

# Access Now submission to the Department of Industry, Innovation and Science's Paper Artificial Intelligence: Australia's Ethics Framework

MAY 2019

#### **TABLE OF CONTENTS**

I. INTRODUCTION	2
II. WHAT THE DISCUSSION PAPER GETS RIGHT	2
The Threat to 'Anonymized' Data	3
A Strong Stance on Accountability	3
III. THE NEED FOR A HUMAN-RIGHTS-BASED ETHICS FRAMEWORK	4
The Danger of Ethics without a Foundation in Human Rights	4
The Problem of Generating Net-Benefits	4
Unintended Harm is Still Harm	5
IV. HARMONISING THE DISCUSSION PAPER WITH THE HUMAN RIGHTS PERSPECTIVE	6
Lessons from the European Commission's Ethics Guidelines for Trustworthy Al	6
Revising the Eight Principles: Hierarchy and Human Rights	8
V. ADDITIONAL RECOMMENDATIONS TO MITIGATE THE MOST DETRIMENTAL POTENTIA	۱L
IMPACTS OF AI ON AUSTRALIAN SOCIETY	9
VI. CONCLUSION	12



#### I. INTRODUCTION

Access Now is committed to defending and extending digital rights on the frontlines of technological change. We recognise that the direction taken by developments in Artificial Intelligence will have a significant impact on human rights, both in the digital world and beyond it. Whether that impact is positive or negative will depend on the efforts taken in the present to ensure that AI is developed in a way that builds in human rights standards from the beginning.

Over the past year, Access Now have made several important contributions to the debate on AI, with reports on <u>Human Rights in the Age of Artificial Intelligence</u> and <u>Mapping Regulatory Proposals for Artificial Intelligence in Europe</u>, along with a host of <u>blog posts</u> and submissions to various initiatives. We also provided feedback to the Australian Human Rights Commission's 2018 <u>consultation on Artificial Intelligence</u>. On top of this, our European Policy Manager, Fanny Hidvégi, is one of the 52 members of the European Commission's High-Level Expert Group on AI, and thus one of the contributors to the <u>Ethics Guidelines for Trustworthy AI</u>. She was also the only civil society representative invited to participate at the <u>G7 multistakeholder conference on AI</u> in December 2018.

We support initiatives that make the effort to promote ethical and human-centric AI, but we remain convinced that such goals can only be achieved by placing human rights front and centre.

Australia faces a unique challenge in this space as there is no affirmative right to privacy in Australian law. Nor do Australians have the ability to file a lawsuit against an individual, entity, or the government for a violation of their privacy. The current patchwork of privacy legislation seeks to regulate government and private use of data but is incomplete when we compare it to international standards. It is also inadequate to address the challenges brought by AI as it focuses too strongly on the economic utility of users' data, a concept which overshadows the need to protect individual rights.

There is an urgent need to address these gaps in Australian legislation as we move forward in the digital era. Australian decision-makers should seize this opportunity to build a framework that protects individual rights and shields minority and vulnerable groups from any potential adverse impacts of AI and emerging technologies.

#### II. WHAT THE DISCUSSION PAPER GETS RIGHT

Firstly, we will address some areas where the Australian Ethics Framework is particularly successful in addressing the challenges posed by Artificial Intelligence.



#### The Threat to 'Anonymized' Data

Data anonymization is all too often seen as a universal solution to privacy and data protection concerns. We are pleased to see that the discussion paper acknowledges the limitations of anonymizing personal data in stating that:

"The use of AI enabled devices and networks that can collate and predict data patterns has heightened the risk of being able to identify individuals in what was considered a de-identified dataset."

While anonymization of data should continue to be a best practice when it comes to protecting individual's rights, it should be supported by a number of complementary data protection measures such as data minimization and consent for processing in order to be truly effective.

#### A Strong Stance on Accountability

Regarding the ethical principles outlined in the discussion paper, we particularly commend the strong stance on accountability taken by the eighth principle which states that:

"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."

