Australian Human Rights Commission

Human Rights and Technology Issues Paper

Global Partners Digital Submission

About Global Partners Digital

The advent of the internet – and the wider digital environment – has enabled new forms of free expression, organisation and association, provided unprecedented access to information and ideas, and catalysed rapid economic and social development. It has also facilitated new forms of repression and violation of human rights, and intensified existing inequalities.

Global Partners Digital (GPD) is a social purpose company dedicated to fostering a digital environment underpinned by human rights and democratic values. We do this by making policy spaces and processes more open, inclusive and transparent, and by facilitating strategic, informed and coordinated engagement in these processes by public interest actors.

GPD's Submission

GPD welcomes the Australian Human Rights Commission's (AHRC) consideration of the impact of technology on human rights, and this open consultation process. We share the AHRC's belief that technology "should be shaped by human values, to protect and promote rights and freedoms",[1] a position reflecting the well-established and accepted international norm that "the same rights that people have offline must also be protected online".[2]

This submission aims to complement that work by expanding on the AHRC's initial examination of how the international human rights framework set out in section 3.1 of the Issues Paper is engaged by the rise of new technologies, particularly those related to the internet. This submission focuses on the impact of such new technologies on Article 19 of the International Covenant on Civil and Political Rights - the right to freedom of expression – answering questions 1, 3(b) and (c), 4, 5, and 6(a), (b) and (d).

[1] Australian Human Rights Commission, Human Rights & Technology, available at: https://tech.humanrights.gov.au/.

[2] This position was most recently reaffirmed by the UN Human Rights Council in Resolution 38/7, *The promotion, protection and enjoyment of human rights on the Internet*, UN Doc. A/HRC/RES/38/7, 17 July 2018.

Question 1: What types of technology raise particular human rights concerns? Which human rights are particularly implicated?

The impact of new technologies, and the internet in particular, on human rights cannot be understated and many of these impacts have been identified by the AHRC and set out in sections 3.1 and 3.4 of the Issues Paper. While new technologies have the potential to impact almost *all* human rights, this consultation response focuses in particular on the impact that they have on the right to freedom of expression. As such, and so to add to the examples in the Issues Paper, we would highlight new technologies such as online platforms and communication services, which have created a paradigmatic shift in our ability to enjoy and exercise many of our human rights, not only freedom of expression, but associated rights such as the rights to freedom of association and assembly. Coupled with this, the development of security tools, such as encryption, has enhanced our ability to exercise privacy online, providing safe and secure spaces for those who would risk persecution or discrimination were they to express themselves publicly.

Notwithstanding these benefits, there are particular concerns from a freedom of expression perspective over the ability of governments and private companies to control what content is and is not allowed online, what information appears when we search for information, and how we receive it. The use of algorithms and automated decisionmaking by platforms can lead to censorship, with particular impacts upon minority and vulnerable groups. Surveillance and other forms of monitoring people's online activities and behaviour represents not only a risk to the right to privacy, but has the potential to create a 'chilling effect' whereby individuals limit how they exercise their rights to freedom of expression, association and assembly.

There are a range of new technologies and by-products of news technologies that have the potential to restrict the enjoyment and exercise of the right to freedom of expression. These include:

- **Automation and algorithmic filtering to regulate content**: The regulation of online content by online platforms, particularly social media and search platforms, through automation and algorithmic filtering poses clear and well-documented risks to the right to freedom of expression, in particular for minority groups who are disproportionately affected by content removals.[3]

- **Artificial Intelligence, including machine learning**: The rise of techniques such as video surveillance, facial recognition and behaviour analysis by public authorities and private companies most obviously has the potential to infringe upon the right to privacy, but can also lead to self-censorship, altered behaviour in public places and private communications alike, effectively creating a 'chilling effect' on how individuals exercise their right to freedom of expression.[4]

Additionally, there are particular issues - related to new technologies - that can directly affect the right to freedom of expression, among them:

[3] See, for example, UN Human Rights Council, *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, UN Doc. A/HRC/32/38, 11 May 2016, Para 55; and Global Partners Digital, *A Rights-Respecting Model of Online Content Regulation by Platforms, May 2018*, pp. 10 and 22, available at: https://www.gp-digital.org/wp-content/uploads/2018/05/A-rights-respecting-model-of-online-content-regulation-by-platforms.pdf.

