



**Electronic Frontiers**  
AUSTRALIA

W [www.efa.org.au](http://www.efa.org.au)  
E [email@efa.org.au](mailto:email@efa.org.au)  
T [@efa\\_oz](https://twitter.com/efa_oz)

8 March 2019

Mr Ed Santow  
Australian Human Rights Commission  
GPO Box 5218  
**SYDNEY NSW 2001**

**By Email: [tech@humanrights.gov.au](mailto:tech@humanrights.gov.au)**

**RE: Submission to the Artificial Intelligence: Governance and Leadership White Paper**

This submission is made on behalf of Electronic Frontiers Australia (“EFA”) in response to the Artificial Intelligence: Governance and Leadership White Paper which was released in January 2019 (“the White Paper”).

This submission ought to be read in the context of the joint submission by the Queensland Council for Civil Liberties (“QCCL”), EFA and the Australian Privacy Foundation to the Human Rights and Technology Project dated 2 October 2018, with particular reference to the sixteen (16) key recommendations made therein.

It is additionally relevant to note that, due to the writer’s position as a Vice President of the QCCL, this submission is in substantially similar terms to the submission made on behalf of the QCCL.

We trust that this submission is of assistance to the Commission.

Please do not hesitate to contact us to discuss this submission.

Yours sincerely,

Angus Murray  
Chair – Electronic Frontiers Australia – Policy Committee  
Email: [angus@irishbentley.com.au](mailto:angus@irishbentley.com.au) | Phone: 0405 715 427

## **Summary**

1. It is prudent to reiterate that technology is not good or bad – it is merely a tool.
2. It is EFA's position that the proper approach to the implementation and regulation of artificial intelligence ("AI") firstly requires a comprehensive legislative framework that protects and promotes human rights. Once this framework has been established, the implementation and regulation of artificial intelligence must be non-exhaustively underpinned by the following principles:
  - a. privacy by design;
  - b. transparency;
  - c. creating an informed society;
  - d. decision-making metrics that are human-reviewable; and
  - e. sustainability.
3. We have responded to the specific questions posted in the White Paper.

### **What should be the main goals of government regulation in the area of artificial intelligence?**

4. Firstly, it is important to make a broad definition for AI. In the writer's view, AI is a term loosely used to describe computational processes that are used to create non-biological intelligence that can be used to accomplish complex tasks. This can occur in two contexts – AI that is focused on a specific task (such as autonomous vehicles<sup>1</sup> or spam filtering<sup>2</sup>) or, AI that identifies solutions when presented with unknown or unfamiliar variables or environments (such as AlfaGo Zero and Alfa Zero<sup>3</sup>).
5. The rapid progression of AI models and technology has created a situation where AI is used prolifically, and the Global Challenges Foundation 2018 Annual Report lists this rapid uptake of AI technology as a potential global catastrophic risk<sup>4</sup>.
6. It is trite that AI requires regulation; however, due to the complex and broad nature of the actual and potential application of AI, it is difficult to create an exhaustive basis for its regulation. Indeed, by the very nature of the creation and application of AI, it is only possible to express a broad series of bases for regulation.
7. In EFA's submission, this must start with a comprehensive legislative framework that protects and promotes human rights and civil liberties. The protection of human rights must occur before the development of AI and its ancillary applications in political advertising, law enforcement, social services, medicine and emerging fields. The rationale for this proposition is simply that human rights, being inalienable fundamental rights, must be properly protected before any legislative or regulatory progression to ensure that a base-line of protection is established. This is particularly important for Australia as we presently lack an enforceable Federal human rights framework.

---

<sup>1</sup> See: <https://medium.com/datadriveninvestor/artificial-intelligence-and-autonomous-vehicles-ae877feb6cd2>.

<sup>2</sup> See: <https://www.wired.com/2015/07/google-says-ai-catches-99-9-percent-gmail-spam/>.

<sup>3</sup> See: <https://deepmind.com/blog/alphago-zero-learning-scratch/>; <https://www.chess.com/news/view/google-s-alpha-zero-destroys-stockfish-in-100-game-match>.

<sup>4</sup> <https://globalchallenges.org/en/our-work/annual-report/annual-report-2018> at pp. 42-43.

8. Once a human rights framework has been properly established and implemented (noting that the Queensland parliament has recently passed the *Human Rights Act 2018* and an enforceable framework could be projected to a Federal level), the main goals of regulation ought to be privacy by design, transparency, creating an informed society, decision-making metrics that are human-reviewable and sustainability.
9. We have elaborated on each of these goals; however, it is important to note that each of these goals are complex and require nuanced consideration that is underpinned by the necessity for an enforceable human rights framework.

