

7 June 2019

T +61 2 9223 5744 F +61 2 9232 7174  
E [info@governanceinstitute.com.au](mailto:info@governanceinstitute.com.au)  
Level 10, 5 Hunter Street, Sydney NSW 2000  
GPO Box 1594, Sydney NSW 2001  
W [governanceinstitute.com.au](http://governanceinstitute.com.au)

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

By email: [artificial.intelligence@industry.gov.au](mailto:artificial.intelligence@industry.gov.au)

Dear Sir/Madam

## **Artificial Intelligence: Australia's Ethical Framework**

Governance Institute of Australia (Governance Institute) is the only independent professional association with a sole focus on whole-of-organisation governance. Our education, support and networking opportunities for directors, company secretaries, governance professionals and risk managers are unrivalled.

Our members have primary responsibility for developing and implementing governance and risk frameworks in public listed, unlisted and private companies. They are frequently those with the primary responsibility for dealing and communicating with regulators such as the Australian Securities and Investments Commission (ASIC) and the Australian Prudential Regulation Authority (APRA). In listed companies, they have primary responsibility for dealing with the Australian Securities Exchange (ASX) and interpreting and implementing the Listing Rules. Our members have a thorough working knowledge of the Corporations Act 2001. Our members also play an important role in external reporting by public listed, unlisted and private companies. We have drawn on their experience in providing our feedback.

Governance Institute welcomes the opportunity to comment on the Discussion Paper for Artificial Intelligence: Australia's Ethics Framework (Discussion Paper).

We commend the Government on the Discussion Paper. We support the ongoing discussion of how artificial intelligence (AI) opportunities need to be balanced with risks. We consider that a set of ethical AI principles would greatly aid public trust as these technologies continue to penetrate communities and would also ensure the right mix of innovation and the avoidance of potential harms.

We make the following comments on the Discussion Paper questions to assist in informing the Government's approach to AI ethics in Australia:

### **1. Are the principles put forward in the Discussion Paper the right ones? Is anything missing?**

We note that there is a considerable volume of documents already published on principles, guidelines and frameworks for AI ethics, including many identified within the Discussion Paper, and several attempts<sup>1</sup> to consolidate the growing amount of material being developed globally. There is significant variation however, in the principles identified. This variation is seen in the number of principles, the level of detail and the language used. One area of commonality is that almost all of this material provides for the ability to voluntarily commit to the relevant principles,

guidelines or framework, rather than providing recommendations for operationalising, overseeing or enforcing them.

We suggest that before looking at the individual principles, there should be consideration about how the principles are structured. We believe it would aid in understanding the principles, if they were grouped into categories or themes. This would not only allow a smaller number of categories which may make understanding the overall intention of the principles easier, but also provide an opportunity for including a larger number of more detailed principles within those groupings. To further explain the detailed principles, a number of individual ethical AI issues could also be captured. The following structure<sup>2</sup> is suggested:

- Core Principle
  - Supporting Principles
    - Ethical AI Issues

An example of this structure may be:

#### Example Principle Structure



We suggest two different approaches for developing the core principles. The first is based on the medical ethical principles of respect for “autonomy”, “non-maleficence”, “beneficence” and “justice.” Many of the existing AI ethics documents are based on these four principles, often with a fifth of “explicability.”

The proposed eight principles could be mapped against these core principles as follows:

### Ethical Principles Map

CORE PRINCIPLE	PROPOSED PRINCIPLES	
BENEFICENCE Do Good	1 Generates net benefits	
NON-MALEFICENCE Do No Harm	2 Do no harm	
AUTONOMY Treat everyone equally	4 Privacy protection	7 Contestability
JUSTICE Act fairly	3 Regulatory and legal compliance	5 Fairness
EXPLICABILITY Be transparent	6 Transparency & explainability	8 Accountability

An alternative approach to categorising the proposed principles into a collection of core principles could be to consider the key components of governance: “stewardship”, “transparency”, “accountability”, “integrity” and “risk management”, or STAIR.