It is essential to maintain clear accountability in the deployment and use of AI systems. Since certain types of AI systems can infer their own logic and decision-making parameters, this allows companies and governments alike to try to deflect responsibility and blame onto black box algorithms or otherwise inscrutable systems.

The clear requirement on accountability is also in line with the Council of Europe Human Rights Commissioner's recent publication on AI. The paper states that "[r]esponsibility and accountability for human rights violations that occur in the development, deployment or use of AI Systems must always lie with a natural or legal person, even in cases where the measure violating human rights was not directly ordered by a responsible human commander or operator."

This is particularly important in cases where public bodies make use of private contractors to develop AI systems. In such cases of public private partnerships, governments must ensure that the contract guarantees transparency and provides avenues for individuals to gain insight and understanding into the system should they wish to do so. Consequently, systems that cannot be subjected to appropriate standards of transparency and accountability should not be used.

<sup>&</sup>lt;sup>1</sup> Unboxing artificial intelligence: 10 steps to protect human rights (2019), https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64.



#### III. THE NEED FOR A HUMAN-RIGHTS-BASED ETHICS FRAMEWORK

#### The Danger of Ethics without a Foundation in Human Rights

Although Access Now broadly supports initiatives to develop an ethics framework for AI, we believe that any such framework must be based on, and be fully compatible with, the existing, established, and binding human rights framework. As noted in Section 2.1.2 of the draft Ethics Framework, Australia is a signatory to seven core international human rights agreements. In spite of that, Australia exists without constitutional protections for human rights or a national human rights charter. Instead, Australia has an incomplete patchwork of human rights protections found in the Australian constitution<sup>2</sup> and in the constitutions of a few states and territories,<sup>3</sup> common law, statutory law, and the inherent respect for civil liberties built into a democratic system of government.<sup>4</sup>

Beyond these formal commitments, the discussion paper also acknowledges the human rights framework as the culmination of centuries of ethical thinking, and states that its "ethics framework for AI is not about rewriting these laws or ethical standards, it is about updating them to ensure that existing laws and ethical principles can be applied in the context of new AI technologies."

Unfortunately, despite this overt acknowledgement of the pre-eminence of the human rights framework, the discussion paper's ethical principles begin from a utilitarian basis of generating net benefits. The consequence of accepting that as the central value of the discussion paper is that overall societal benefits would always prevail regardless of the individual harm. This approach is in conflict with the essence of Australia's human rights commitments. Instead of just "updating" the human rights framework, the position taken in this document undermines some of its core principles. This, combined with the lack of a robust distinction between legal and human rights obligations and ethical principles, leads to a number of serious confusions and conflicts which we outline below.

#### **The Problem of Generating Net-Benefits**

The discussion paper's divergence from the human rights framework is first and foremost the result of the privileged position given to the principle of generating net-benefits. The definition given in the document states that any "AI system must generate benefits for people that are greater than the costs." The fact that this principle comes before all others means that the document explicitly endorses a utilitarian approach to playing off harms against benefits. But how can such a net-benefit be calculated? Are enormous profit margins for private corporations enough to justify the <u>erosion of privacy standards for the general public</u>? And if not, are the harms generated by an AI system

<sup>&</sup>lt;sup>2</sup> The Australian Constitution provides for just five rights: the right to vote, protection against the acquisition of property on unjust terms, the right to a trial by jury, freedom of religious observance, and a prohibition of discrimination on the basis of state of residency. In 1992, the High Court also recognized that the right of political communication is also constitutionally protected.

<sup>&</sup>lt;sup>3</sup> Victoria and the Australian Capital Territory both have human rights charters based on the International Covenant of Civil and Political Rights (ICCPR).