### *Privacy by Design*

10. It is uncontroversial that the operation of AI requires immense data sets. This information, where in aggregate or disaggregate, often inherently includes significant volumes of personal and sensitive information that either directly relates to an individual or could be used to reasonably identify individual(s). Our concern is that AI has the potential to “observe” individuals, including individuals in private settings, by processing significant data sets and drawing assumptions about individuals or groups of individuals.
11. Presently, AI can be used to identify individuals via biometrics including facial recognition and footsteps<sup>5</sup>. This complex data-matching requires access to and use of personal information and, although AI may be cheaper than traditional means for aggregating and extrapolating data, it places technology in a position of great power and arguably fundamentally disregards the right to privacy. It is our respectful view that the progression of AI models used to identify, and profile, individuals is unlikely to slow down and requires prompt and clear regulatory guidance. This is particularly important as AI draws from data sets that contain inherent bias (such as racial profiling potential criminal offenders).
12. It is our submission that a key requirement of AI regulation is that AI models require compliance with the principles of privacy by design. A useful starting point for the Commission would be to recommend that AI development and application (as a controller) be compliant with Art. 25 of the *General Data Protection Regulation*.

### *Transparency*

13. On first blush, transparency of AI models and their application is relatively straight forward. However, international examples demonstrate that AI is increasingly an unavoidable part of our daily lives. AI is being used in platforms like Facebook, Netflix, Google and Amazon which show us different advertisements, or recommend different shows/movies for us to watch.
14. These suggestions are based on what the algorithms “think” that we should enjoy or would be of interest to the individual user. However, it gets more problematic when algorithms start being used to determine mortgage rates or criminal propensity<sup>6</sup> or sway political engagement, as these algorithms are unable to be audited or tested for reliability<sup>7</sup>. This raises significant issues for the protection of human rights.

---

<sup>5</sup> Omar Costilla-Reyes, Ruben Vera-Rodriguez, Patricia Scully, Krikor Ozanyan, 'Analysis of Spatio-Temporal Representations for Robust Footstep Recognition with Deep Residual Neural Networks' (2019) 41 2 *IEEE Transactions on Pattern Analysis and Machine Intelligence* at pp. 285-296.

<sup>6</sup> See for example: *Wisconsin -v- Loomis* 881 NW 2d 749 (Wis 2016).

<sup>7</sup> Els, A. (2017). "Artificial Intelligence as a Digital Privacy Protector". *Harvard Journal of Law & Technology*, 31(1). Retrieved from: <http://jolt.law.harvard.edu/assets/articlePDFs/v31/31HarvJLTech217.pdf>.

15. This becomes particularly concerning as AI becomes increasingly capable of mimicking natural language<sup>8</sup> and targeting individuals with content.
16. It is our submission that a regulatory response ought to occur that requires transparency when AI is being used to target, advertise or otherwise interact with individuals or society generally.

#### *Creating an Informed Society*

17. In addition to the principles of privacy by design and transparency, it is our respectful submission that Australians ought only to make decisions where they are firstly provided with sufficient information and knowledge that enables informed consent and decision making.
18. It is our submission that, broadly, AI has the potential to distort information and create distrust. For example, on 14 February 2019, OpenAI released an AI model<sup>9</sup> that enabled the creation of natural language prose that was so convincing that it could be used to mass-produce “fake news”<sup>10</sup>. It is clear from examples such as the Cambridge Analytica scandal that people are susceptible to “psychological manipulation, entrapment techniques and fake news campaigns”<sup>11</sup> and that these techniques have the (alleged) potential to influence democracy.
19. We appreciate that regulation to prevent malicious applications of AI in advertising is a complex discourse; however, it is our submission that this issue may be alleviated by an increased focus on education.
20. An increase in both user and creator education mitigates the potential for misuse and naïve acceptance of information.
21. It is our submission that government regulation in the area of AI ought to extend to an expansion of education to include the use and consequence of AI and advanced technology in both the school system and university courses.

#### *Decision-Making Metrics that are Human-Reviewable*

22. In addition to an enforceable human rights framework, the Commission ought to consider the means to introducing a regulatory requirement that AI models be translatable into human-reviewable decisions.
23. We appreciate that this recommendation will face technological barriers. For example, the human understanding and extrapolated meaning from Soft Actor-Critic AI Models which rely on maximised entropy and acting as randomly as possible<sup>12</sup>. However, it is important that, as AI and automated decision making becomes more prevalent, there must remain the ability to review such decisions.
24. Indeed, “[t]he legal conception of what constitutes a decision cannot be static; it must comprehend that technology has altered how decisions are in fact made and that aspects of, or the entirety of, decision making, can occur independently of human mental input.”<sup>13</sup>

<sup>8</sup> See for example: <https://www.technologyreview.com/s/612975/ai-natural-language-processing-explained/>.