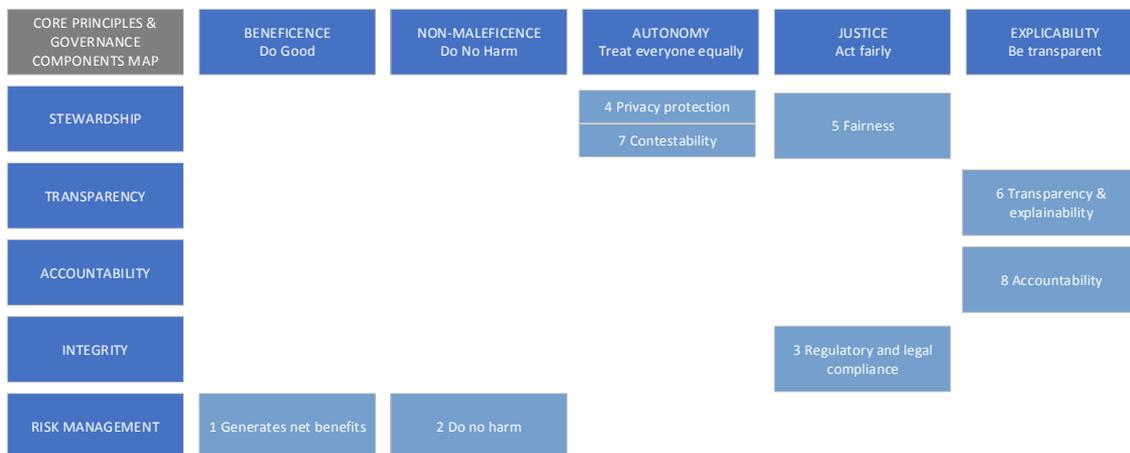
Mapping the proposed principles against these components could follow the structure below:

### Governance Components Map

GOVERNANCE COMPONENT	PROPOSED PRINCIPLES		
STEWARDSHIP	4 Privacy protection	5 Fairness	7 Contestability
TRANSPARENCY	6 Transparency & explainability		
ACCOUNTABILITY	8 Accountability		
INTEGRITY	3 Regulatory and legal compliance		
RISK MANAGEMENT	1 Generates net benefits	2 Do no harm	

The two approaches might be combined into a core principles map showing the proposed principles plotted against an ethical framework and a governance framework to identify missing principles. The vast array of principles contained within the other published documents could be analysed and considered for inclusion within this diagram:

### Combined Core Principles Map



### 2. Do the principles put forward in the Discussion Paper sufficiently reflect the values of the Australian public?

The first proposed principle “Generates Net Benefits” suggests that benefits should be greater than costs. We question the wording used here, noting that “costs” may be considered by some in terms of purely financial costs and suggest “non-benefits” may be preferable. We presume that the intention here is that all “costs” which are non-benefits, including financial, environment, social etc should be weighed up against benefits of any proposed AI system.

### 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?

We suggest that consideration be given to the entire AI value chain, not just customers and suppliers. A principle or obligation for organisations to inquire into the AI supply chain to, for example, identify black box algorithms used within suppliers’ AI systems may be helpful. In the same way, identifying how the AI system may be used is important, to for example, not aid terrorist organisations or those who engage in modern slavery.

### 4. Would the proposed tools enable you or your organisation to implement the core principles for ethical AI?

We agree that the proposed tools are appropriate and comprehensive although we believe that a significant focus on impact assessments is of vital importance.

### 5. What other tools or support mechanisms would you need to be able to implement principles for ethical AI?

We suggest that a website with a variety of tools would be helpful, including:

- case studies where organisations have successfully implemented systems that address each of the proposed principles, including examples of small, medium and large organisations

- template documents for impact assessments and risk assessments
- references to the appropriate Australian and/or International standards
- assessment and decision making diagnostic tools which can assist organisations in their decision making.

**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?**

We would like to draw your attention to the UK Institute for Ethical AI & Machine Learning<sup>3</sup> that proposes that the regulation of responsible AI be approached through four phases:

- By Principle - empowering individual through best practices and applied principles
- By Process - empowering leaders through practical industry frameworks and applied guides
- By Standards - Empowering entire industries through contributions to industry standards
- By Regulation - Empowering entire nations through case studies and breakthroughs

We suggest a similar phased approach as a practical model for applying the principles through to enforcing regulation is a sensible way forward.

Another model to consider applying within an AI ethics framework is the Australian Institute of Health and Welfare Five Safes Framework<sup>4</sup>. This is a risk assessment framework for data access including safe people, safe projects, safe settings, safe data and safe outputs. Aligning the components of this framework with an AI ethics framework may also be helpful.

If you have any questions concerning this submission or would like to discuss any aspect please don't hesitate to contact our General Manager, Policy and Advocacy, Catherine Maxwell.

Yours sincerely



Megan Motto  
CEO

<sup>1</sup> <https://algorithmwatch.org/en/project/ai-ethics-guidelines-global-inventory/>;  
<http://alanwinfield.blogspot.com/2019/04/an-updated-round-up-of-ethical.html>;

<sup>2</sup> A similar hierarchical structure of principles, values and rights is shown in Figure 2 on page 6 of the Draft Ethics Guidelines for Trustworthy AI at  
[https://ec.europa.eu/newsroom/dae/document.cfm?doc\\_id=57112](https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=57112)

<sup>3</sup> <https://ethical.institute/index.html>

<sup>4</sup> <https://www.aihw.gov.au/about-our-data/data-governance/the-five-safes-framework>