[4] Article 19 and Privacy International, *Privacy and Freedom of Expression in the Age of Artificial Intelligence*, 2018, p. 8, available at: https://www.article19.org/wp-content/uploads/2018/04/Privacy-and-Freedom-of-Expression-In-the-Age-of-Artificial-Intelligence-1.pdf.

- **Filtering**: The filtering of online content by governments, through firewalls or other tools, or with the assistance of the private sector, can lead to restrictions on what information people can receive, and what people can say online. While some may be permissible, some forms of content protected by the right to freedom of expression may also be restricted. As well as limiting that right for those within the particular jurisdictions, but also the rights of those outside of that jurisdiction who want to communicate or share information with them.[5]

- **Network disruptions**: Network disruptions – intentional state or state-sanctioned shutdowns, disruptions or other limitations of the internet, social media or other form of electronic communication – have increased dramatically across the world, and represent a clear restriction on the ability of individuals to communicate and seek and receive information.[6]

- **Inappropriate content regulation laws and regulations**: Although legislation and other forms of regulation can be used to tackle unlawful and harmful content online, inappropriate legislation and regulation, particularly those related to intermediary liability, can lead to unjustified limitations on freedom of expression and access to information.[7]

More indirectly, the right to freedom of expression through can also be limited by:

- **Inappropriate intellectual property frameworks and enforcement**: While there is a difficult balance to maintain between the protection and enforcement of intellectual property rights, and the right to freedom of expression, the potential harmful effects on the latter through inappropriate intellectual property frameworks and their enforcement has been highlighted recently through debates regarding the EU's new Copyright Directive.[8]

- **The erosion of net neutrality**: If internet service providers (ISPs) are given the ability to block or slow down certain types of traffic, they are effectively given the power to restrict access to online content. An ISP could, in theory, restrict access to sites or sections of the internet if its customers (or the sites themselves) do not pay a higher fee for access. This would fundamentally alter the nature of the internet as an enabler of the right to freedom of expression on an equal basis.

Example: The LGBTIQ Community

As stated above the process of blocking, filtering and removing content can have a disproportionate effect on minority and vulnerable groups. In some cases, initiatives that are intended to protect individuals from harmful or inappropriate sexual material may end up affecting other vulnerable groups such as lesbian, gay, bisexual and trans (LGBTIQ) individuals. Discussion of sexuality can share some of the features of sexual content and so automatic filtration or removal of content can cause particular harm to LGBTIQ individuals who need

[5] See, for example, UN Human Rights Council, *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, UN Doc. A/HRC/32/38, 11 May 2016, Paras 46-47.
[6] *Ibid*., Para 48.
[7] See, for example, Global Partners Digital, *A Rights-Respecting Model of Online Content Regulation by Platforms, May 2018*, pp. 29-31, available at: https://www.gp-digital.org/wp-content/uploads/2018/05/A-rights-respecting-model-of-online-content-regulation-by-platforms.pdf.
[8] See, for example, Rankin, J. "EU votes for copyright law that would make internet a 'tool for control'", *The Guardian*, 20 June 2018.

anonymity when seeking information, as well as in countries where LGBTIQ individuals are persecuted or subject to discrimination.

Question 3: How should Australian law protect human rights in the development, use and application of new technologies?

Australian law can protect human rights by translating existing, internationally agreed human rights law and standards into any regulation governing new technologies and by mainstreaming existing human rights protections into laws, policies and procedures dealing with technology.

Question 3(b): What can we learn about the need for regulating new technologies, and the options for doing so, from international human rights law and the experiences of other countries?

International human rights law

International human rights law provides a rich, practical basis for assessing the impact of new technologies. When taken alongside the norms and standards that complement treaty provision themselves, the international human rights framework offers guidance on how to ensure protection of the right to freedom of expression, particularly when seeking to balance that right against other legitimate state and public interests.

The value of the framework is especially pertinent in guiding the regulation of new technologies because it is comprised of established norms that have been negotiated and agreed internationally, and are thus universal, making them particularly well-suited for the internet and other digital technologies which are global in nature and use.