<sup>9</sup> See: <https://blog.openai.com/better-language-models/>.

<sup>10</sup> <https://www.technologyreview.com/s/612960/an-ai-tool-auto-generates-fake-news-bogus-tweets-and-plenty-of-gibberish/>.

<sup>11</sup> See: <https://www.cnn.com/2018/03/21/facebook-cambridge-analytica-scandal-everything-you-need-to-know.html>.

<sup>12</sup> See: Tuomas Haarnoja et al, ‘Soft Actor-Critic Algorithms and Applications’ (2019) 2 *Cornell University* (available at URL: <https://arxiv.org/abs/1812.05905>).

<sup>13</sup> See: *Pintarich v Deputy Commissioner of Taxation* [2018] FCAFC 79 at [49] per Kerr J.

25. It is our submission that the Commissioner ought to ensure that the reviewability of AI decision making is included as a regulatory goal.

#### *Sustainability*

26. Finally, it is our submission that there is an additional human rights context that must apply when considering the regulation of AI. This is the environmental impact of AI.

27. Big data is often discussed in the context of “the cloud” with little regard for the energy and environmental cost of creating, operating and disposing of complex computation systems.

28. It is our submission that a goal of the regulation of AI ought to be the protection of sustainable and environmentally acceptable technology.

#### **What existing bodies play an important role in this area and what are the gaps in the regulatory system?**

29. We respectfully repeat paragraphs 4 to 28 of this submission and reiterate that the largest gap that presently exists in Australia is the lack of a Federal enforceable human rights framework.

#### **Would there be significant economic and/or social value for Australia in establishing a Responsible Innovation Organisation?**

30. EFA respectfully repeats that, as a starting point, the Commission ought to recommend the introduction of a Federal enforceable human rights framework (or endorse the need for such a framework).

31. EFA appreciates the suggested functions and powers of a Responsible Innovation Organisation (“RIO”) expressed at Part 5.2 of the White Paper; however, it is EFA’s position that this would be best addressed via a certification scheme that causes the implementation of and requires:

- a. privacy by design;
- b. transparency;
- c. consumer and creator education;
- d. decision-making metrics that are human-reviewable; and
- e. sustainability.

32. It is EFA’s submission that social value must be placed before economic value and the discourse of human rights and technology ought not become a discussion about economic merit that loses sight of the underlying and fundamental human rights that the AI economic will inherently be built upon.

#### **Under what circumstances would a RIO add value to your organisation?**

33. We respectfully repeat paragraphs 30 to 32 of this submission.

#### **How should the business case for a RIO be measured?**

34. It is EFA's submission that social value must be placed before economic value and the discourse of human rights and technology ought not become a discussion about economic merit that loses sight of the underlying and fundamental human rights that the AI economic will inherently be built upon.
35. It is our position that any measure of success must primarily relate to the promotion and protection of human rights and civil liberties.

**If Australia had a RIO:**

36. At the outset, and in the interest of completeness, EFA's primary recommendation is that the Commission ought to recommend the introduction of a Federal enforceable human rights framework (or endorse the need for such a framework). A RIO *may* be beneficial; however, the introduction of this organisation should be subordinate to a recommendation to introduce a Federal enforceable human rights framework.

**What should be its overarching vision and aims?**

37. A certification scheme that causes the implementation of and requires:
- a. privacy by design;
  - b. transparency;
  - c. consumer and creator education;
  - d. decision-making metrics that are human-reviewable; and
  - e. sustainability.

**What powers and functions should it have?**

38. A RIO should function as an industry accreditation and public awareness body that serves to certify and recognise best-practice in the development and use of AI.

**How should it be structured?**

39. EFA considers that a RIO must be a stand-alone organisation that is free from perceived or actual bias or conflict that may arise from association or affiliation to other bodies and overseen by the Human Rights Commissioner.

**What internal and external expertise should it have at its disposal?**

40. A RIO would not function without a board of remunerated experts that are randomly selected from a meritocratic pool of potential members available for bespoke appointment.
41. A RIO's decision making processes in relation to accreditation ought to be public.

**How should it interact with other bodies with similar responsibilities?**

42. A RIO should be mandated to interact with other bodies as and where technical or expert reference and responsibility is required and desirable.

**How should its activities be resourced?**

43. A RIO should only be resourced via public funding and certification application fees.

**How should it be evaluated and monitored?**

44. A RIO decision making processes in relation to accreditation ought to be public and an annual accreditation report tabled before Parliament and publicly disseminated.

45. EFA appreciates the opportunity to make this submission and please do not hesitate to contact Mr Angus Murray should you require any further information or comment.