<sup>&</sup>lt;sup>4</sup> "Human Rights in Australia," Australian Human Rights Commission, <a href="https://www.humanrights.gov.au/education/students/get-informed/human-rights-australia">https://www.humanrights.gov.au/education/students/get-informed/human-rights-australia</a>.



permissible if they only affect <u>recipients of social welfare</u> but save money for others? The other issue with making that calculation is how potential benefits are measured against the potential harms: how likely or how foreseeable must the benefits be to be given more weight than the harms, and what evidence is necessary for making such claims?<sup>5</sup>

Many such questions are bound to plague any framework which places the calculation of net-benefit as its foremost ethical principle. By contrast, the human rights framework begins from the idea that certain harms are never permissible, no matter what the benefit is. In practice, of course, situations do arise where unfortunate trade-offs need to be made between rights which come into conflict. When such situations do arise, however, international human rights law can refer to its well-developed standards and institutions as well as a universal framework for safeguards. Moreover, such trade-offs are seen as exceptions from the human rights perspective, rather than the primary mode of operation as suggested by the principle of generating net-benefits.

Another problem with the principle of generating net-benefits is that it provides no answer to the question of whose benefit matters most. Can the benefit of one group cancel out the harms done to another? One of the "Key Points" listed in Section 6.6 states that:

"AI-enabled surveillance technologies should consider "non-instrumentalism" as a key principle—does this technology treat human beings as one more cog in service of a goal, or is the goal to serve the best interests of human beings?"

This idea of non-instrumentalism is arguably in direct conflict with the principle of generating net-benefits: if we follow the former, must we not state that harm done to one person or group can never be justified as a means to another's benefit? Instead of leaving the calculation of net-benefits up to the ethical whims of interested parties and spurious utilitarian calculations, the human rights framework offers clear safeguards against such instrumentalism and robust procedures to deal with conflicts and trade-offs. In Australia's context, the approach should be heavily weighted towards extreme caution when considering such net-benefit calculations, as trade-offs for the general population are most likely to have detrimental impact on the indigenous population and vulnerable communities.

#### **Unintended Harm is Still Harm**

Beyond this issue of calculating one group's harm against another's benefit, the draft framework runs into another serious problem with its conception of harm. Its second ethical principle is to **do no** 

<sup>&</sup>lt;sup>5</sup> See, for example the lack of clarity in the Google AI Principles when they say that they "will proceed where we believe that the overall likely benefits substantially exceed the foreseeable risks and downsides", <a href="https://ai.google/principles/">https://ai.google/principles/</a>

<sup>&</sup>lt;sup>6</sup> See, for example, the discussion of the human rights impact of AI in the Berkman Klein Center's report on Artificial Intelligence and Human Rights: <a href="https://cyber.harvard.edu/publication/2018/artificial-intelligence-human-rights">https://cyber.harvard.edu/publication/2018/artificial-intelligence-human-rights</a>

<sup>&</sup>lt;sup>7</sup> For a detailed discussion of the human rights approach to proportionality and trade-offs, see the article from Bernadette Somody, Máté Dániel Szabó, and Iván Székely, 'Moving away from the security–privacy trade-off The use of the test of proportionality in decision support':

http://archive.ceu.hu/sites/default/files/publications/somody-szabo-szekelymoving-away-trade-off.pdf



**harm** and states that: "Civilian AI systems must not be designed to harm or deceive people and should be implemented in ways that minimise any negative outcomes." There are multiple issues here, such as why this principle is limited to "civilian AI systems", but the most serious issue is that this principle only covers systems designed to harm or decieve people. If <u>recent research</u> into the impact of AI systems has shown anything, it is that they are liable to cause <u>serious unintended harms</u>.

It is not enough to avoid designing systems that set out to harm people. Those developing or using an AI system must do everything possible to ensure that nobody is intentionally or unintentionally harmed as a consequence of its operation. There needs to be an understanding that there will be instances where AI should not and will not be available as a modus operandi for this reason alone.

## IV. HARMONISING THE DISCUSSION PAPER WITH THE HUMAN RIGHTS PERSPECTIVE

The issues outlined above are just some of the clear and predictable consequences of having a set of ethical principles that lack a strong grounding in human rights. To overcome these confusions and conflicts, we recommend that the discussion paper's ethical principles be brought into harmony with the fundamental principles of the human rights framework which we will present here through the European Union's work in developing guidelines for trustworthy AI.