While the core international human rights treaties make no explicit reference to new technologies, their interpretation and elaboration, particularly by the UN Human Rights Council, Treaty Bodies and Special Procedures, provides a wide source of material demonstrating the connections between new technologies and human rights, the risks that can result from an absence of – or poor – regulation, and the obligations on states to ensure that human rights are protected in the digital environment. These include:

- UN Human Rights Council Resolution on the promotion, protection and enjoyment of human rights on the Internet;[9]
- UN Human Rights Council Resolution on the promotion and protection of human rights in the context of peaceful protests;[10]
- UN Human Rights Council Resolution on accelerating efforts to eliminate violence against women and girls: preventing and responding to violence against women and girls in digital contexts;[11]

[9] UN Human Rights Council, *The promotion, protection and enjoyment of human rights on the Internet*, UN Doc. A/HRC/RES/38/7, 17 July 2018.
[10] UN Human Rights Council, *The promotion and protection of human rights in the context of peaceful protests*, UN Doc. A/HRC/RES/38/11, 16 July 2018. The UN Human Rights Council resolution recognise that although an assembly has generally been understood as a physical gathering of people, human rights protections, including for the rights to freedom of peaceful assembly, of expression and of association, may apply to analogous interactions taking place online.
[11] UN Human Rights Council, Accelerating efforts to eliminate violence against women and girls: preventing and responding to violence against women and girls in digital contexts, UN Doc. A/HRC/RES/38/5, 17 July 2018.

- UN Human Rights Council Resolution on the Right to Privacy in the Digital Age;[12]
- UN Human Rights Committee General Comment No. 16: Article 17 (Right to privacy);
- UN Human Rights Committee General Comment No. 34: Article 19: Freedoms of opinion and expression;
- Report of the United Nations High Commissioner for Human Rights: Promotion, protection and enjoyment of human rights on the Internet: ways to bridge the gender digital divide from a human rights perspective;[13]
- Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (a human rights approach to platform content regulation);[14]
- Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (on the roles played by private actors engaged in the provision of Internet and telecommunications access);[15]
- Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (ways in which the information and communications technology sector implicates freedom of expression);[16] and
- Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (encryption and anonymity);[17]

Example: Online Content Regulation by Platforms

In societies where the offline exercise of the right to freedom of expression is constrained by censorship or state regulation of the media, online platforms may be one of the only ways that individuals are able to exercise that right. However, platforms develop their own Terms of Service in order to determine what content they will and will not allow which may be based on their own values, or on vague or arbitrary criteria. This can result, with documented instances, in the removal of content which is protected by the right to freedom of expression, or removals which disproportionately affected minority and vulnerable groups. The scale of inappropriate content regulation is not fully known, partly as a result of the lack of any meaningful transparency about moderation decisions from the online platforms themselves. This lack of transparency also reinforces the difficulty of ensuring awareness of when and why mistakes have been made.

In setting out a framework for online content regulation by platforms, international human rights law and standards provide a clear basis for determining what content should and should not be permitted and can serve as a meaningful check on both government and private sector decision making. A number of human rights-based models for online content regulation by platforms, as well as the appropriate national legal and regulatory frameworks that should complement and facilitate them, have been put forward by, *inter alia*, the UN Special Rapporteur

[12] UN Human Rights Council, *The right to privacy in the digital age*, UN Doc. A/HRC/RES/34/7, 7 April 2017.
[13] UN Human Rights Council, *Report of the United Nations High Commissioner for Human Rights: Promotion, protection and enjoyment of human rights on the Internet: ways to bridge the gender digital divide from a human rights perspective*, UN Doc. A/HRC/RES/35/9, 5 May 2017.
[14] UN Human Rights Council, *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, UN Doc. A/HRC/38/35, 6 April 2018.
[15] UN Human Rights Council, *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, UN Doc. A/HRC/35/22, 30 March 2017.
[16] UN Human Rights Council, *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, UN Doc. A/HRC/32/38, 11 May 2016.
[17] UN Human Rights Council, *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, UN Doc. A/HRC/29/32, 22 May 2015.

on the promotion and protection of the right to freedom of expression and opinion, and civil society organisations including GPD and Article 19.[18]

Lessons from other countries

Although innovative regulatory regimes can be used to address the potential harmful effects of new technology, global experience in recent years has demonstrated that to be effective, new regulatory regimes must be accompanied by strict human rights protections to ensure that they do not have adverse or unintended effects on the individual or society as whole. This is particularly true when seeking to tackle the harms caused by unlawful and harmful content online.

