government has a role in minimising such negative impacts. For example, Government can be a skills enabler through education and training around areas such as digital capability (including in AI), cyber security and privacy.

A.2 Role of regulation

At this early stage of Australia's involvement in AI, positive measures from Government are critical. More can be done to make us globally competitive. Regulation is an important area that could make or break the growth of an industry at its early stages of development. The extent to which AI is regulated can act as an investment barrier and diminish our attractiveness relative to other jurisdictions.

In some areas of regulation in response to modern technology, we have been alarmed by heavy handed interventions that seek to eliminate some forms of risk rather than manage them, while ignoring the risks and costs to innovation and the economy. For example, the recently passed encryption law risks substantial damage to the security and credibility of Australia's connected systems and products and the businesses and people who use them. Such measures not only add costs to international business, but risk curtailing innovation and limiting the benefits of digitalisation to businesses and their customers.

Part of this regulatory response could be due to criticism and concern that regulators are generally not moving fast and flexibly enough to adapt and respond to the pace of technological change. There could also be a lack of understanding of the broader context, such as: the technology; business models; the effect of globalisation; and the role of the different government regulators and other agencies in this environment.

While it is legitimate to question the impact of specific emerging technologies such as AI, the alarm and reactive response from parts of government and the public highlights a wider issue: the role of government in managing the social risks and disruptions associated with new technology.

It is important that regulators are mindful that we have been through similar experiences before with other technological advances like automobiles, telephones and cameras, and more broadly industrial revolutions. As history and experience has shown with these technologies, as the public became more exposed to their presence and practicality, they not only accepted it, but embraced the positive impact that these technologies have had on their lives. Many initial concerns and fears were resolved or proved groundless, and regulation focussed on specific genuine and continuing risks, such as traffic safety or interception of telecommunications. AI is yet to reach that full public comfort, and similar concerns are being expressed about other emerging technologies such as drones, robots and driverless vehicles.

While regulation has a role in addressing reasonable public concerns such as around security, safety, privacy and environmental issues, there are also often alternative approaches to the regulatory "stick". Regulatory barriers should only be introduced where there are clear net community benefits.



The Australian Industry Group
51 Walker Street
North Sydney NSW 2060
Australia
ABN 76 369 958 788

Depending on the identified policy issue, regulation may be an option to address that issue, as well as non-regulatory measures. The issues need to be understood and developed further before an appropriate policy response can be considered.

Notwithstanding our general concerns about traditional regulatory approaches to emerging technologies, we welcome the Discussion Paper's reference to the range of pre-existing laws and regulations that govern AI. Understanding where these laws and regulations fit in this discussion will assist in determining the effectiveness of the current laws and regulations, as well as gaps that can be improved upon (whether through regulation or non-regulatory mechanisms) to address industry and community concerns.

A.3 International comparisons

From an international perspective, Australia is not a leader in AI, facing strong global competition to make the most of AI where it is still behind its peers overseas in terms of AI investment.

The United States and China are reported to be leading in AI investment.⁷ An independent report for the UK Government also acknowledges that the "UK and other countries are generally seen as behind the US and China in terms of scale of investment and activity".⁸ See the Figure below.

While behind the US, the UK is still a significantly larger investor in AI than Australia. Last year, the UK government announced the AI Sector Deal where the government, industry and academia will contribute almost £1bn to support the AI sector.⁹

The Australian Government can play a vital role to improve our global competitiveness in AI. It is therefore positive that Innovation and Science Australia has recommended that priority should be given to the "development of advanced capability in artificial intelligence and machine learning in the medium- to long-term to ensure growth of the cyber-physical economy".¹⁰ It is also positive that the Government announced funding of approximately \$30 million "to develop the artificial intelligence and machine learning capabilities of Australian businesses and workers. This will include funding for Cooperative Research Centre Projects with a focus on artificial intelligence, and a national ethics framework to address standards and codes of conduct for adopting such technology in Australia".¹¹

On this note, we welcome the work of Data61 and its partners to develop this Discussion Paper, as well as its consideration of how ethical AI is approached to overseas. Nevertheless, when comparing jurisdictions on aspects of AI such as ethics, governance and regulation, we emphasise the need to also consider the level of investment in developing an AI industry and the workforce. In other words, different aspects of AI cannot be considered, developed and implemented in isolation, without understanding how it fits into part of a broader national strategy. Australia is a relatively small investor in AI and success will need considerable support from Government – not through free rein for rogue AI operators, but careful consideration of any domestic practices against global best practice approaches and the extent of AI industry support overseas.

⁷ Australian Government, Innovation and Science Australia, "Australia 2030: Prosperity through innovation – A plan for Australia to thrive in the global innovation race" (November 2017), pp. 16-17.