### Lessons from the European Commission's Ethics Guidelines for Trustworthy AI

Perhaps the most prominent initiative to develop an ethical framework for AI has come from the European Commission's High-Level Expert Group on AI (HLEG) with its "Ethics Guidelines for Trustworthy AI". Two aspects of this document are of particular relevance to the Australian Ethics Framework: firstly, its clarity on the distinction between legal obligation and ethical aspiration; secondly, the approach it takes to deriving its ethical principles from a human rights foundation.

In terms of the differentiation between legal compliance and ethical aspiration, the HLEG Guidelines clearly state that AI systems must first and foremost be legally compliant. Such compliance must take into account not only national law, but also UN Human Rights treaties and the Council of Europe Conventions (such as the European Convention on Human Rights). Only once this basic compliance has been assured should one begin to speak about the aspiration to ethical principles.

In a similar, but clearer manner than this draft of the Australian Ethics Framework, the HLEG Guidelines acknowledge the preeminence of human rights for the development of an ethical framework:



"Respect for fundamental rights, within a framework of democracy and the rule of law, provides the most promising foundations for identifying abstract ethical principles and values, which can be operationalised in the context of AI."

On this basis, the HLEG derives four ethical principles, with no implicit hierarchical order, from the fundamental rights framework:

- > Respect for human autonomy
- > Prevention of harm
- > Fairness
- > Explicability

Although all four principles contain much that is worthy of emulation, this draft of Australia's Ethics Framework for Artificial Intelligence could significantly benefit from adopting certain aspects of the principles of **respect for human autonomy** and **prevention of harm**.

Regarding respect for human autonomy, the HLEG's guidelines state:

"Humans interacting with AI systems must be able to keep full and effective self-determination over themselves, and be able to partake in the democratic process. AI systems should not unjustifiably subordinate, coerce, deceive, manipulate, condition or herd humans. Instead, they should be designed to augment, complement and empower human cognitive, social and cultural skills."

In contrast to the eight principles in the Australian discussion paper, the EU document begins with a commitment to upholding human rights and explicitly stating the lines that cannot be crossed.

With respect to the prevention of harm, the HLEG's guidelines clearly state that "AI systems should neither cause nor exacerbate harm or otherwise adversely affect human beings." In contrast to the Australian Framework, this conception of the harms caused by AI goes beyond systems *designed* to harm people, and takes account of the diverse manners in which such systems can harm people.

Moreover, the EU document makes the following strong statement about the need to consider both historical patterns of discrimination and asymmetries of power when considering the potential of AI systems to cause harm:

"Vulnerable persons should receive greater attention and be included in the development, deployment and use of AI systems. Particular attention must also be paid to situations where AI systems can cause or exacerbate adverse impacts due to

<sup>&</sup>lt;sup>8</sup> High-Level Expert Group on AI, Ethics guidelines for trustworthy AI, p.9.

<sup>&</sup>lt;sup>9</sup> High-Level Expert Group on AI, Ethics guidelines for trustworthy AI, p.12.



asymmetries of power or information, such as between employers and employees, businesses and consumers or governments and citizens."<sup>10</sup>

The exact framing of this approach should steer the debate on AI in Australia, making sure that the interests of individuals - especially those of vulnerable persons - remain central to all AI development and deployment. Seeing the adverse impact of even simple automated decision making, such as with Centrelink's welfare financing or "robo debt", should serve as a cautionary tale for all government services and a stark warning of the impacts emerging technologies can have on vulnerable groups.<sup>11</sup>

#### **Revising the Eight Principles: Hierarchy and Human Rights**

In light of the criticisms outlined above, and the positive alternative examples presented, we propose that the following revisions be made to both the order and the content of the 8 ethical principles underlying the discussion paper.