Example: NetzDG in Germany

In recent years, a number of governments have attempted to grapple with regulating online content, and - although no single approach has met with universal support - one clear theme that has emerged is that any that regime aiming to regulate online content must both incorporate international human rights standards and provide accountability mechanisms that are consistent with those standards.

One notable example of a failure to ensure such safeguards is the Network Enforcement Act (or NetzDG) adopted in Germany in 2017. The NetzDG sought to eliminate hate speech and other forms of unlawful content on online platforms, imposing fines of up to 50 million euro on large platforms for failure to take down "manifestly unlawful" content within 24 hours. Since its introduction, Twitter started deleting posts from a far-right politician who referred to "barbaric, gang-raping hordes of Muslim men"[19] as well as other controversial or satirical – but lawful – tweets, with one of the users implicated pointing out that before its inception she had "tweeted things that were significantly more extreme" without being blocked.[20]

Regulation like the NetzDG forces private businesses to make decisions about what is legal or illegal – which is primarily the role of courts and other public authorities. Restrictions, including the removal of content, should only take place following a clear, transparent and rights-respecting process, with appropriate accountability and the possibility of an independent appeal process. And tight time limits and high sanctions risk incentivising the removal of content which might in fact be lawful. In cases where platforms are unsure whether content is unlawful or harmful, many will play it safe and simply delete the content rather than risk a fine.

There is a further danger with such legislation and regulation since while the state may have strong domestic human rights frameworks which mitigate some of the risks, there is a pattern of states across the world adopting the legal frameworks of others to deal with concerns, but without any equivalent effective human rights protections, exacerbating the risks. As Eileen

[18] UN Human Rights Council, *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, UN Doc. A/HRC/38/35, 6 April 2018; Global Partners Digital, *A Rights-Respecting Model of Online Content Regulation by Platforms*, May 2018, available at: https://www.gp-digital.org/wp-content/uploads/2018/05/A-rights-respecting-model-of-online-content-regulation-by-platforms.pdf; and Article 19, *Self-regulation and 'hate speech' on social media platforms*, 2018, available at: https://www.article19.org/wp-content/uploads/2018/03/Self-regulation-and-%E2%80%98hate-speech%E2%80%99-on-social-media-platforms_March2018.pdf.
[19] BBC News, "German AfD MPs under fire for anti-Muslim New Year's Eve messages", *bbc.co.uk*, 2 January 2018, available at: https://www.bbc.co.uk/news/world-europe-42537656.
[20] Scott, M. and Delcker, J., "Free speech vs. censorship in Germany", *Politico*, 6 January 2018, available at: https://www.politico.eu/article/germany-hate-speech-netzdg-facebook-youtube-google-twitter-free-speech/.

Donahoe has identified, democratic values are at risk of serious erosion when content regulation regimes such as the German NetzDG law are translated into jurisdictions that do not mainstream human rights protections and do not use the international human rights framework as an effective basis for public dialogue:

"Within two weeks of the adoption of the German law, the Russian Duma proposed a copy-cat bill, with multiple explicit references to the German law as its model. The Russian version, like the German original, compels social media companies to take down vaguely defined "illegal" content within twenty-four hours or face severe penalties. The official justification for the law was to prevent use of digital networks for "illegal" purposes. In Russia, this can mean anything that challenges the authoritarian rule of Vladimir Putin. Russia's cynical use of Germany's example should raise alarm bells for all democratic actors."[21]

Question 3(c): What principles should guide regulation in this area?

It is essential that any legislative or regulatory frameworks developed which apply to online platforms or communications services, or other new technologies which engage the right to freedom of expression, facilitate, rather than undermine the enjoyment of that right.

One form of troublesome regulation which has been seen elsewhere, and is being proposed in different jurisdiction, is that which attaches liability to platforms for content which is available on them, which can lead to a 'chilling effect' in which platforms either become reluctant to host or otherwise make available content, or are overly zealous in removing content which might be harmful.

There are, at present, a range of liability regimes which fall within three broad categories, outlined in the below table.