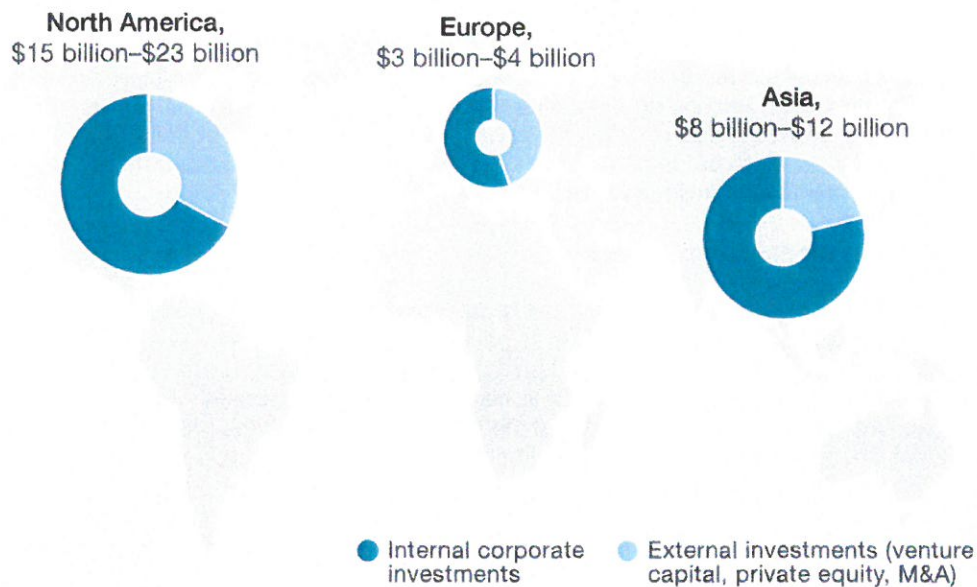
⁸ Professor Dame Wendy Hall and Jérôme Pesenti, Independent Report for the UK Government, "Growing the Artificial Intelligence industry in the UK" (October 2017), pp. 39-40.

⁹ UK Government, "Tech sector backs British AI industry with multi million pound investment" (April 2018).

¹⁰ Australian Government, Innovation and Science Australia, "Australia 2030: Prosperity through innovation – A plan for Australia to thrive in the global innovation race" (November 2017), p. 52.

¹¹ Minister for the Department of Industry, Innovation and Science, "Budget 2018 – New opportunities and jobs for Australian industry" (May 2018).

Artificial-intelligence investment, 2016



Source: McKinsey Global Institute, “10 imperatives for Europe in the age of AI and automation” (Report, October 2017).

A.4 Identifying the problem

Before developing potential solutions to tackle concerns with AI, it is critical that any issues around AI are properly understood. If properly developed, core principles for AI could be used as a guide to assess these issues and develop appropriate responses.

For example, the Australian Human Rights Commission (AHRC) and World Economic Forum (WEF) recently consulted on its White Paper about AI governance and leadership. We appreciated that the White Paper aimed “to identify how Australia can simultaneously foster innovation and protect human rights through the application of new technologies, in particular AI”. However, the White Paper appeared to leap almost immediately to a solution – in this case, a Responsible Innovation Organisation (RIO) – without further unpacking the problem that it was trying to address or considering alternative or existing measures that could be utilised or improved upon.

In that context, we recommended that more substantive work will be needed to understand the human rights issues before they can be considered further. In particular, the AHRC is concurrently consulting on identifying issues relating to human rights and technology (published in its Issues Paper in July last year). We therefore suggested that it might be more prudent for the AHRC to complete that piece of important work before proceeding with the consideration of potential options.

Alternatively, if there was a view to proceed with a RIO, a possible configuration of the RIO could be to consider fundamental questions and issues around AI to inform either the work of existing bodies or the development of a new regulatory response. A contemplative RIO without new regulatory powers could be a useful initial (and perhaps only) step.

Furthermore, we noted that there are a range of possible concerns with AI (e.g. economic disruption and resulting social impact, existential threats, transhumanism) which do not seem to be in the AHRC’s scope. A strong human rights dimension within a broader “governing and thinking through AI” body or regime could make more sense than a body that only tied together one cluster of issues.



The Australian Industry Group
51 Walker Street
North Sydney NSW 2060
Australia
ABN 76 369 958 788

Notwithstanding the above, we note that the Department's Discussion Paper has included useful case studies to assist in identifying problems with AI. To make the most of these case studies, it would be constructive to also include case studies that demonstrate what is considered to be best practice (to the extent that they exist).

A.5 Collaboration between existing bodies

Like other types of emerging technologies, various issues could arise from AI including with respect to standards, education and training, cyber security and privacy, and innovation. To a certain extent, this is acknowledged in the Discussion Paper. In this regard, there will be a need for more collaboration or integration of work between the relevant bodies, industry and the community on AI. These various existing bodies include the ACCC, AHRC, Data61, Fair Work Commission, IEEE, Office of the Australian Information Commission (OAIC), Standards Australia, Industry Growth Centres, Cooperative Research Centres, and government bodies procuring AI-related projects such as the Digital Transformation Agency (DTA).