- 1. The principle of generating net-benefits should be replaced by a first principle which states that any alleged benefits of an AI system must not come at the cost of possible human rights violations. In the design, development and deployment of AI, international human rights law applies. Therefore, any systems which could lead to the violation of human rights should meet these standards, no matter what financial or other benefits they may bring. Further legal and policy steps are necessary to identify areas where the development and/or deployment of AI systems should not be permitted.
- 2. The principle of **do no harm** should be substantially revised to state that AI systems must neither cause harm to human beings, nor exacerbate existing harms, whether intentionally or not. Moreover, following the HLEG Guidelines, consideration must be given to the existence of historical patterns of discrimination and asymmetries of power.
- 3. The "ethical principle" on Regulatory and Legal Compliance should not be considered an ethical principle, but should be a condition that precedes any consideration of ethics. It is important to make a clear differentiation between legal obligation and ethical aspiration, and legal compliance should explicitly include Australia's obligations to uphold and promote human rights standards.

<sup>&</sup>lt;sup>10</sup> High-Level Expert Group on AI, Ethics guidelines for trustworthy AI, p.12.

<sup>11</sup> 



# V. ADDITIONAL RECOMMENDATIONS TO MITIGATE THE MOST DETRIMENTAL POTENTIAL IMPACTS OF AI ON AUSTRALIAN SOCIETY

Arguably the largest gap in human rights protections for the development of AI is that there is no affirmative right to privacy in Australia. The *Privacy Act* regulates the collection and use of personal information via a set of "Privacy Principles" that apply to most federal government agencies, as well as businesses and nonprofits with an annual turnover of over \$3 million, with some exceptions.<sup>12</sup> It defines personal information as "information or an opinion, whether true or not, and whether recorded in a material form or not, about an identified individual, or an individual who is reasonably identifiable."<sup>13</sup>

The Privacy Principles include a number of laudable provisions, such as requiring entities to inform individuals about why their personal information is being collected, how it will be used, and to whom it will be disclosed, allowing individuals the option of not identifying themselves or using a pseudonym (with exceptions), mandating the collection of personal information "only if necessary," allowing individuals to request access to their personal information, as well as request that the information be corrected if inaccurate.

Despite this, there are a number of important shortcomings of the Privacy Act. First, it does not address the ability of AI to easily re-identify "anonymized" information. Information in anonymized datasets does not qualify as personal information according to the definition, and therefore is not subject to any of the protections stipulated by the Privacy Principles. Second, there is no requirement for entities to obtain consent prior to collecting personal information. This removes the right of individuals to decide whether or not they are comfortable with the entity's disclosed use of their data. Third, it does not apply to state or territory governments, although some states have passed their own privacy legislation. Fourth, it does not address the privacy threats of surveillance. Although the Privacy Act covers the Australian Federal Police and the Border and CrimTrac, it does not cover most law enforcement and intelligence agencies, which are arguably the entities most likely to commit major breaches of privacy. Further, where it does apply to law enforcement agencies there are many exemptions and carve outs that allow for near limitless data collection. And finally, the Privacy Act only provides for limited civil redress via a complaints mechanism overseen by an Information Commissioner. These gaps make the Privacy Act nearly useless in protecting Australians against the privacy and data protection risks posed by AI.

The Data Sharing and Release Act is a new piece of legislation that would erode the already insufficient Privacy Act even further. Whenever individuals interact with the government, data are created about their activities. As part of its modernization efforts, the Australian government would

<sup>&</sup>lt;sup>12</sup> The following businesses are subject to the privacy act regardless of their size: private sector health providers, businesses that sell or purchase information, credit reporting bodies, and government contractors.

<sup>13</sup> https://www.oaic.gov.au/privacy-law/

<sup>&</sup>lt;sup>14</sup> https://digitalrightswatch.org.au/2018/05/14/the-state-of-digital-rights/



like to be able to capitalize on all the data it collects. The government states that "better use of public sector data can help us improve government services for Australians and ensure our programs and policies are informed by evidence." The Data Sharing and Release Act seeks to make it easier for government agencies to share data with each other, allowing any government entity to access any and all the information the government holds about individuals, and also permitting the government to share data with "trusted" third parties and researchers. And while it is certainly good to use data analysis to create evidence-based policy, the proposed law lacks any privacy and security protections. It merely states that the risks should be "appropriately managed." It also reflects an ignorance of how seemingly innocuous data points, when combined and analyzed by ML systems, can quickly reveal intimate details about people's lives.