Liability regime	Summary	Examples
Strict liability	Platforms are held liable for unlawful or harmful content made available by users on their platforms, even if they are not aware of the content.	Thailand (Section 15 of the Computer Crimes Act 2007)
Conditional liability/ 'safe harbour'	Platforms are not held liable for unlawful or harmful content made available by users on their platforms provided they do not have any knowledge of the content or, if they do have knowledge, have acted expeditiously to remove that content.	European Union (Article 14 of the E-Commerce Directive)
Broad immunity	Platforms are, as a general rule, not held liable for unlawful or harmful content made available on their platforms, even if they are aware of the content. Some limited exceptions may exist, such as for certain specified crimes or intellectual property.	USA (Section 230 of the Communications Decency Act)

[21] Donahoe, E., "Protecting Democracy from Online Disinformation Requires Better Algorithms, Not Censorship", 2017, *Council on Foreign Relations*, available at: https://www.cfr.org/blog/protecting-democracy-online-disinformation-requires-better-algorithms-not-censorship.

'Strict liability' regimes are the most likely to result in overly broad restrictions of freedom of expression, as they require the platform proactively to monitor and remove content, even without notification. However, even 'safe harbour' or 'conditional liability' regimes can be problematic particularly where the conditions under which liability will be held are such that they require a platform to make determinations about the lawfulness of content, to remove content within short time limits or impose high sanctions for a failure to take down content. In such circumstances, there is a clear incentive on platforms to 'play it safe' and remove ambiguous content so as to avoid liability and potential fines or other sanctions. One example of such a liability regime is the NetzDG in Germany discussed above in our response to question 3(b).

While we do not consider that intermediaries should never be liable for content which is made available on their platforms, we consider that there must be sufficient limitations and safeguards in place when it comes to attaching liability to ensure that risks to freedom of expression through incentives to remove content are effectively mitigated. We believe that such a regime is feasible through compliance with the following principles:

- First, the development of any regulation which attaches liability to platforms should be open, inclusive and transparent. The development process should include consultation with all relevant stakeholders and states should consider undertaking a human rights impact assessment to understand the impact that the legislation may have on human rights.

- Second, the regulation itself should be consistent with the principle of legal certainty. This means that it should be accessible, and sufficiently clear and precise for platforms, users and other interested groups to be able to regulate their conduct in accordance with the law.

- Third, the regulation should not directly or indirectly impose a general obligation on platforms to monitor third party content where they do nothing more than host that content, or transmit or store it, whether by automated means or not. Further, the regulation should not attach strict liability to a platform for hosting unlawful content as this would, de facto, require such monitoring.

- Fourth, the regulation should not directly or indirectly impose liability on platforms for third party content where they do nothing more than host that content, or transmit or store it, whether by automated means or not, and have no actual knowledge of specific content thereby hosted, transmitted or stored. Indeed, the regulation should explicitly exempt platforms from liability in such circumstances.

- Fifth, the regulation should not attach liability to platforms for failing to restrict lawful content.

- Sixth, the regulation should not provide any incentives to remove content which may be lawful, such as via unrealistic timeframes for compliance, or the imposition of disproportionate sanctions for non-compliance.

Question 4: In addition to legislation, how should the Australian Government, the private sector and others protect and promote human rights in the development of new technology?

In order to ensure the promotion and protection of human rights in the development of new technology, governments should work with the private sector and civil society to develop and implement frameworks for the promotion and protection of human rights in the private sector, where most such technology is developed. Three examples of how can be achieved include:

- Effective implementation of UN Guiding Principles for Business and Human Rights
- Development and implementation of sectoral standards
- Incorporating human rights considerations into public procurement processes

Effective Implementation of UN Guiding Principles on Business and Human Rights

Although the UN Guiding Principles for Business and Human Rights (UNGP) affirm the well-established principle that the primary human rights obligations rest with governments, they also affirm that private businesses have a responsibility to respect human rights. In light of the speed of technological change, the pace of innovation of products and services by the private sector, the importance of the full and effective implementation of the UNGP cannot be overstated. The UNGP constitute a ready-made and internationally agreed set of norms for guiding private sector decisionmaking on issues that impact upon human rights, and set out the role that states and businesses have in respecting and upholding human rights, under the "Protect, Respect and Remedy" framework.