Appendix B: General comments directly relevant to the Discussion Paper

The following are general comments that were raised by one Ai Group member about the Discussion Paper.

AI in General

There is currently a significant amount of confusion and there are different levels of understanding by members of the public, industry, and government as to what AI is, what it can do and what are the risks of its deployment. We believe the discussion paper does not contribute to clarifying this issue for the audience before it discusses ethics. We believe it should.

It is worth noting that AI is not some far-off futuristic concept. It is with us today and developers, users and societies need to deal with the issues now. Development of ethical guidelines is a task that needs to be addressed and a discussion that needs to be had forthwith.

AI Definition

The definition of AI used in the paper seems too open as it applies to artefacts such as automation systems, any computer algorithm that solves a mathematical problem, any computer program following a normative decision process, a robot, and even appliances (for example, a toaster). Perhaps the focus may better be split into:

- 1) *Artificial agents*
- 2) *Machine learning as part of a decision process.*

Item 1 refers to technology that is able to “perceive” its environment, “analyse” and “decide” how to act to modify the environment – even adapt or learn. This aligns more with the traditional view of how scientists think of technology in the field of AI [1,2]. Item 2 considers mostly data-analytics technology that extracts information from data in order to inform decision making, whether the latter is automated or made by humans [6]. These two classes of AI bring completely different issues in regards to ethics and law.

Ethics and Law

It is surprising that ethics has been dissociated from the Law, since some issues in these areas are very much interrelated as discussed in [3,4]; for example, whether an AI-based system is considered a “tool” and an “agent” in the legal sense, which can bring significant challenges in terms of liability, and also the ethics of its deployment. Indeed, some of the core principles put forward in the paper relate to legal aspects.

Ethics of AI versus ethics of human users and designers

The core principles being put forward in the discussion paper appear to take a view that it is AI that seems unethical, whereas, the ethical issues referred to in most of the examples in the paper arise due to human behaviour using AI without understanding its limitations, and the impact of those limitations. This would be like talking about the ethics of an instrument, (say a knife,) instead of concentrating on how people use the instrument.

Perhaps an additional view could be considered that those who deploy and make use of AI-enabled tools in areas that impose risk to others should have a licence to make sure they understand the technology and how to use it as well as the potential impact, for the same reason that using certain machines requires a licence. It is not the fault of the AI system that it was trained with poor data and therefore cannot generalise its predictions accurately. Also, the fact that a neural network has been trained to examine patterns about individuals and relate them to the likelihood of a credit default is not the main concern; rather, the issue is the blind use by humans of such a tool to make a decision about access to credit.



The Australian Industry Group
51 Walker Street
North Sydney NSW 2060
Australia
ABN 76 369 958 788

Decisions about AI-enabled Autonomy

Over the past ten years, our organization has been conducting work on methods and technology to provide information to stakeholders who make decisions about safety and performance of autonomous systems, including AI-enabled autonomy. For example, regulators face decisions about adapting the legal-regulatory environment as well as certifying systems, actuaries face decisions about insurance of operations, end users face decisions about system acquisition for particular operations and missions, and finally, developers face decisions about system design and configuration that impact performance and safety. We take the view that the understanding of autonomous systems' behaviours and the quantification of uncertainty about these behaviours is the crucial aspect. This draws parallels between the assessment of autonomous-system behaviours and the assessment of human behaviours. Indeed, when we assess human operators as part of their qualification process, we analyse their behaviours against agreed standards, and we do not look into brain activity or other physiometric information as a means to explain the behaviours. We argue that perhaps AI should be treated in a similar way.

Behavioural approach

Ethics refers to standards on how to "behave" within different societies. Similarly, certain aspects of the law (for example, negligence workplace health and safety legislation) are described also in terms of "behaviours" [3,4]. Therefore, a supporting view seems to be emerging that a behavioural approach to the design and assessment of autonomy may be an appropriate path to follow to characterise safety, performance and even ethics and legal aspects in a way that parallels how we approach these issues with humans.

While some behaviours can be related to performance and others to safety, behaviours related to ethics may also be incorporated and assessed in the same way that regulators assess behaviours related to safety.

References

- [1] Russel, S. J. and Norvig, P. (2011) Artificial intelligence - A modern approach. Pearson, third edition.
- [2] Sutton, R. S. and A. G. Barto (2018) Reinforcement Learning: An Introduction, Second Edition, MIT Press, Cambridge, MA.
- [3] Chopra, S. and White, L. F. (2014) A Legal Theory for Autonomous Artificial Agents. The University of Michigan Press, 2014.
- [4] Calo, R., Fromkin, A., and Kerr, I. (2016) Robot Law. Edward Elgar Publication.
- [5] Leenes, R. and Lucivero, F. "Laws on robots, laws by robots, laws in robots: Regulating robot behaviour by design." Law, Innovation and Technology, 2014. 6(3):193–220.
- [6] French, S., J. Maule, and N. Papamichail (2009) Decision Behaviour, Analysis and Support. Cambridge University Press.