Currently, such use of individuals' data potentially conflicts with the over 500 existing data secrecy and confidentiality provisions across existing Australian law. Notably, it violates the Privacy Act and the Australian Privacy Principles, which state that government agencies cannot use individuals' administrative data for secondary purposes unrelated to providing them with a respective service. However, if passed, the Data Sharing and Release Act would override any conflicting legislation, for both government and non-governmental entities alike.<sup>17</sup> Additionally, the bill would instate data sharing by default. There would be no ability for Australians to opt-out of having their data being shared across the government and with third parties. With this bill, the government is clearly communicating its view that your data belongs to the government because they collect it, as well as continuing carelessness in its approach to risk management of technology projects.

The following recommendations, some of which Access Now made in its reports on human rights in the age of artificial intelligence and on the state of digital rights in Australia, would substantially mitigate the most detrimental potential impacts of AI on Australian society.

1. **Strengthen the powers of the Australian Human Rights Commission and other regulators.** The Australian Human Rights Commission, which oversees Australia's compliance with its international human rights obligations, can hear complaints and resolve breaches of federal anti-discrimination law. However, it has no broad legal authority to act as an arbiter of human rights protections. Like the international human rights system from which it stems, its power lies in calling out government misdeeds if they wish to affect change. <sup>19</sup>

<sup>&</sup>lt;sup>15</sup> https://www.pmc.gov.au/resource-centre/public-data/issues-paper-data-sharing-release-legislation

<sup>16</sup> https://www.eigenmagic.com/2018/07/30/tlir-data-sharing-and-release-issues/

<sup>&</sup>lt;sup>17</sup> https://independentaustralia.net/life/life-display/-the-data-sharing-and-release-act-is-coming-for-your-data,11761

<sup>&</sup>lt;sup>18</sup> Access Now, Human Rights in the Age of Artificial Intelligence

<sup>(</sup>https://www.accessnow.org/cms/assets/uploads/2018/11/Al-and-Human-Rights.pdf) and Human Rights in the Digital Era: An International Perspective on Australia

<sup>(</sup>https://tech.humanrights.gov.au/sites/default/files/inline-files/6%20-%20AccessNow\_0.pdf)

<sup>&</sup>lt;sup>19</sup> The AHRC has begun thinking about its role in protecting human rights in the age of AI, and how it might promote responsible innovation. See

https://www.humanrights.gov.au/our-work/rights-and-freedoms/publications/artificial-intelligence-governance-and-leader ship for more information.



The specific human rights risks of AI in Australia are in many ways a continuation of existing threats to digital rights. The lack of legal human rights protections has enabled the erosion of certain digital rights, most notably the right to privacy. In July 2018, Access Now released a report examining Australia's approach to human rights in the digital age. The report concluded that Australia appears more than willing to undermine human rights as it struggles to adapts to the challenges of the digital era.<sup>20</sup>

The powers of the Australian Human Rights Commission, similarly to the Australian Information Commissioner or the Commonwealth Ombudsman, are limited to monitoring and enforcing protections which are provided for by law. As long as those protections remain deficient in supporting meaningful rights for individuals in the digital space, so will the action of regulatory bodies. Australia must therefore introduce binding legislation to protect human rights against the challenges posed by emerging technologies such as AI.