The government of Australia has endorsed the UNGP but has not yet developed a National Action Plan (NAP) which is one of the primary means that a government can show its commitment to the UNGP and set out concrete steps on how they will be implemented in the national framework. The government of Australia should revisit its decision not to develop a NAP, and set out a process for the development of one, explicitly including recognition of the human rights impacts of new technologies developed by the private sector.
At the same time, the comprehensive acceptance and implementation of the UNGP by the private sector in Australia would represent a significant step in ensuring that human rights were protected and promoted in the development of new technology. To support the private sector, the Australian government or another relevant actor should develop specific guidance on the application of the UNGP to new technologies. Examples of similar guidance developed in other jurisdictions exists, such as the European Commission's ICT Sector Guide on Implementing the UN Guiding Principles on Business and Human Rights.[22]

Development and Implementation of Sectoral Standards

Private businesses can also help ensure the protection and promotion of human rights through working together to develop industry or sectoral standards or commitments.

Example: Global Network Initiative Principles on Privacy and Freedom of Expression

The Global Network Initiative (GNI) is a multi-stakeholder platform (including companies, human rights and press freedom organisations, academics and investors).[23] GNI participants work together in two mutually supporting ways: by implementing the GNI Principles and adhering to their Implementation Guidelines and by collectively advocating to governments and

[22] European Commission, *ICT Sector Guide on Implementing the UN Guiding Principles on Business and Human Rights*, 2013.
[23] Global Network Initiative, available at: https://globalnetworkinitiative.org/.

international institutions for laws and policies that promote and protect freedom of expression and privacy. These GNI instruments provide a framework for responsible company decision-making in support of free expression and privacy rights, wherever they operate.

In its preamble, the GNI Principles explicitly recognise that "ICT companies have the responsibility to respect and promote the freedom of expression and privacy rights of their users. ICT has the potential to enable the exchange of ideas and access to information in a way that supports economic opportunity, advances knowledge and improves quality of life. By implementing these Principles, ICT companies can also work to protect, promote and support human rights, including through improved responsible decision-making, shared learning and multi-stakeholder collaboration".

Incorporating Human Rights Considerations into Public Procurement Processes

Governments can also help reduce the risk of human rights abuses by the private sector by including an assessment of companies' adherence to international human rights standards in their public procurement and tender processes. Such objectives may be achieved, for example, by including "social clauses" in public procurement contracts.[24]

Examples of good practice include:

- **Denmark**: Denmark's National Action Plan: Implementation of the UN Guiding Principles on Business and Human Rights, includes as one of its goals, the inclusion of more voluntary social clauses in connection with public tenders, and to develop cases demonstrating how companies and municipalities work with social clauses in practice.[25]

- **The Netherlands**: The Netherlands' National Action Plan on Business and Human Rights highlights that: "companies supplying the government with goods and services are required to respect human rights" under the "social conditions" of the existing national sustainable procurement policy included in all central government EU contract award procedures which have been in place since 1 January 2013.[26]

- **Norway**: The Norwegian National Action Plan on Business and Human Rights explicitly connects government procurement to the state duty to protect human rights and to promote respect for human rights by companies they transact with, and highlighted its consultation on amending national public procurement legislation to include a provision stating that contracting authorities should have adequate procedures for ensuring social responsibility in connection with.[27]

- **Finland**: The Finnish National Action Plan on Business and Human Rights includes, as one of three key aims, the application of social criteria in public procurement, and proposes a range of measures to support this objective, including updating the state

[24] International Learning Lab on Public Procurement and Human Rights, *Public Procurement and Human Rights: A Survey of Twenty Jurisdictions*, July 2016, available at: http://www.hrprocurementlab.org/wp-content/uploads/2016/06/Public-Procurement-and-Human-Rights-A-Survey-of-Twenty-Jurisdictions-Final.pdf.
[25] Government of Denmark, *Danish National Action Plan – Implementation of the UN Guiding Principles on Business and Human Rights*, 2014, p. 97.
[26] Government of Netherlands, *National Action Plan on Business and Human Rights*, 2013, p. 17.
[27] Government of Norway, *National Action Plan on Business and Human Rights*, 2015, p. 25.

procurement manual's "responsibility themes" and producing a report on high-risk product groups.[28]

Question 5. How well are human rights protected and promoted in AI-informed decision making? In particular, what are some practical examples of how AI-informed decision making can protect or threaten human rights?

The benefits of grounding decisions in mathematical equations, machine learning and deep learning appear to be promising in many areas and, if used appropriately, AI-informed decisionmaking has the potential to offer a number of benefits to society. At present, algorithms and AI-informed decisionmaking are now widely used by both government and the private sector to make decisions that affect different aspects of people's lives - from the assessment of social security entitlements, to risk assessments within the criminal justice system and the allocation of credit by banks.