- 2. Immediately repeal the Data Sharing and Release Act and the Identity Services Bill. Both bills ignore the high likelihood of human rights violations, and risk institutionalizing mass, unchecked public surveillance if they are passed.
- 3. Conduct a comprehensive inquiry into the impacts of AI and automated decision making on indigenous Australians, with a view to ensuring such technologies are used to benefit, rather than harm, indigenous communities. Any systems which are likely to have an impact on indigenous communities should involve representatives from the communities in the process of stakeholder consultation in a meaningful way.
- **4. Sign the International Principles on the Application of Human Rights to Communications Surveillance.**<sup>21</sup> The insertion of AI into Australia's unchecked expansion of domestic surveillance is perhaps the single biggest threat to the rights of Australians. Abiding by these principles would check the worst abuses, and enable the government to protect national security without infringing upon human rights.
- 5. Update the Privacy Act and Privacy Principles to provide Australians with affirmative rights to privacy and data protection, and address the unique risks posed by Al. First, without a right to privacy, there are too many gray areas that allows entities to get away with privacy violations. Second, comprehensive data protection legislation, like the General Data Protection Regulation in the EU, can anticipate and mitigate many of the human rights risks posed by Al. Because data is the engine of Al, any law that mandates protection of personal data will necessarily implicate Al systems. Particularly helpful provisions include adopting and implementing the data minimisation and purpose limitation principles and establishing clear

<sup>&</sup>lt;sup>20</sup> Access Now, *Human Rights in the Digital Era: An International Perspective on Australia:* https://tech.humanrights.gov.au/sites/default/files/inline-files/6%20-%20AccessNow\_0.pdf

<sup>&</sup>lt;sup>21</sup> Necessary and Proportionate, *International Principles On The Application Of Human Rights To Communications Surveillance*: https://necessaryandproportionate.org/principles



legal basis for collecting and processing data, including opt-in consent. Access Now has a detailed guide on how to create a data protection framework that respects human rights.<sup>22</sup>

**6. Address the specific AI related human rights harms.** In addition to adopting and enforcing a general human rights framework, all stakeholders must ensure that the design, deployment and development of AI systems is individual centric and respects human rights. Because the application of AI systems covers such a large and diverse field, any approach to dealing with human rights risks and to preventing the foreseeable detrimental impacts of AI will need to be sector-specific to some extent. Access Now has developed specific recommendations for high standards for government use of AI, meaningful human control, data protection and non-discrimination, and more.

#### VI. CONCLUSION

The intention behind the creation of an ethics framework to guide AI development in Australia is laudable, but <u>ethics guidelines can only be a first step</u>. The European Commission introduced the concept of "trustworthy AI" in its guidelines as a voluntary framework to achieve legal, ethical and robust AI. Just as the European Commission must now follow through with policy recommendations and binding, legal frameworks to ensure that "trustworthy AI" is not just an empty brand name, Australia should pursue a similar effort.

Without comprehensive protections for personal data and privacy, there can be no hope for a rights-respecting AI framework. The threat to vulnerable groups in particular cannot be ignored. As the consultation paper correctly notes, there is a legislative and regulatory gap which needs to be addressed as Australia considers further use of AI and emerging technologies.

#### The response to the consultation was prepared by:

Daniel Leufer, daniel.leufer@accessnow.org Fanny Hidvegi, fanny@accessnow.org Lucie Krahulcova, lucie@accessnow.org

For any questions or to connect with us about our work in Australia, please reach out to Lucie Krahulcova, Access Now's Australia and Asia Policy Analyst.

<sup>&</sup>lt;sup>22</sup> See https://www.accessnow.org/cms/assets/uploads/2018/01/Data-Protection-Guilde-for-Lawmakers-Access-Now.pdf

<sup>&</sup>lt;sup>23</sup> Access Now, Human Rights In The Age Of Artificial Intelligence:

https://www.accessnow.org/cms/assets/uploads/2018/11/Al-and-Human-Rights.pdf

<sup>&</sup>lt;sup>24</sup> Access Now, Mapping Regulatory Proposals For Artificial Intelligence In Europe,

https://www.accessnow.org/cms/assets/uploads/2018/11/mapping regulatory proposals for AI in EU.pdf

<sup>&</sup>lt;sup>25</sup> The Toronto Declaration,

https://www.accessnow.org/cms/assets/uploads/2018/08/The-Toronto-Declaration\_ENG\_08-2018.pdf