The potential for AI to impact on human rights varies depending on the application of the technology. The following table represents a summary of examples taken from Business for Social Responsibility's paper, "Artificial Intelligence: A Rights-Based Blueprint for Business Paper No. 2"[29] and Article 19 and Privacy International's paper "Privacy and Freedom of Expression in the Age of Artificial Intelligence".[30]

Sector	Protect	Threaten
Financial Services Example: The use of algorithms to inform decisionmaking about clients, e.g. whether to provide access to credit, what rate to provide it at, the pricing of insurance of products based on a more sophisticated understanding of risk.		**Privacy**: The collection and use of large amounts of personal data for analysis. **Non-discrimination**: Decisions relating to credit and insurance determined by algorithms risks perpetuating or exacerbating existing forms of discrimination and inequality based on characteristics.
Healthcare Example: Using AI to provide more effective evidence-based treatment plans and learn which types of treatments will be most effective for different patients.	**Right to the highest attainable standard of health / non-discrimination**: Tackling health inequalities by improved access to information, analysis and guidance for patients.	**Privacy**: AI requires access to data contained in medical records which will often be highly sensitive, personal, and confidential information.

[28] Government of Finland, *Ministry of Employment and the Economy, National Action Plan for the Implementation of the UN Guiding Principles on Business and Human Rights*, 2014, p. 14.
[29] BSR, *Artificial Intelligence: A Rights-Based Blueprint for Business: Paper 2: Beyond the Technology Industry*, August 2018, available at: https://www.bsr.org/reports/BSR-Artificial-Intelligence-A-Rights-Based-Blueprint-for-Business-Paper-02.pdf.
[30] Article 19 and Privacy International, *Privacy and Freedom of Expression in the Age of Artificial Intelligence*, 2018, available at: https://www.article19.org/wp-content/uploads/2018/04/Privacy-and-Freedom-of-Expression-In-the-Age-of-Artificial-Intelligence-1.pdf.

Sector	Protect	Threaten
Retail Examples: Product design, product functionality, manufacturing, checkout, targeted advertising, store design, and distribution.		**Privacy**: The collection and use of consumer data. The development of AI-enabled products also pose risks to privacy, especially for children. **Non-discrimination**: Targeted advertising can be discriminatory.
Transport and logistics Example: Self-driving vehicles and unmanned aircraft systems (i.e., drones).		**Privacy**: Self-driving vehicle services may collect data about who is travelling where, when, and with whom. **Privacy**: Drones may use recording and sensory devices. **Right to life**: An autonomous which is hacked could put the passengers' lives in danger.
Security Example: The increased use in techniques such as video surveillance, facial recognition, behaviour analysis etc., by public authorities and private companies.		**Privacy**: Mass surveillance is a disproportionate interference with the right to privacy. **Freedom of expression**: Mass surveillance and AI can also have a chilling effect on exercise of the right to freedom of expression.

Question 6: How should Australian law protect human rights in respect of AI-informed decision making? In particular:

Question 6(a): What should be the overarching objectives of regulation in this area?

AI-informed decisionmaking raises a host of human rights considerations and any regulatory responses should therefore be guided by international human rights law and standards, for the reasons set out earlier in this submission. We therefore believe that an overarching objective of regulation in the area of AI-informed decisionmaking should be to ensure that it is developed and used in a manner which enables, rather than harms, the enjoyment and exercise of human rights.

Question 6(b): What principles should be applied to achieve these objectives?

We would point to a number of sets of principles which have been developed in recent months and years which indicate the approach that governments and businesses should take with regards to the regulation and use of AI to ensure consistency with human rights.
These include:

- The Toronto Declaration: Protecting the right to equality and non-discrimination in machine learning systems;[31]
- The 10 Beliefs for the responsible development and use of AI set out in BSR's paper, 2Artificial Intelligence: A Rights-Based Blueprint for Business Paper 1: Why a Rights-Based Approach?";[32]
- The Asilomar AI Principles (particular the section on Ethics and Values);[33]
- The Partnership on AI;[34]
- Microsoft's AI Principles;[35] and
- Google's principles for the responsible development of AI.[36]

From these, a number of consistent themes have emerged which we would summarise as follows:

- The design, development and use of AI systems should be done in ways that fully respects international human rights law and standards.

- Special attention should be paid to nexus between the private sector and the public sector, where private sector AI systems are used by public authorities or to deliver public services.

- Where a state uses AI systems, it should:
    - thoroughly investigate the systems for discrimination and other risks to human rights prior to its development or acquisition and on an ongoing basis throughout the lifecycle of the systems;
    - ensure and require accountability and maximum possible transparency around public sector use of AI systems. This must include ensuring that the systems are sufficiently understood so that their impact on affected individuals and groups can be effectively scrutinised by independent entities, responsibilities established, and actors held to account; and
    - take steps to ensure public officials are aware of and sensitive to the risks of discrimination and other rights harms in machine learning systems, and create mechanisms for independent oversight, including by judicial authorities when necessary.

- Private companies should, in designing, developing and applying AI systems:
    - identify potential discriminatory outcomes;

[31] The Toronto Declaration: Protecting the right to equality and non-discrimination in machine learning systems, available at: https://www.accessnow.org/cms/assets/uploads/2018/08/The-Toronto-Declaration_ENG_08-2018.pdf.
[32] BSR, Artificial Intelligence: A Rights-Based Blueprint for Business Paper 1: Why a Rights-Based Approach?, available at: https://www.bsr.org/en/our-insights/report-view/artificial-intelligence-a-rights-based-blueprint-for-business.
[33] The Asilomar AI Principles, available at: https://futureoflife.org/ai-principles.
[34] Partnership on AI, available at: https://www.partnershiponai.org/.
[35] Microsoft AI principles, available at: https://www.microsoft.com/en-us/ai/our-approach-to-ai.
[36] AI at Google: our principles, available at: https://blog.google/technology/ai/ai-principles/.

- take effective action to prevent and mitigate discrimination and track responses; and
- be transparent about efforts to identify, prevent and mitigate against discrimination in machine learning systems.

- Those whose human rights have been adversely impacted, even if unintentionally, by the use of AI systems should have access to grievance mechanisms and effective remedies. This may include, for example, creating clear, independent, visible processes for redress following adverse individual or societal effects, and designating roles in the entity responsible for the timely remedy of such issues subject to accessible and effective appeal and judicial review.

Question 6(c): What can we learn from how other countries are seeking to protect human rights in this area?

Currently, no country has specific legislation regulating AI and its use. Instead, aspects of AI and its use tend to be regulated, if at all, by more general legal frameworks relating to freedom of expression, data protection, consumer protection, media and competition. There may also be sectoral regulation and standards. While they are few, there are some examples of how human rights are protected through these forms of regulation and other standards when it comes to AI.

Example: European Union Declaration of Cooperation on Artificial Intelligence

In April 2018, 25 European countries signed a Declaration of Cooperation on Artificial Intelligence.[37] While a fuller ethical framework on AI developments, is expected to be published by the European Commission by the end of 2018, the Declaration of Cooperation recognises the need to develop "an adequate legal and ethical framework, building on EU fundamental rights and values, including privacy and protection of personal data, as well as principles such as transparency and accountability". The Declaration also commits the signatories to "ensure that humans remain at the centre of the development, deployment and decision-making of AI, prevent the harmful creation and use of AI applications, and advance public understanding of AI".

Example: European Union General Data Protection Regulation

The European Union's General Data Protection Regulation[38] contains a number of provisions providing for safeguards for data subjects when automated decisions are made affecting them. Most significantly, Article 22(1) provides that data subjects have a right "not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her". While there are some, limited exceptions to this general right, even where these apply, there must still be "suitable measures to safeguard the data subject's rights, freedoms and legitimate interests".

To complement this right, Articles 13(2) and 14(2) set out a general requirements for data controllers, when obtaining personal data, whether from a data subject or otherwise, to inform data subjects of the existence of any automated decision-making as well as "meaningful information about the logic involved, as well as the significance and the envisaged consequences

[37] EU Declaration on Cooperation on Artificial Intelligence, available at: https://ec.europa.eu/jrc/communities/community/digitranscope-digital-transformation-and-governance-human-society/document/eu-declaration.
[38] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

of such processing for the data subject". Further, under Article 15(1), data subjects have the right "to obtain from the controller confirmation as to whether or not personal data concerning him or her are being processed, and, where that is the case, access to the personal data and (…) the existence of automated decision-making and (…) meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject".