Developing Standards for Artificial Intelligence: Hearing Australia's Voice

University of Melbourne Response

July 2019

Overview

The University of Melbourne welcomes the opportunity to respond to Standards Australia's Discussion Paper, *Developing Standards for Artificial Intelligence: Hearing Australia's Voice*. With global processes regarding the standardisation of Artificial Intelligence (AI) currently underway, it is timely that Standards Australia is seeking feedback from stakeholders on how standards and related material can support AI in Australia.

The University broadly supports the need for standardisation and acknowledges the role that standards, as an adaptive form of regulation, can play in enabling interoperability, supporting the adoption of new technologies, and ensuring quality and public trust in products and services. The Discussion Paper highlights the work Standards Australia is doing to give voice to Australian stakeholders in global consultations on standardisation, and the University is supportive of this.

This submission does not attempt to address all the issues raised in the Discussion Paper. Instead, we offer some high-level comments concerning the broad approach that ought to be taken to the development of AI standards. These comments address four key areas:

- the need for international harmonisation;
- the importance of developing standards which ensure safe, trustworthy and reliable AI, while not hindering innovation and the development of new technology;
- the need for the AI Standards Roadmap to integrate with related work such as the Australian Competition and Consumer Commission's Digital Platforms Inquiry, the Australian Human Rights Commission's consultation on Human Rights and Technology and the AI Technology Roadmap prepared by Data61; and,
- the importance of adopting a multi-disciplinary approach to responding to the challenges and opportunities presented by AI.

We would welcome the opportunity to discuss these issues further with Standards Australia or to otherwise assist with the development of the AI Standards Roadmap.

For more information, or to discuss the submission, Professor Liz Sonenberg, Pro Vice-Chancellor (Research Infrastructure and Systems) and Pro Vice-Chancellor (Digital and Data), can be contacted at l.sonenberg@unimelb.edu.au or on (03) 8344 4447.

General Comments

1. International harmonisation

International harmonisation ought to be central to the development of standards for AI, with proper integration essential to Australia enjoying the full benefits of new technologies. Achieving international harmonisation will help promote a wider and more interoperable market, while also helping to ensure consumers can have confidence and trust in AI technology.

The University therefore welcomes Standards Australia's intention to work constructively with the International Organisation for Standardization (ISO) and the International Electrotechnical Commission's (IEC) Joint Technical Committee 1's Subcommittee 42 – Artificial Intelligence (JTC1/SC42). The focus of JTC 1/SC42 on areas including trustworthiness, bias in AI systems, risk management, big data reference architecture, and governance implications of the use of AI reflects elements of AI that are already being considered through a range of mechanisms in Australia.

Championing an Australian voice on global standards

The University supports Australia's active engagement in the development of international standards in AI, with a view to these being appropriate for adoption in Australia.

Standards Australia's decision to create a mirror committee to JTC1/SC42 brings an Australian voice to conversations around global standards. It provides on opportunity to express matters that may be particular to Australia, including insights from recent consultations and inquiries and, in turn, ensure that Australia is best placed to leverage the learnings from international innovation, ethics and regulation of AI. It is essential that this committee can proactively communicate information on international standards development back to Australian industry, government and academia. This should be a feature of the AI Standards Roadmap.

Learning from international experience

The AI Standards Roadmap must also have regard for the work being conducted by other countries in relation to standards development and frameworks for AI. By taking a proactive approach to understanding what is being done around the world, Australia may draw learnings or be better placed to anticipate and respond to changes. We note the National Institute of Standards and Technology (NIST) in the United States recently released its draft *Plan for Federal Engagement in Developing AI Technical Standards and Related Tools*.[1] The draft plan recommends that the U.S should "strategically engage with international parties to advance AI standards for U.S. economic and national security needs," and the same can be said for Australia. As highlighted by Intel in its response to NIST's request for information, international standards are especially important for areas that benefit from a consistent global approach such as those related to technical interoperability, reliability, safety and trustworthiness.[2] Not being involved in processes to develop international standards risks Australia's interests being overlooked and may result in disadvantage to both Australian industry and consumers.

---

[1] National Institute of Standards and Technology, June 2019, 'U.S. Leadership in AI: Plan for Federal Engagement in Developing AI Technical Standards and Related Tools - Draft Plan for Consultation.' Available from: https://www.nist.gov/sites/default/files/documents/2019/07/02/plan_for_ai_standards_publicreview_2july2019.pdf Accessed July 22, 2019

[2] Intel, June 2019. 'Intel response to NIST RFI: Developing a Federal AI Standards Engagement Plan.' p3, Available from https://www.nist.gov/sites/default/files/documents/2019/06/11/nist-ai-rfi-intel-001.pdf Accessed July 22, 2019

Recommendations

The University of Melbourne recommends:

- Standards Australia should continue to act as Australia's voice in international processes regarding standardisation of AI, helping to develop international standards that are appropriate for adoption in Australia.
- The AI Standards Roadmap should include a mechanism to monitor the work on AI being conducted globally and in other countries, ensuring Australia is well-placed to anticipate and respond to changes.

2. Innovation and standards development

Standards should aim to protect Australian consumers by establishing requirements or guidelines for functional, reliable and trustworthy AI systems, while still fostering an environment of innovation and a competitive marketplace for new technologies.

Ensuring timely implementation of standards

The timing of standards development is significant, and this should be reflected in the AI Standards Roadmap. If standards are applied where technology is in too early a stage of development or where not enough robust research has been conducted, then standards are at risk of not being fit-for-purpose, being ignored and hindering innovation. If standards are developed too late, they may be difficult to apply retrospectively to existing technologies already operating in the market. Once developed, standards need to be regularly reviewed and updated to ensure they keep pace with evolving technology and public expectations.[3]

Deepening research and understanding of AI

A theme apparent in international approaches to standards development is the need for more research in specific areas for inclusion in standards.[4] A recommendation in the NIST draft paper, for example, is to "promote focused research to advance and accelerate broader exploration and understanding of how aspects of trustworthiness can be practically incorporated within standards and standards-related tools."[5] In their response to NIST's earlier request for information, Intel also notes a number of areas requiring research to inform standardisation. These include understanding specific attacks to understand trustworthiness requirements for AI, data security specific to machine learning based AI and societal issues such as explainability, transparency and ethics to understand the requirements for technical standards.[6]

We would note that universities are well-placed to conduct this type of research that may in turn help inform standards development. Universities have an important role to play in the responsible development of new technology. Publically funded research can be a key enabler of innovation and can investigate ways to address challenges that are associated with emerging technologies such as AI.

[3] NIST draft plan for consultation, p6, (cited previously)
[4] See, ISO/IEC JTC 1/ SC42 and NIST Draft plan for consultation p17 (cited previously)
[5] NIST draft plan for consultation, p17 (cited previously)
[6] Intel response to NIST RFI pp17 -19 (cited previously)

Recommendations

The University of Melbourne recommends:

- The AI Standards Roadmap should have regard to timing of standards development, with a view to ensuring any implementation enhances, rather than stifles, innovation.
- The AI Standards Roadmap should identify areas where more research is required prior to standards being developed and note opportunities where Australian researchers can assist in addressing these gaps.

3. Integration with related work

It is important that the AI Standards Roadmap is properly integrated into the other relevant work being undertaken on artificial intelligence. As discussed, this includes ensuring that Australia's AI standards are appropriately aligned with those of other countries. In addition, attention should be given to integration on a domestic level. The Discussion Paper notes that this project has been commissioned alongside the AI Ethics Framework and the AI Technology Roadmap, both to be prepared by Data61. It also acknowledges the Australian Human Rights Commission's consultation on Human Rights and Technology. The University of Melbourne has provided responses to several of these consultations, which are attached to this submission. While the focus differs across each of these projects, they nonetheless address a number of common issues concerning the challenges of AI. There is a need to approach each of the projects in a holistic manner.

Articulating human rights

The relationship between the human rights framework being developed by the Australian Human Rights Commission and the AI Standards Roadmap is particularly important: there should be clarity around this relationship. By definition, a human rights framework articulates a set of requirements that all relevant actors are stringently obligated to comply with. It therefore differs from a set of standards which are "*voluntary* documents" with which there "is no requirement for the public to comply."[7] This suggests that the human rights framework is the more fundamental of the two. This should be made clear in the AI Standards Roadmap itself. For example, it is important that the roadmap does not imply that specifications relating to automated decision-making prescribed by the human rights framework are merely voluntary.

Integrating with standards for foundational and related technologies

A related issue concerns the integration of the AI standards to those of related technologies. As Intel notes in its response to the NIST Federal Engagement Plan, the "current success of machine-learning based AI is predicated on significant advancements in foundational technologies, including computing power, storage, networking bandwidth, battery power, software and hardware architectures and many others."[8] There exists already a relatively mature standards framework that supports interoperability and trustworthiness for these technologies. Consideration should be given to how the AI standardisation framework is to interact with these, and to the extent to which these standards can be "extended to support evolving requirements for AI, including emerging standards that address governance and other aspects for trustworthy AI systems."[9]

[7] Standards Australia Discussion Paper page 13, emphasis added
[8] Intel response to NIST RFI, page 5 (cited previously)
[9] Intel response to NIST RFI, page 6 (cited previously)

Coordinating with the ACCC Digital Platforms Inquiry

The AI Standards Roadmap should also align with the work being conducted by the Australian Competition and Consumer Commission (ACCC) as part of its Digital Platforms Inquiry, something which is not referenced in the Standards Australia Discussion Paper. In its recently released report, the ACCC makes a number of recommendations in relation to promoting consumer protection in data collection and developing automated decision-making technologies. The recommendations include strengthening protections in the Privacy Act, broader reform of Australian privacy law, a statutory tort for serious invasions of privacy and prohibition against unfair contract terms.[10] The Australian Consumer Law and suggested reforms in the report may provide 'safety net' protection for consumers while standards are being developed and complement standards that do exist. In developing its AI Standards Roadmap, Standards Australia should have regard for the issues and recommendations raised in this work and should monitor the pending response to the ACCC report from the Australian Government.

Defining Artificial Intelligence

In previous submissions to public consultations on AI and new technologies, the University of Melbourne has raised concerns about the difficulty of defining AI for the purposes of regulation and law-making. We again make this point in relation to the development of the AI Standards Roadmap. There is a lack of a clear or widely accepted definition of AI, which makes the development of AI specific standards challenging. In addition, from a legal perspective, the distinction between AI algorithms and other decision-making algorithms is irrelevant, and any regulatory response should be independent of algorithmic implementation details.

A focus instead on standards for 'automated decision-making' or even more generally 'software' may provide for a cleaner definition. This would encompass techniques used in AI such as data-driven machine learning, data mining, optimisation, search and rule-based systems. As an example, in considering the OECD AI principles referenced in the Discussion Paper, if one was to substitute 'software' for AI, then these principles will still make sense. Such considerations should feed into the development of the AI Standards Roadmap and Standards Australia's input into the development of international standards.

Recommendations

The University of Melbourne recommends:

- Standards Australia has due regard to the related work being undertaken in Australia in relation to Artificial Intelligence in the development of its AI Standards Roadmap.
- The AI Standards Roadmap makes clear its relationship with the human rights framework being developed by the Australian Human Rights Commission and explicitly note that specifications prescribed by this framework are not voluntary.
- Standards Australia considers more broadly defining the scope of applicability of the roadmap analysis, adopting terminology such as 'automated decision-making' rather than 'AI'.

[10] Australian Competition and Consumer Commission, July 2019. *Digital Platforms Inquiry. pp456 -501.* Available from: https://www.accc.gov.au/system/files/Digital%20platforms%20inquiry%20-%20final%20report.pdf Accessed July 2019.

4. A multi-disciplinary approach

In our contributions to previous public consultations on AI and related technology, the University of Melbourne has emphasised the importance of adopting a multi-disciplinary approach to responding to the challenges and opportunities associated with new technologies. We have also reflected such an approach in our own engagement with these consultations. Each of our written submissions have drawn from experts from a range of fields within the university, whose insights are relevant to these challenges and opportunities.

This point is again important in the context of the development of the AI Standards Roadmap. The Discussion Paper touches on issues that are legal, ethical and economic as well as technological. The development of the roadmap and subsequent work on standards should therefore draw on the knowledge of researchers from a broader set of fields. For example, promoting accessibility for persons with disabilities ought to be a key aim in establishing a regulatory framework that shapes the development and use of new technology. While this issue does not surface in the Discussion Paper, it is important that considerations relating to disability access are reflected in the development of standards.

Recommendations

The University of Melbourne recommends:

- The development of the AI Standards Roadmap and further work on standards should incorporate a multi-disciplinary approach and draw on the expertise of researchers across a broad set of fields.
- Standards developed for AI should promote accessibility for persons with disabilities.

Summary of recommendations

International harmonisation

International harmonisation ought to be central to the development of standards for AI, with proper integration essential to Australia enjoying the full benefits of new technologies.

The University of Melbourne recommends:

- Standards Australia should continue to act as Australia's voice in international processes regarding standardisation of AI, helping to develop international standards that are appropriate for adoption in Australia.
- The AI Standards Roadmap should include a mechanism to monitor the work on AI being conducted globally and in other countries, ensuring Australia is well-placed to anticipate and respond to changes.

Innovation and standards development

Standards should aim to protect Australian consumers by establishing requirements or guidelines for functional, reliable and trustworthy AI systems, while still fostering an environment of innovation and a competitive marketplace for new technologies.

The University of Melbourne recommends:

- The AI Standards Roadmap should have regard to timing of standards development, with a view to ensuring any implementation enhances, rather than stifles, innovation.
- The AI Standards Roadmap should identify areas where more research is required prior to standards being developed and note opportunities where Australian researchers can assist in addressing these gaps.

Integration with related work

The AI Standards Roadmap should be integrated into other relevant work being undertaken on artificial intelligence. This includes aligning Australia's AI standards with those of other countries and, domestically, with the ACCC Digital Platforms Inquiry, the AI Technology Roadmap prepared by Data61 and the Australian Human Rights Commission's work on Human Rights and Technology.

The University of Melbourne recommends:

- Standards Australia have due regard to the related work being undertaken in Australia in relation to Artificial Intelligence in the development of its AI Standards Roadmap.
- The AI Standards Roadmap makes clear its relationship with the human rights framework being developed by the Australian Human Rights Commission and explicitly note that specifications prescribed by this framework are not voluntary.
- Standards Australia considers more broadly defining the scope of applicability of the roadmap analysis, adopting terminology such as 'automated-decision making' rather than 'AI'.

A multi-disciplinary approach

The Discussion Paper touches on issues that are legal, ethical and economic as well as technological. The development of the roadmap and subsequent work on standards should therefore draw on the knowledge of researchers from a broader set of fields.

The University of Melbourne recommends:

- The development of the AI Standards Roadmap and further work on standards should

incorporate a multi-disciplinary approach and draw on the expertise of researchers across a broad set of fields.

- Standards developed for AI should promote accessibility for persons with disabilities.

Contributors to this submission

Professor Mark Hargreaves, Pro Vice-Chancellor (Research Collaboration & Partnerships)

Associate Professor Tim Miller, Academic, School of Computing and Information Systems

Professor Jeannie Paterson, Academic, Melbourne Law School

Professor Liz Sonenberg, Pro Vice-Chancellor (Research Infrastructure and Systems) and Pro Vice-Chancellor (Digital and Data)

Dr Paul Barry, Adviser, Policy and Government Relations, Chancellery

Lauren Smith, Adviser, Policy and Government Relations, Chancellery

Attachments

The University of Melbourne provided responses to related consultations on AI and new technologies. The following responses are attached to this submission:

- Artificial Intelligence: Australia's Ethics Framework, University of Melbourne Response, June 2019 (Attachment A)
- Artificial Intelligence: Governance and Leadership White Paper 2019, Response from the University of Melbourne, March 2019 (Attachment B)
- Human Rights and Technology, Response to Australian Human Rights Commission Issues Paper, October 2018 (Attachment C)

Attachment A:

Artificial Intelligence: Australia's Ethics Framework

University of Melbourne Response

June 2019

Executive Summary

The University of Melbourne welcomes the opportunity to respond to the Department of Industry, Innovation and Science's Discussion Paper, *Artificial Intelligence: Australia's Ethics Framework*. The continuous development of new technologies has considerable ethical implications, entailing both risks and opportunities. We welcome the interest in this area shown by the Department and by the Data61 division within the CSIRO.

The Discussion Paper recognises that emerging technologies come with risks but also generate opportunities. We support the attempt to account for both. While there is often an understandable focus on the dangers that AI and related technologies represent, it is important that we not lose sight of the potential benefits that these technologies offer. In many cases, there is a clear ethical imperative to make use of automated systems e.g. where automated vehicles are likely to significantly reduce road fatalities. A key challenge in developing an AI ethics framework is responding to the risks posed by new technologies without undermining innovation in this area, thereby depriving Australia of the benefits. The Discussion Paper has attempted to achieve this balance.

We do, however, argue that there are significant issues with the framework described in the Discussion Paper, largely relating to the Paper's narrow focus in some areas. This submission addresses some of these issues. The following comments do not represent a comprehensive statement on what an ethics framework should look like. Instead, this submission outlines specific issues that the Discussion Paper has either overlooked or mis-characterised. These include:

- The discussion of 'privacy' captures only a subset of the ethical and legal issues associated with AI and related technologies and fails to acknowledge the other relevant public inquiries underway in this area.
- The discussion of individual consent for the use of data requires further development. There are a number of important considerations relating to consent that are not addressed in the Discussion Paper.
- The discussion of the significance of AI and related technologies to Indigenous Australians and to persons with a disability needs to be further developed.

As well as raising these and other issues with the content of the Discussion Paper, comment on each of the "Core principles for AI" identified in the Discussion Paper – and some suggested additional principles – is included in the Appendix of this submission.

Prior to coming to a response to the Discussion Paper, we offer a broad overview of the University of Melbourne's engagement with the legal and ethical challenges associated with AI and related technologies, outlining some of the key points made in our contribution to public consultations in this area.

We would welcome the chance to further discuss these issues in more detail with the authors of the report and with the Department of Industry, Innovation and Science.

For more information, please contact Professor Mark Hargreaves, Pro Vice-Chancellor (Research Collaboration & Partnerships) on 03 8344 4447 or m.hargreaves@unimelb.edu.au.

The University of Melbourne and Digital Ethics

The University of Melbourne has a deep engagement with the challenges associated with the evolution of new technologies. We refer to the University's response to the Australian Human Rights Commission's (AHRC) *Human Rights and Technology* Consultation Paper in 2018, and in response to the AHRC and World Economic Forum's *Artificial Intelligence: Governance and Leadership White Paper* earlier this year. In each case, the University's response drew from a community of researchers from across a range of fields. Key points made in those submissions include the following.

Human Rights by default and design

In our submission to the AHRC's 'Human Rights and Technology' Consultation Paper, the University promoted the general principle of 'Human Rights by default and by design' as a way of thinking about the human rights-related implications of new technologies. The motivating insight behind this principle is that human rights-related considerations ought to inform the development of new technology from the beginning. This contrasts with an approach that seeks to ensure that already mature technologies are brought into line with a human rights standard.

This basic principle – defined broadly in terms of 'ethics' rather than the narrower category of 'human rights' – is largely consistent with the approach suggested in the Discussion Paper. A clearly articulated principle such as this may nonetheless be useful in highlighting the integrated response that is needed in developing an ethical framework for the use of AI and related technologies. An ethics by default and design approach would help to ensure that the relevant considerations influence development in an ongoing way, instead of retro-fitting already developed technologies (while acknowledging that retrofitting will in some cases be necessary for established technologies). An ethics by design principle also underscores the importance of bringing a range of viewpoints to bear upon the design of new technologies, including users who are especially vulnerable to technology that is poorly designed.

A multi-disciplinary approach

As noted, the University has adopted a multi-disciplinary approach to investigating the legal and ethical challenges associated with AI and related technologies. This broad approach is appropriate for the work Data61 is undertaking to develop an ethics framework for AI. The challenges raised by emerging technology are not merely or primarily technological challenges. The response should therefore not be limited to those working in technology or related fields. In addition to technology researchers, an ethics framework should draw from a community that includes the humanities and social sciences, legal scholars and disability researchers.

A Technology Commissioner

The University of Melbourne suggests consideration of a new Chair – a Technology Commissioner – being established within the Human Rights Commission to have oversight of this area. The legal framework for protecting human rights in the context of new technology is dispersed across a range of Acts and instruments, including the Privacy Act, various anti-discrimination laws, and Australian Consumer Law. Attempting to re-build this legal framework from the ground up is neither desirable nor practically feasible. Moreover, in view of the rapidly changing context that laws and regulations need to grapple with, a "set and forget" approach to this area is inappropriate. A new Chair would enable a cohesive approach across Government to engaging with these challenges, working with industry and the research sector.

Comment on the Discussion Paper

As noted above, the University of Melbourne argues that there are considerable gaps in the ethics framework presented in the Discussion Paper, largely owing to a relatively narrow approach to some of the issues that it raises. The following comments address some of these gaps. The intention in these comments is not to identify everything that should be included in an ethics framework for AI, but to identify specific areas where the ideas raised in the Discussion Paper need further development.

A focus on 'AI'

The Discussion Paper makes 'AI' its specific area of focus, by implication placing beyond scope non-AI forms of technology. We have argued elsewhere that drawing a line between AI and non-AI technology for the purposes of a legal, regulatory or ethical framework is problematic.[1] Firstly, the distinction is on a basic level ethically relevant. AI is just one form of digital technology. There is no reason to single it out against other forms which raise the same ethical challenges relating to privacy, fairness, the avoidance of harm, etc. Secondly, creating a set of AI-specific guidelines may encourage organisations to seek to categorise a given application as something other than AI as a way of avoiding the relevant requirements.

Given this, an ethical framework should be developed independently to algorithmic implementation details, remaining agnostic with respect to the type of technology in question. Instead of a framework that is specific to AI, using a broader category such as 'automated decision-making' will help to avoid the problems identified above.

Privacy

Privacy-related issues are of central importance to the development of an ethical framework for AI and related technologies. While the Discussion Paper addresses privacy, there are significant flaws in its treatment of this issue. We refer to Salinger Privacy's submission in response to the Discussion Paper, which identifies a number of these flaws.[2]

For example, the Discussion Paper identifies 'privacy protection' as one of eight core principles for inclusion in the ethics framework. The broader articulation of that principle appears to assume that privacy law is only (or primarily) concerned with "private data". Since privacy law covers a much broader set of issues than this, couching the discussion in terms of private data risks sidelining important privacy-related considerations.

Anonymisation and re-identification

The possibility of individuals being re-identified through anonymised data is a crucial issue in the context of a discussion of privacy (raised in section 3.3 of the Discussion Paper). This possibility underscores the key point that de-identifying or anonymising data does not obviate the need to obtain the data owner's consent before sharing. Unfortunately, de-identification of detailed unit-record level data does not work without substantially reducing the information content of the data. Recent episodes in Australia have shown that sincere efforts at de-identification were insecure.

The Discussion Paper notes that the Government has sought to address this issue through the Privacy Amendment (Re-identification Offence) Bill 2016 which would prohibit the re-identification of data.[3] Unfortunately, the proposed amendments to *The Privacy Act* could deliver the worst of both worlds.

---

[1] See University of Melbourne, 2018, *Response to Australian Human Rights Commission Issues Paper*, p.7. https://about.unimelb.edu.au/__data/assets/pdf_file/0017/60146/UoM_submission_Human_Rights_and_Technology_Issues_Paper.pdf

[2] Johnston, Anna, 2019, "The ethics of artificial intelligence: start with the law" (Salinger Privacy). https://www.salingerprivacy.com.au/2019/04/27/ai-ethics/

[3] pp.30-31.

The amendments would prevent an open examination of problems, thus making them less likely to be discovered by Australian researchers. Bad actors could nonetheless continue to exploit opportunities to re-identify individuals. This represents a major risk. Even data that are not made open may be distributed to entities (such as an insurers or employers) with strong incentives to identify individuals. Hence, anonymisation should not allow the data holder to share other people's data without their consent.

Digital platforms inquiry

We also note that the Australian Competition & Consumer Commission's (ACCC) 'Digital platforms inquiry' is currently active and due to deliver its final report by June 2019. A number of the points raised in the Inquiry's Preliminary Report are directly relevant to privacy issues associated with the collection and storage of data. For example, in the context of the "information asymmetry between digital platforms and consumers", the ACCC offers the preliminary finding that:

> […] consumers are generally not aware of the extent of data that is collected nor how it is collected, used and shared by digital platforms. This is influenced by the length, complexity and ambiguity of online terms of service and privacy policies. Digital platforms also tend to understate to consumers the extent of their data collection practices while overstating the level of consumer control over their personal user data.[4]

While information asymmetry is clearly relevant to the issues of privacy and consent, this issue is not addressed explicitly in the Discussion Paper. More generally, the ethical framework under construction will be more robust the better integrated it is with other work being done in this area.

Consent and "reasonable expectation"

The issue of consent is crucial to the conversation on the ethical use of AI and related technology. While this issue surfaces in the context of the discussion on privacy and data breaches, it is noteworthy that consent does not feature explicitly in the principles proposed in the Discussion Paper. A principle of transparency insists that people should be informed when an algorithm that impacts them is being used, but that consenting to that use is a further issue.

The discussion would also benefit from engaging with the conceptual question of what it means for an individual to consent to their data being used. The idea of a 'reasonable expectation' may be useful in capturing what is morally significant in any agreement for the use of an individual's data, emphasising that it should only be used for purposes and under conditions that those affected have reason to expect and accept as appropriate. This approach would help avoid the problems with a narrow approach that focuses on a discrete act of consent performed at a specific time:

- **Power asymmetries:** There are cases where an individual has little choice but to consent to the suggested use of their data, e.g. where an application is required for work. A discrete act of consent is clearly insufficient in these circumstances. Couching consent in terms of 'reasonable expectations' makes it clear that *what* an individual is consenting to must itself be reasonable.
- **Information asymmetries:** As noted above, an individual may "consent" to their data being stored and used without properly understanding the nature of this use. Defining consent in terms of reasonable expectations provides some protection in the face of information asymmetries.
- **Secondary uses of data:** Data that are collected for a given purpose may be put to additional or new uses at a later stage. An ethical framework should make clear the conditions under which such uses are appropriate. The notion of 'reasonable expectations' is useful in helping to define these conditions. The Discussion Paper comes close to this point in noting the requirement from

[4] Australian Competition & Consumer Commission, 2018, *Digital Platforms Inquiry: Preliminary report*, p.8.

The Privacy Act that consent be "current and specific".[5] However, this issue is worth addressing directly in view of the scope for secondary use of data.

- **Data use affecting others:** The storage and use of data often impact not only the user themselves but also others, e.g. friends, family etc. The use of reasonable expectation helps to safeguard the rights and interests of other affected parties.

Indigenous Australians and AI

Section 6.5 of the Discussion Paper addresses Indigenous communities and their relationship to AI, identifying three interrelated issues for consideration: the need to comply with Indigenous cultural protocols; the importance of AI development and use being guided by cross-cultural collaborative approaches; and the need for transparency that ensures that "Indigenous people and organisations are clear about how AI learning is generated and why this information is used to inform decisions that affect Indigenous estates and lives."[6]

This discussion would benefit from a sharper focus on the active role that Indigenous communities should play in the development of new technology. The Discussion Paper has a significant focus on the ways in which Indigenous persons and communities may be impacted by emerging technology, with much less attention given to the role of Indigenous persons as users and potential innovators. (The reference to "cross-cultural collaborative approaches" points to the importance of Indigenous community involvement in the collection and analysis of data). An ethics framework should emphasise the need to support Indigenous *participation* and *agency* in shaping the AI agenda, recognising the importance of including Indigenous communities in tech development and in the design of policies, and the matter of Indigenous data sovereignty.[7] AI and related technologies have a key role to play in advancing solutions to complex issues affecting Indigenous Australians. The opportunities will go unrealised if Indigenous people are not included in the development of these technologies.

We should also note that Indigenous communities are especially vulnerable to privacy-related risks that come with (for example) the collection and storage of data on individual persons. The risk of individuals being re-identified (see below) through anonymised data is heightened when dealing with minority groupings and with sparsely distributed populations. The heightened risks for Indigenous communities and other marginalised groups is worth addressing directly in an ethics framework.

Accessibility

The impact of new technologies on persons with a disability is an issue that deserves more attention than it receives in the Discussion Paper. In discussing the issue of fairness, the Paper rightly identifies persons with a disability as one of the cohorts of Australians who are potentially vulnerable to discrimination where algorithms are biased or rely on unrepresentative input data.[8]

While important, issues relating to discrimination are only part of the picture when it comes to the ethical implications of new technologies on persons with a disability. An additional set of issues concern the need for 'accessible technology' i.e. ensuring that new technologies are accessible to persons with a disability. In our submission to the AHRC consultation on Human Rights and Technology, the University of Melbourne noted that digital technology is proving to be a powerful

[5] p.28.

[6] Discussion Paper, p.56.

[7] See, for example, Kukutai and Taylor (eds.) 2016, *Indigenous Data Sovereignty: Toward An Agenda*, Canberra, ANU Press; Indigenous Data Sovereignty Symposium 2017, Indigenous Studies Unit University of Melbourne https://mspgh.unimelb.edu.au/research-groups/centre-for-health-equity/indigenous-studies/indigenous-data-sovereignty-symposium ; Indigenous Data Sovereignty Communique 2018, Maiam nayri Wingara Indigenous Data Sovereignty Collective and the Australian Indigenous Governance Institute)

[8] See p.39 and p.41.

enabler for persons with a disability.[9] However, if poorly designed, new technology can further contribute to the marginalisation of those with a disability.[10] In addition to discrimination-related issues, the need for inclusive design warrants emphasis in an ethics framework for AI.

Accountability and collaborative research

Section 7.1.6 discusses the relationship between business and academia, and appropriately identifies the benefits promised by deeper ties between industry and research. Industry-research collaboration should have a central place in Australia's innovation agenda and is central to the ethical development of AI and related technology.

However, the Discussion Paper does not address the accountability-related issues that are associated with collaborative research. If we are to incorporate "ethics by design", starting when projects are initiated, then there needs to be clarity at the beginning of funding relationships about which party is accountable for the ethical use of the research product. The basic point that technology is often put to new uses – i.e. different from that initially intended – is relevant here, as researchers involved in development may be unaware of those uses. The lines of responsibility need to be clearly drawn in such cases.

Robustness

An important ethical dimension for AI and machine learning omitted in the Discussion Paper concerns the robustness of automated systems. While minute changes to a small number of an image's pixels may not alter a human's perception of the image, such changes regularly fool state-of-the-art machine learning systems.[11] A number of related attacks on AI systems are well known in the study of adversarial machine learning[12] and have been demonstrated in the real world.[13] In many such cases, human decision-makers would not make these errors, and would be able to continue to make good decisions in situations that are unlike those they have seen before. In machine learning, related issues of poor data hygiene lead to overfitting systems to the data they have been trained on, and broader problems of replication crises widely publicised in the sciences.[14] This is further compounded by the fact that AI systems struggle to identify that they are presented with a problem that they are not trained for. Comprehensive frameworks for ethics in AI must recognise the importance of robustness and the need to support it through practices that require human oversight and intervention.

---

[9] See *Response to Australian Human Rights Commission Issues Paper*, pp.11-12.

[10] See, for example, Haxton, Nancy, 2017, "Blind groups push for CBA to find solution to 'inaccessible' touchscreen EFTPOS terminals", ABC. https://www.abc.net.au/news/2017-07-28/blind-groups-push-for-solution-to-inaccesible-eftpos-terminals/8751366

[11] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. "Intriguing properties of neural networks." *arXiv preprint arXiv:1312.6199*(2013).

[12] Anthony D. Joseph, Blaine Nelson, Benjamin I. P. Rubinstein, and J. D. Tygar. *Adversarial Machine Learning*. Cambridge University Press, 2018.

[13] Alexey Kurakin, Ian J. Goodfellow, and Samy Bengio. "Adversarial Examples in the Physical World." In *Artificial Intelligence Safety and Security*, pp.99-112. Chapman and Hall/CRC, 2018.

[14] Andrew Gelman, and Eric Loken. "The statistical crisis in science: data-dependent analysis--a" garden of forking paths"--explains why many statistically significant comparisons don't hold up." *American Scientist* 102, no. 6 (2014): 460-466.

Contributors to this submission

Dr Chris Culnane, Lecturer, School of Computing and Information Systems

Mr Mark Fallu, Digital & IT Advisor, Chancellery Research

Mr Assyl Haidar, Director, Digital and Data, Chancellery Research

Professor Mark Hargraves, Pro Vice-Chancellor (Research Collaboration & Partnerships)

Professor John Howe, Director, Melbourne School of Government

Associate Professor Reeva Lederman, Academic, School of Computing and Information Systems

Associate Professor Tim Miller, Academic, School of Computing and Information Systems

Professor Scott McQuire, Academic, School of Culture and Communication

Professor Jeannie Paterson, Academic, Melbourne Law School

Professor Megan Richardson, Academic, Melbourne Law School

Associate Professor Ben Rubinstein, Academic, School of Computing and Information Systems

Professor Liz Sonenberg, Pro Vice-Chancellor (Digital & Data)

Associate Professor Mark Taylor, Deputy Director, Centre for Health, Law and Emerging Technologies (HeLex @Melbourne)

Associate Professor Vanessa Teague, Academic, School of Computing and Information Systems

Professor Monica Whitty, Professor of Human Factors in Cyber Security, School of Culture and Communication

Appendix

Comment on "Core principles for AI"

	Principle	Comment
Principles identified in Discussion Paper	Generates net-benefits	Given the uncertainty typically associated with new technology, there are likely to be cases where an innovation is reasonable but where an organisation cannot be sure that it will generate benefits greater than the costs. The definition of 'net-benefit can arguably be arbitrary and contested in any event – a 'net benefit' from whose perspective, for example. The key point may be best captured in terms of an intent to cause harm and of foreseeable consequences. The 'Do no harm' principles articulates this point.
	Do no harm	Support.
	Regulatory and legal compliance	Support. We note that this leaves open the extent to which the legal and regulatory framework itself is adequate.
	Privacy protection	While the principle itself is supported, the ethical issues related to privacy are broader than those addressed in the Discussion Paper.
	Fairness	Support. We note, however, that 'fairness' is discussed exclusively in terms of non-discrimination. While this is important, it is only one aspect of fairness. There are separate questions as to whether the use of automated systems is fair in the circumstances e.g. in a criminal justice setting. Also, there is a need to determine what counts as 'unfair discrimination' and who determines this.
	Transparency and explainability	Support.
	Contestability	Support.
	Accountability	Support. We note that there is a need to address the accountability issues arising out of collaborative research.
Suggested additional principles	Consent/Reasonable expectations	The storage and use of data should be dependent upon the consent of individuals *and* should reflect reasonable expectations of users and other affected parties.
	Accessibility	New technology should be accessible to persons with disability and should not further contribute to their marginalisation.
	Enforceability	Where the framework identifies legal and regulatory requirements, there should be mechanisms for enforcement and penalties for non-compliance.
	Remedy/recourse mechanism	Those unfairly affected by AI should have a cost-effective, timely and non-litigious remedy available to them.

Attachment B:
Artificial Intelligence: Governance and Leadership White Paper 2019

Australian Human Rights Commission and World Economic Forum

Response from The University of Melbourne

18 March 2019

Executive Summary

The University of Melbourne welcomes the opportunity to respond to the Australian Human Rights Commission and World Economic Forum's *Artificial intelligence: Governance and Leadership White Paper*. This submission builds on the University's response in 2018 to the Commission's *Human Rights and Technology Issues Paper* and associated policy roundtables hosted at the University.

The University commends the parallel focus on regulatory considerations, in light of the scope, complexity, pace of change and potential social influence of innovative technologies. As per our response to the Commission's human rights paper, the University recommends a multi-faceted and co-regulatory response to protecting human rights and preventing or minimising social harm, coordinated across relevant bodies and underpinned by the principal of 'human rights by design and default'.

An appropriate response to the flow of new technologies is the establishment of a framework that attends to the genuine risks without stifling innovation. Ideally, the framework should promote sustained comparative advantage in ethical design. Existing regulatory arrangements in the domains of privacy and consumer protection, as well as human rights, need to be elaborated and improved to properly accommodate the changing opportunities and risks of novel technologies. The University has therefore recommended mapping of existing regulators to establish the extent of change needed and strengthening of regulatory powers, functions and resources.

When societal damage, or risk of harm, is unambiguously a consequence of innovation, there is a case for regulatory intervention. The overarching objective for Australian law and regulation should be the provision of a regulatory space that promotes and fosters beneficial innovation while protecting against social harm. Enhanced functions of existing regulatory bodies, or in the alternative, a carefully calibrated new RIO with specific and non-duplicative functions, may be a useful supplement to Australia's existing frameworks. Additionally, a specialist advisory body should be established to support responsible regulation and provide high-level advice and engagement on innovative technologies to the public, governments and other regulatory entities.

For more information, please contact Professor Mark Hargreaves, Pro Vice-Chancellor (Research Collaboration & Partnerships) on 03 8344 4447 or m.hargreaves@unimelb.edu.au.

Recommendations

1. **The use of innovative technologies, including AI and ML, should be subject to greater oversight and regulatory responsibility to safeguard social expectations and protections including fairness, accountability, transparency, inclusiveness and non-discrimination in the use of those technologies.**

2. **Existing regulatory bodies, and relevant legal and regulatory frameworks, in Australia should be examined to determine if enhancing their powers and mandates would address regulatory requirements relating to innovative technologies such as AI. The mapping exercise should look at regulatory functions as well as diverse applications of automated decision-making technologies.**

3. **A specialist advisory organisation on AI and innovative technologies should be established to coordinate with other regulatory bodies and provide key supports such as;**

    a. **Providing community education and advice;**

b. **Reviewing and recommending on regulatory needs;**

c. **Providing technological and social analytical capability;**

d. **Conducting self-enabled inquiry into emerging issues;**

e. **Conducting community and cross-sector (including private sector) engagement on AI responsibilities;**

f. **Drawing on international experience/expertise and promoting relevant Australian examples and practices;**

g. **Providing expert guidance to Australian governments and public sector;**

h. **Developing a well-regarded and reliable 'trust' mark or equivalent identifier.**

4. **Inclusivity should be a core aim of a specialist advisory organisation or a RIO, which can be ensured by using a 'universal design' approach to the formation of any new regulatory organisation that treats impact groups as core stakeholders, rather than as exceptions.**

5. **A specialist advisory organisation or RIO should have access to internal and external expert guidance, such as a panel drawn from research and development sectors, to ensure currency of information and breadth of networked expertise.**

1: What should be the main goals of government regulation of AI?

Comments on definition

At the outset, the University raises a definitional concern about the use of 'artificial intelligence (AI) and machine learning (ML)' as the technological focus of this inquiry into regulation. While the White Paper rightly identifies AI as a key driver in current industrial transformation and investment, the fields that are relevant to the themes of this inquiry – and the proposal for a Responsible 'Innovation' Organisation (RIO) – expand beyond AI and ML. A RIO described as such would have relevance to a wide and fast-moving set of innovative technologies, digital or otherwise (e.g. cyber-physical, with combined physical and computational elements), including diverse forms of computing, automation, nanotech, Internet of Things (IoT), advanced materials, biotechnology, and others. Many data protection regimes around the world have preferred a broader or more technology neutral terminology that can keep pace with scientific advancements.[1]

In the case that narrower AI capabilities are identified for regulation to avoid social harm in that specific area, an applicable terminology could be 'automated decision-making'. However, too narrow a definition runs a practical risk of regulatory avoidance, in which researchers or companies choose to strategically rename new technologies to remain outside regulation. Additionally, narrowly aimed regulation – such as rules or standards triggered by the use of big data set-based AI – risk being outpaced when AI develops to the point it can be trained with much smaller data sets. As previously

[1] For a broader approach to the complex linkages between technologies and regulatory structures, see the analysis of 'Responsible Research and Innovation': a 'transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view to the (ethical) acceptability, sustainability and societal desirability of the innovation process and its marketable products (in order to allow a proper embedding of scientific and technological advances in our society)'. Von Schomberg, 2013, p. 63 referenced in:
https://www.frontiersin.org/articles/10.3389/fpls.2018.01884/full

raised in the University's response to the Commission's human rights paper, from a legal perspective the distinction between AI algorithms and other decision-making algorithms is irrelevant, and any regulatory response should be independent of algorithmic implementation details.

Australian needs in a global context

As noted in the White Paper, Australia already lags well behind in terms of Australian companies making sustained investments in AI. There is therefore a risk that heavy-handed regulation will make Australian-led technological advancements less relevant in a world context. At the same time there is growing interest in technologies that create public value.

The University therefore supports the development of a carefully calibrated response which would include an Australian specialist advisory organisation with global reach and connections, that can tap into the significant body of work on innovation, ethics and regulation being undertaken by industry, think tanks, universities, international organisations, and civil society groups globally, and deepen engagement with international partners without duplicating efforts. It is also important that it should reflect uniquely Australian needs and concerns, given the fundamental social protection aspect of its function. For instance, the impacts and interests of technological innovation on Australian indigenous people is not necessarily replicated by overseas algorithmic practices and models of regulation.

Fostering responsibility and public protections

The AI4People report launched in November 2018, *Ethical Framework for Good AI Society*, provides an excellent summary of the opportunities, risks and principles for ethical use of technology. Describing AI as enhanced, or even improved and multiplied, human agency, the report observes:

> The larger the number of people who will enjoy the opportunities and benefits of such a reservoir of smart agency "on tap", the better our societies will be. Responsibility is therefore essential, in view of what sort of AI we develop, how we use it, and whether we share with everyone its advantages and benefits. Obviously, the corresponding risk is the absence of such responsibility.[2]

In developing organisational and/or regulatory responsibility in the Australian context, the goals should accord with social values such as:

- Fairness;
- Accountability;
- Transparency;
- Inclusiveness;
- Non-discrimination.

Goals such as the above would usefully set a baseline for how AI should operate in Australian civil society. These include that AI does not discriminate against citizens, does not bully, coerce or exploit citizens, and does not cause harm to citizens. There is an important power dynamic that gives rise to the necessity of fostering technological responsibility in the public and private sectors; for the most part, citizens are most likely to be affected by automated decision-making but have little ability to influence the content and application of AI.

[2] http://www.eismd.eu/wp-content/uploads/2019/02/Ethical-Framework-for-a-Good-AI-Society.pdf page 8.

Promoting innovation and market protections

It is important that in regulating to achieve these goals, innovation is not stifled. Alongside protection of the public, regulation should facilitate and ensure the appropriate functioning of a competitive marketplace for new technologies. A well-designed regulatory structure could set up a lasting framework for ways in which innovative breakthroughs can be taken to market. Equally, it should be remembered that innovation can be fostered and encouraged within a regulatory regime that encourages best practice, rather than allowing a proverbial 'race to the bottom' in the AI context.

2: Considering how artificial intelligence is currently regulated and influenced in Australia, (a) What existing bodies play an important role in this area? (b) What are the gaps in the current regulatory system?

Domestically and internationally, institutions and companies are looking closely at the expansive social implications, including altered pressures on regulatory arrangements, brought on by the advancement of AI and innovative technologies. In recent years the OECD has been investigating the digital transformation of economies and societies.[3] UNESCO is leading a program on technological transformations with a focus on the Sustainable Development Goals and human impacts.[4] The Institute of Electrical and Electronics Engineers (IEEE), one of the world's largest industry standards bodies, has convened numerous global standards projects on AI and in 2018 launched an Ethics Certification Program for Autonomous and Intelligent Systems.

While these projects are looking at distinct aspects of AI, regulation (and its limitations) is emerging as a theme. Jurisdictions outside Australia are moving to regulate aspects of AI and serve as a useful comparison. For instance in mid-late 2018 the California legislature became the first US state to pass bills regulating rapidly-developing technologies including the IoT, AI, and chatbots. Europe already has in place a comprehensive General Data Protection Regulation (GDPR), which include provisions for data protection by design and the right of data subjects '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' (subject to certain limited exceptions).

In Australia, the existing regulatory bodies that play a role, to various degrees, in this area include:

- The Australian Competition and Consumer Commission (ACCC);
- The Office of the Australian Information Commissioner;
- The Australian Securities and Investments Commission;
- Office of the National Data Commissioner;
- Specific industries regulators, such as the Actuaries Institute and Australian Communications and Media Authority (ACMA);[5]
- The Australian Human Rights Commission.

The legal framework for protecting human rights in the context of new technology is dispersed across a range of Acts and Instruments, including the Privacy Act, anti-discrimination laws, and the Australian Consumer Law. Policy platforms such as the National Innovation and Science Agenda, supported by government offices like Data61 and the Office of the Chief Scientist, would also have relevance.

[3] http://www.oecd.org/going-digital/ai/

[4] https://en.unesco.org/artificial-intelligence

[5] ACMA have a significant role in relation to the NBN as well as technology companies operating in the media-communications space (e.g. live-streaming on Facebook).

Enhance regulatory powers in existing bodies

The University notes the comments in the White Paper excerpted from contributors to the Commission's earlier human rights paper, including the ACCC's recommendations, highlighting the need for greater regulatory oversight of digital platforms. However, the University does not, at this stage, recommend the establishment of an entirely new regulatory body to absorb all functions necessary to meet this need in the particular context of AI.

As a first step, the University recommends due consideration be given to whether enhancing and expanding the functions and powers of existing regulatory bodies could meet the new regulatory needs relating to innovative technologies. Many of the regulatory bodies listed above are already working collaboratively on challenges relevant to their sectoral remit or overlapping their remits. Establishing a new national RIO without considering workable alternatives in the present regulatory constellation could lead to duplication and impediments to innovation, which would be counterproductive to the intent of the reforms. With carefully calibrated enhancements (including in Australian law), and appropriate resourcing to cover the new areas of responsibility, existing regulators could efficiently and effectively step in to collaboratively cover regulatory gap.

Along these lines, the AI Now Institute's report in December 2018 recommended that governments should regulate AI by expanding the powers of sector-specific agencies to oversee, audit, and monitor these technologies by domain. Looking at the US, the report argued that 'a national AI safety body or general AI standards and certification model will struggle to meet the sectoral expertise requirements needed for nuanced regulation. We need a sector-specific approach that does not prioritize the technology but focuses on its application within a given domain'.[6] The University submits that this observation is relevant to present Australian circumstances.

The University's previous submission to the Commission's human rights inquiry outlined numerous ways in which Australia's existing legal framework could be updated to manage the challenges and maximise the opportunities created by advances in technology. It also flagged the useful comparative framework of the European Union's GDPR, which is significantly more advanced than the equivalent regulations in Australia and may be used as a guide to updating Australia's legal and regulatory arrangements.

Role for a specialist advisory organisation

While not a national RIO in the form sketched out by the White Paper, the University suggests there is scope and need for establishing a new specialist entity with deep technical expertise, so far as it does not replicate or overlap the remit of existing regulatory bodies. A specialist organisation in this vein could lead a coordinated model with other regulators, as well as provide the following supports:

- A platform for community education and advice;
- A mandate to review, monitor and recommend on responsible regulation;
    - This should include identifying what powers regulators need over time, supporting them to build specialist skills and looking ahead to help them anticipate emerging issues for their sectors;
- Technical/technological capability, to assist other regulators where the workings of innovation technologies are opaque or are not well understood;
- Social science and diverse disciplinary capability, to contextualise and humanise the social impact of innovative technologies;

[6] https://ainowinstitute.org/AI_Now_2018_Report.pdf page 4.

- Ability to conduct self-enabled inquiries into emerging issues;
- Mandate to facilitate community and cross-sector engagement;
- Provision of expert guidance to Australian, State and Territory governments, to assist with internal public service use and design of innovative technologies, and to assist with the standards required by governments in procurement and partnership;
- Strong links to draw on international experience/expertise and promote relevant Australian examples and practices;
- Potentially, ability to conduct *in camera* inquiries where a public issue has arisen in regard to commercially held algorithmic function; and
- Potentially, development of a well-regarded and reliable trust mark or verification to appropriately signal public confidence in the fairness and safety of a product or process impacted by innovative technology.

Map existing regulation to maintain coherence

Before introducing a new specialist entity or RIO in the terms of the White Paper, it will be important to carefully map the scope and function of existing regulatory regimes and bodies. This will be an opportunity to ensure a good fit between what is already in place and any new initiatives, to ensure coherence across the sector and avoid duplication.

As raised earlier, AI more narrowly defined is a decision-making capability in software that is application and scenario specific. The mapping exercise should therefore also look at application areas, e.g. advertising and direct marketing; job automation; robotics and others, to identify if there is an existing regulatory regime that provides appropriate coverage.

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

Improved regulation of innovative technologies would add social value for Australia by improving the general public's trust and familiarity with AI and related technologies. There is currently a lack of general education about technology and its regulation, which contributes to limited public support for its uses. There is also a perception in some parts of the population that the Australian Government is not well-equipped to work with cutting-edge digital technologies; a specialist advisory organisation could raise the public perception of the government's data handling capabilities.

There could be significant social and economic value in Australia taking a leading role in this area, rather than catching up with, or adapting, the approach taken elsewhere globally. It is worth observing that private and public institutions are moving swiftly in the area of AI skills and social impacts in other countries. For instance, in an effort to keep pace with China and the United States, the United Kingdom (UK) Government has recently announced plans to fund over 1,000 places for Masters and PhD students to train in AI technologies at the cost of approximately £115 million. Businesses, including Google DeepMind, BAE Systems and Cisco, have also pledged to help fund 200 new AI Masters courses at UK universities.[7] A carefully calibrated regulatory response in the immediate term, with appropriate review and revision, designed to enhance Australian capabilities could enable Australia to stay relevant in a global context and gain social and economic benefits of doing so.

[7] Sam Shead, 'U.K. Government To Fund AI University Courses With £115m', *Forbes*, 22 February 2019: https://www.forbes.com/sites/samshead/2019/02/20/uk-government-to-fund-ai-university-courses-with-115m/#5bfce7c4430d1/4

4: Under what circumstances would a Responsible Innovation Organisation add value to your organisation directly?

A specialist regulatory organisation could provide an additional and valuable channel for the University's research and educational expertise to flow out beyond the University to influence and inform public policy development. This is discussed in more detail below at (6d). A strong regulatory framework designed to support innovative breakthroughs being taken to market could also assist the University to manage risks that adversely impact our core businesses of teaching and research, while facilitating the productive expression of university-generated IP.

A high-functioning RIO or specialist advisory entity could, in turn, assist the university sector by providing informed forecasts on how teaching and public education services can be adapted or deployed to reflect and/or help people manage emerging technologies, trends and issues.

5: How should the business case for a Responsible Innovation Organisation be measured?

The University is silent on this point.

6: Features of a Responsible Innovation Organisation

This section of the White Paper sought feedback on the potential powers, functions, aims, connections, resourcing and evaluation of a RIO.

Inclusivity as a core aim (6a)

Inclusivity should be a core aim of any regulatory intervention. In a recent article published by Nature, an observation was made in relation to in the use of AI in healthcare in the UK:

> As with the advent of the car, many serious implications will be emergent, and the harshest effects borne by communities with the least powerful voices. We need to move our gaze from individuals to systems to communities, and back again. We must bring together diverse expertise, including workers and citizens, to develop a framework that health systems can use to anticipate and address issues. This framework needs an explicit mandate to consider and anticipate the social consequences of AI – and to keep watch over its effects.[8]

The White Paper noted the importance that the RIO's approach and governance be inclusive, with 'special attention given to those who are particularly affected by new technologies and most susceptible to the threats associated with them'. The University recommends a universal design approach could effectively inform the design of an inclusive RIO, by placing the perspectives of all people as central to the design process rather than as exceptional, and at the same time maintaining reasonable opportunities for human intervention and oversight of the design processes.

The universal design approach ensures that people with disability, children and young people, older people, and people from CALD backgrounds are contemplated as core stakeholders, rather than being viewed as special or vulnerable groups whose perspectives and needs are an add-on to the RIO's core model or business.[9] While some Federal, State/Territory, local governments are starting to identify

[8] Melanie Smallman, 'Policies designed for Drugs won't work for AI', *Nature*, Vol 567, 7 March 2019, p7.

[9] See https://www.wired.com/2017/03/voice-is-the-next-big-platform-unless-you-have-an-accent/

and encourage the application of universal design principles, particularly in relation to the built environment, the most prominent current example is the National Disability Strategy 2010-2020.[10]

As outlined in the University's response to the human rights Issues Paper, universal design principles are appropriate to the development of AI to ensure systematic bias or prejudice is not, even inadvertently, built into datasets or processes. Universal design should also be reflected in the design of regulatory frameworks, as stated by the UN Convention on the Rights of Persons with Disabilities (preamble para (o) and Article 33).

Powers and functions (6b)

Whether established in an existing regulator or in a new RIO, there must be enhanced supports created for people to seek redress from technology-driven harms by ensuring people's complaints are fairly handled and remedies enforced. As mentioned in other parts of this submission, an innovative technologies regulator could also create value for AI applications in Australia by developing a well-regarded and reliable trust mark or verification.

Structure (6c)

If a new RIO as sketched out by the White Paper is ultimately recommended by the Commission, a useful starting model is the ACCC. The ACCC is an effective regulator as it achieves a broad cross-sector reach, conducts stakeholder consultations, balances regulatory oversight with specialist expertise, and has a suite of powers allowing effective enforcement of the regulatory regime it administers. The consumer protection regime allows for both the ACCC and individual consumers to bring actions to protect their rights. Notably, the Australian Government fully funds the ACCC. A regulatory body that is co-funded by industry would need to be carefully structured to ensure its independence and avoid perception of regulatory capture.

Internal and external expertise (6d)

The University strongly recommends a specialist advisory organisation or a RIO should have inbuilt and external assets of computing technical expertise, including people from outside the Sydney/Canberra/Melbourne axis. As stated earlier in this response, it is also crucial to have social science skills that will help technologists and bureaucrats understand the impact of innovative technologies in the community.

The Australian university sector has a key role to play in the trajectory and impact of new technologies as a key nexus point of education, research and development. Publicly funded research is an enabler of innovation in the broader economy and is increasingly global and multi-disciplinary. Universities and partners in industry and elsewhere, being familiar with the rate of technological change and disciplinary complexity, offer a wealth of expertise to respond to these challenges as they arise.

The University recommends an External Advisory Group be established as part of the RIO to draw in expertise from relevant sectors, including universities. This would be consistent with the approach taken for other regulatory bodies. For instance, ASIC takes guidance from several advisory panels including an External Advisory Panel, which channels senior level advice from the financial services industry and other sectors.

[10] See https://www.humanrights.gov.au/sites/default/files/NDS%20PDF.pdf p 30.

References

https://www.frontiersin.org/articles/10.3389/fpls.2018.01884/full

http://www.eismd.eu/wp-content/uploads/2019/02/Ethical-Framework-for-a-Good-AI-Society.pdf

http://www.oecd.org/going-digital/ai/

https://en.unesco.org/artificial-intelligence

https://ainowinstitute.org/AI_Now_2018_Report.pdf

https://www.forbes.com/sites/samshead/2019/02/20/uk-government-to-fund-ai-university-courses-with-115m/#5bfce7c4430d1/4

https://www.wired.com/2017/03/voice-is-the-next-big-platform-unless-you-have-an-accent/

https://www.humanrights.gov.au/sites/default/files/NDS%20PDF.pdf

List of contributors from The University of Melbourne

Dr Greg Adamson, Associate Professor and Enterprise Fellow is Cyber Security, Melbourne School of Engineering

Mr Doron Ben-Meir, Vice-Principal (Enterprise)

Professor Bruce Bonyhady, Executive Chair and Director, Melbourne Disability Institute

Professor Mark Hargreaves, Pro Vice-Chancellor (Research Collaboration & Partnerships)

Dr Yvette Maker, Senior Research Associate, Melbourne Social Equity Institute

Professor Scott McQuire, School of Culture and Communication

Professor Bernadette McSherry, Foundation Director, Melbourne Social Equity Institute

Associate Professor Tim Miller, School of Computing and Information Systems

Associate Professor Jeannie Paterson, Melbourne Law School

Professor Megan Richardson, Melbourne Law School

Associate Professor Ben Rubinstein, Senior Lecturer, School of Computing and Information Systems

Dr Rajeev Samarage, Senior Data Scientist, Melbourne Institute of Applied Economics and Social Research

Professor Liz Sonenberg, Pro Vice-Chancellor (Digital & Data)

Associate Professor Mark Taylor, Deputy Director, Centre for Health, Law and Emerging Technologies (HeLEX @Melbourne)

Professor Tony Wirth, School of Computing and Information Systems

Ms Kelly Farrow, Policy and Government Relations Adviser, Chancellery

Attachment C:
Human Rights and Technology

Response to Australian Human Rights Commission Issues Paper

October 2018

Executive Summary

The University of Melbourne welcomes the Australian Human Rights Commission's 'Human Rights and Technology' Issues Paper. The Paper makes a timely contribution to the conversation around how human rights are to be respected and promoted in the context of responsible technological advancement.

The University held a series of three roundtable events in September 2018, drawing together leading researchers from various disciplines to discuss both the Issues Paper and the human rights-related challenges associated with emerging technologies. The comments provided below are the product of those roundtable discussions. A key aim is for the submission to reflect the multi-disciplinary approach we believe is essential to meeting the challenges generated by new technologies. A list of contributors has been included at the end of the submission.

The submission does not attempt to address all the topics that are discussed in the Issues Paper. Instead, we offer some comments that speak to the general approach that ought to be taken to human rights and technology – outlining the 'human rights by default design' principle and its implication – followed by responses to a selection of the consultation questions contained in the Issues Paper. These responses include some discussion of the threats and opportunities arising from new technology (in reply to Question 2), the challenges relating to AI-informed decision making (in reply to Questions 5, 6 and 7) and issues relating to accessible technology and persons with disability (in reply to Question 8).

We also note that Human Rights Commissioner Ed Santow was invited to one of the roundtable sessions. We greatly appreciate the Commissioner's attendance and his insights into the background of the AHRC's work in this area.

For further information, or to discuss the submission, Professor Liz Sonenberg, Pro Vice-Chancellor (Research Infrastructure and Systems), can be contacted at l.sonenberg@unimelb.edu.au or on (03) 9035 8619.

General comments

Human rights and responsible innovation

Much of what is challenging in this area is the rapid rate of technological change, and the problems that this poses to existing laws intended to safeguard human rights. The starting point for our thinking about these issues is that human rights and responsible digital innovation must *not* be seen as exclusive of each other. As the Issues Paper notes, new technologies are often accompanied by risks, but also promise significant benefits. If developed properly, new technology can help disrupt patterns of exclusion and contribute to Australia's economic and social wellbeing. The appropriate response to the flow of new technologies is to establish a framework that attends to the genuine risks that they pose without stifling their development.

Existing regulatory arrangements in the domains of privacy and consumer rights, as well as human rights, need to be elaborated and improved to properly accommodate the changing opportunities and risks associated with novel technologies. There will be new roles for certification, changed concepts of consent, and an ongoing need for education of consumers, policy makers and regulators. Given the scope and influence of new technology, we encourage a co-regulatory response that draws input from regulators across relevant sectors.

'Human rights by design and default'

The University of Melbourne suggests the adoption of a principle of 'human rights by design and default'. The motivating insight for this principle is that human rights ought to inform the design of new technology from the very beginning. This contrasts with an approach that seeks to correct already developed products and services that would otherwise fall short of a human rights standard. This means that human rights-related considerations should be present in the research that leads to new technologies, right through to product and service development and delivery. It also means that the governance and regulatory framework that provides oversight of new technologies develops in parallel to the technology itself, enabling responsible innovation.

A multi-disciplinary approach

The Issues Paper understandably has a focus on ensuring that the legal and regulatory framework intended to safeguard human rights in Australia remains adequate to the constantly changing digital environment. However, we cannot afford to become narrowly focussed on the legal implications of new technology. The questions that arise in this context are in many cases social, ethical and economic in nature. For example, issues relating to algorithmic bias and discrimination raise questions as to when discrimination between individual cases is 'unfair'; an ethical (as well as a legal) question. The regulation of new digital products ultimately needs to be sensitive to the relevant economic conditions in order to be effective, thus demanding the insights of economists, as well as legal scholars and technical experts. A siloed approach to tackling the implications of new technology will fail. What is called for is a broad, multi-disciplinary response to emerging technologies.

This response will address the way that products and services are designed and provided, and the various arms of the legal and regulatory framework that applies to new technology. It will draw from the research community that extends beyond Technology and Law, into the Humanities and Social Sciences. A broad approach of this kind reflects the scale of the issues raised in the Paper. The following comments explore selected issues in more detail, and flag some areas in which University of Melbourne researchers are already actively engaged.

A new Technology Commissioner

The University of Melbourne suggests consideration of a new Chair – a Technology Commissioner – being established within the Human Rights Commission to have oversight of this area. The legal

framework for protecting human rights in the context of new technology is dispersed across a range of Acts and instruments, including the Privacy Act, various anti-discrimination laws, and Australian Consumer Law. Attempting to re-build this legal framework from the ground up is neither desirable nor practically feasible. Moreover, in view of the rapidly changing context that laws and regulations need to grapple with, a "set and forget" approach to this area is inappropriate. A new Chair would enable a cohesive approach across Government to engaging with these challenges, working with industry and the research sector.

The University of Melbourne would welcome the opportunity to further discuss these topics with the Human Rights Commission.

Response to consultation questions

Threats and opportunities arising from new technology

Question 2: Noting that particular groups within the Australian community can experience new technology differently, what are the key issues regarding new technologies for these groups of people (such as children and young people; older people; women and girls; LGBTI people; people of culturally and linguistically diverse backgrounds; Aboriginal and Torres Strait Islander peoples)?

New technology typically impacts different cohorts in different ways. The Issues Paper rightly gives significant attention to this basic point. It is important that an understanding of the different ways in which particular groups interact with, and are affected by, technological transformation is integrated into the design process and into the policy response. In very broad terms, the two key issues concern (a) access (or lack of access) to new technologies for particular groups, and (b) the effects of those new technologies for particular groups.

In some cases, new technology promises to disrupt patterns of disadvantage, offering unique benefits to members of traditionally marginalised groups. Smart speaker technology represents a mere convenience for many. For people with vision impairment, this technology can provide a means of media access and a level of social connection that would otherwise be unavailable.

Of course, it is also true that technology can adversely impact demographic cohorts that are already disadvantaged, thereby further entrenching inequality. Women, LGBTI people and members of ethnic minorities contend with prejudice that is both harmful in itself and that drives unequal outcomes across a range of areas. There is a danger that prejudice gets 'built in' to new technology, for example where AI-applications unfairly discriminate against individuals based on their membership of a particular group. This brings the risk of compounding unequal outcomes relating, for example, to employment.

Similarly, low socio-economic and regional communities suffer from lower levels of access to education, employment and social amenity. Since lower levels of material wealth mean that members of these communities find themselves on the wrong side of the "digital divide", there is a danger that digital advances exacerbate unequal access. Those at the lower end of the socio-economic scale could find themselves doubly disadvantaged by digital disruption, to the extent that they occupy jobs that are more susceptible to being displaced than other parts of the labour market.

Grappling with the digital divide is not just a case of defining those who have access against those who do not. A range of other factors should be taken into account, including: level of comprehension and self-efficacy of use of digital technologies; general level of digital literacy and; groups that have been either victims of online crimes or who are more susceptible to online harms, or who have a fear of crimes or other harms caused by interacting with digital technologies.

We should also note that groups not yet identified may become disadvantaged by the transition to new technologies. This highlights the importance of publicly funded research to monitor the 'emerging disadvantaged', and to enact interventions that address this.

AI-informed decision making

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?

AI-informed decision making is becoming increasingly prevalent across a range of areas. It is now used for the purposes of credit scoring, recruiting, predictive policing and in the criminal justice

system. The following comments focus on the criminal justice system, but there are significant human rights implications in the use of AI applications across each of these domains.

The application of AI-informed decision making is becoming common in the criminal justice system. There is a risk that the right to liberty and the right to non-discrimination may be compromised if the relevant applications are not developed in a carefully considered manner. In Australia, AI-informed risk assessment tools are widely used for the purposes of risk management in correctional settings and for predictive purposes in post-sentence detention. Post-sentence preventive detention and supervision schemes exist in multiple jurisdictions[1]: these schemes are regarded as problematic, independent of the use of AI-informed risk assessment. While the High Court of Australia has held that post-sentence detention is constitutional,[2] eleven of the thirteen members of the United Nations Human Rights Committee have agreed that such schemes breach Article 9(1) of the International Covenant on Civil and Political Rights.[3]

Such schemes often call upon forensic psychologists and psychiatrists to provide a predictive risk assessment of individual cases. Usually, the assessment involves the use of an algorithm that uses empirically identified risk factors to generate a risk score. Much of the research in this area is focused on the risks concerning future violence: there are more than 200 violence risk assessment tools available.[4]

The use of these tools to inform decisions based on a prediction of the likelihood of future offences is problematic:

- **Transparency**: there are serious concerns relating to the transparency of the algorithms used. Where the particular method by which a decision was arrived at is unclear, there is a danger that the defendant is prevented from challenging the grounds for that decision. This was the central issue in *Loomis v Wisconsin 2017*, discussed in the Issues Paper (p.29).
- **Reliability and accuracy:** there are concerns relating to the reliability and accuracy of AI-informed risk assessment in a courtroom. Assessment outcomes may deliver a "false positive" – a finding that the individual concerned is at risk of harming others when this is not the case – resulting in unnecessary detention.
- **Application of risk assessment across differing populations:** The application of assessment tools to members of particular demographic cohorts can be problematic to the extent that the tool was predominantly developed and tested on members who do not belong to that cohort. In one case, the Supreme Court of Canada found that Correctional Services Canada breached its obligation to take all reasonable steps to ensure that information about an offender that it uses is as accurate as possible.[5] This was on the basis that in assessing the risk of Indigenous offenders, it used actuarial risk assessment tools that were developed and tested on predominantly non-Indigenous populations. The use of such tools in relation to Indigenous offenders in Australia has also been questioned.[6]

---

[1] See in general Bernadette McSherry, *Managing Fear: The Law and Ethics of Preventive Detention and Risk Assessment* (New York: Routledge, 2014).
[2] *Fardon v Attorney-General (Qld)* [2004] 223 CLR 575; Kirby J dissenting.
[3] *Re Fardon v Australia i[2010] Human Rights Committee, Communication No. 1629/2007 UN Doc CCPR/C/98/D/1629/2007* (12 April 2010); *Re Tillman v Australia* [2010] Human Rights Committee, *Communication No. 1635/2007, UN Doc CCPR/C/98/D/1635/2007* (12 April 2010).
[4] Seena Fazel and Achim Wolf, "Selecting a risk assessment tool to use in practice: A 10-point guide" (2018) 21(2) *Evidence Based Mental Health,* 41-43.
[5] See *Ewart v Canada* [2018] CCC 30.
[6] *Attorney-General (Qld) v McLean* [2006] QSC 137, para [26]; *Attorney-General (Qld) v George* [2009] QSC 2, para [33].

These problems point to the need for clear guidelines for – and constraints upon – the use of AI-informed risk assessment in a criminal justice setting. Justice Glazebrook of the New Zealand Court of Appeal has attempted to formulate such guidelines; these could serve as a regulatory model for this area of AI-informed decision-making.[7]

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

a) What should be the overarching objectives of regulation in this area?
b) What principles should be applied to achieve these objectives?

The overarching objective for Australian law should be the provision of a regulatory space that promotes and fosters beneficial innovation consistent with respect for human rights. The approach should be oriented around a basic principle of human rights by design and default, drawing on and expanding beyond the now established idea of 'privacy by design', or 'data protection by design and default' embodied in the EU General Data Protection Regulation (GDPR).

Importantly, the University of Melbourne urges a multi-faceted response to respecting human rights in the context of digital automation. The legal framework is an important part of this response, but it is only part of it. The reply to Question 7 below notes non-legal measures that should be considered to help ensure that human rights are protected and advanced in the face of new technologies. Before coming to these, we offer some general points that ought to guide the approach to developing laws and regulation that deal with new technology.

AI and non-AI technology

The Issues Paper has a strong focus on decision-making using AI, with section 6 of the Paper dedicated to the human rights implications of AI technology. While the interest in AI is understandable given its dramatic increase in recent years, the University of Melbourne cautions against attempts to develop laws that are specific to AI, to the exclusion of other technology. There are two reasons for this:

i. From a legal perspective, the distinction between AI and non-AI technology is irrelevant. AI is a branch of computer science; AI algorithms are just algorithms. There is no reason why a law or a regulation should apply to an AI algorithm but not to other algorithms that are tasked with the same decisions. While there may be additional challenges of opacity applying to modern machine learning techniques, the regulatory response should be independent of algorithmic implementation details.

ii. Attempting to introduce AI-specific regulations is problematic, given the lack of a clear or widely accepted definition of AI. Such an attempt would likely result in the wrong kind of behaviour on organisations who are targeted by those legal rules. Rather than aiming to comply, organisations would invest time and effort arguing that the regulations do not apply to them because what they are doing does not count as 'AI'.

Accuracy

Despite the interest in this area, there is some confusion surrounding what 'accuracy' in AI actually means. In many cases, AI informed decision-making should not be viewed in binary terms, i.e. as issuing either a correct or an incorrect answer to a given question. Such an approach is often neither realistic nor desirable:

- This may be because the AI application is predictive in nature, providing an assessment that is probabilistic and that reflects the *likelihood* of an event taking place, rather than a statement as to whether the event will (or will not) take place.

[7] See *R v Peta* [2007] 2 NZLR 627, CA., paras [51]-[53].

- In other cases, incomplete input data – arising from imprecise measurements, missing observations, human error or linguistic ambiguities – means that a degree of uncertainty is inevitable. This is an issue across many problem areas, including health and finance.

The key point is that such uncertainty can have positive or negative effects. Negative effects should be addressed and/or rectified. On the other hand, sometimes an uncertain measurement can be beneficial as it captures the essence of the problem or 'gold standard', e.g. a range of expert opinions.

From a research standpoint, what is needed are methods to quantify and model uncertain data to optimise the datamining and hence decision support capabilities of the AI system when faced with some inherent uncertainty. There is a further question of whether it is reasonable or appropriate to place expectations on AI-enabled decision making that go beyond that which is expected of human decision making.

Explainable AI

Given both the significance of decisions in which AI technology is now involved and the complicated nature of the applications in question, it is important for individuals affected to understand why AI applications arrive at the decisions they do. The currently active research topic of 'Explainable AI' aims to address the reasons for AI decision making, such that the relevant reasoning can be understood by humans. Explainability is essential to the legitimacy of AI decision making.[8]

Differential privacy

Differential privacy is a relatively new framework for guaranteeing the privacy of individuals in a sensitive dataset, when releasing aggregate statistics or machine-learned models on such data. This should be part of the approach to privacy-by-design in data sharing systems. Differential privacy is yet to be addressed to any significant extent in law or regulation.

Companies such as Google[9] and Apple[10] have deployed products and services using differential privacy, while the U.S. Census Bureau has announced that all releases of the 2020 Census must preserve differential privacy.[11] Ongoing research at Melbourne is seeking to develop new tools for making the framework usable by developers.[12] The framework complements cryptographic protocols for guaranteeing privacy when temporarily storing, transmitting, or processing data in untrusted environments such as over the internet or in a cloud service.

Penalties for failure to comply

To be effective, laws and regulations that constrain the design of digital products and services need to be backed by penalties for organisations that fail to comply. The University of Melbourne supports consumers being given individual rights of redress in circumstances where their data has been misused. An analogy here lies in consumers' right of redress for unsafe or defective products

[8] See Vered, M. and T. Miller (2018) 'What were you thinking?', *Pursuit*. https://pursuit.unimelb.edu.au/articles/what-were-you-thinking

[9] Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. "Rappor: Randomized aggregatable privacy-preserving ordinal response." *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2014.

[10] Apple Computer Inc. "Differential Privacy Overview." Report https://www.apple.com/privacy/docs/Differential_Privacy_Overview.pdf accessed Oct 5, 2018.

[11] John M. Abowd. 2018. The U.S. Census Bureau Adopts Differential Privacy. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '18). ACM, New York, NY, USA, 2867-2867.

[12] Benjamin Rubinstein and Francesco Aldà. "diffpriv: Easy Differential Privacy." open-source R package, https://cran.r-project.org/package=diffpriv

under the Australian Consumer Law, and for various contraventions in the Consumer Data Right, applicable in the first instance to banking.[13]

Trade-offs between conflicting aims

Ultimately, decisions around the use of AI-informed decision making involve trade-offs between desired but conflicting aims. The use of algorithms promises considerable benefits relating to efficiency in decision making and to the mitigation of some of the risks of human error and bias. On the other hand, the same algorithms may raise legitimate concerns regarding accuracy and fairness, particularly when their operations are opaque. In some cases, key objectives may be mutually incompatible, for example where two plausible sets of fairness criteria cannot both be satisfied.[14]

This issue highlights the need for judgement, exercised on a case by case basis, in consideration of the specific circumstances surrounding the application in question, and of its associated risks and benefits. Establishing a specific Chair within the Australian Human Rights Commission, with broad oversight of the relevant issues, to provide judgement on individual cases seems necessary at this stage of AI development. As noted, the Chair could also monitor trends and advise on issues as they arise, and guide policy development.

c) Are there any gaps in how Australian law deals with this area? If so, what are they?

d) What can we learn from how other countries are seeking to protect human rights in this area?

The legal framework in Australia is falling behind international standards and has not kept pace with data use and storage practices that form the basis of AI decision making. Current laws are patchy across a number of areas, in particular concerning privacy and consumer protection.

The Privacy Act 1988 (Cth) (Australia's version of a data protection regime) could usefully be updated in several respects in response to the challenges posed by advances in digital technology. Consideration should be given to:

- expanding the Act's definition of 'personal information' to take explicit account of digital technologies,
- extending responsibilities of data custodians, including in relation to a data subject's rights of access to data, and allowing for data portability (as provided for in the new Consumer Data Right being rolled out in the first instance for the banking, energy and telecommunications sectors[15]),
- providing more elaborated provisions about consent and what that means in this context,
- establishing a legal basis for individuals to object to automated processing, where reasonable,
- elaborating requirements for transparent and explainable automated decision making, and
- providing expressly for 'privacy by design and default' following the model of 'data protection by design and default' in the European Union's General Data Protection Regulation (GDPR).

The GDPR is significantly more advanced than the equivalent regulations in Australia, and may be used as a guide to updating our legal framework around data protection.

Recent cases before the European Court of Justice, the European Court of Human Rights and the United States Supreme Court are drawing from human rights frameworks to address what it means to be a member of a digital society in the context of powerful technology. We anticipate that the process will only continue as individuals and groups find that automated decision making

[13] https://treasury.gov.au/consultation/c2018-t316972/

[14] See Corbett-Davies, Sam, Emma Pierson, Avi Feller and Shared Goel (2016) "A computer program used for bail and sentencing decisions was labeled bias against blacks. It's actually not that clear", *Washington Post*. https://www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas/?utm_term=.a13b72c87b12

[15] See eg https://treasury.gov.au/consultation/c2018-t316972/

increasingly impacts the conditions in which they live. These approaches can offer useful insights for Australia.

Question 7. In addition to legislation, how should Australia protect human rights in AI-informed decision making? What role, if any, is there for:

a) An organisation that takes a central role in promoting responsible innovation in AI-informed decision making?
b) Self-regulatory or co-regulatory approaches?
c) A 'regulation by design' approach?

As noted above, the University of Melbourne encourages a multi-faceted and co-regulatory response to protecting human rights in automated decision making. We envisage a cooperative regulatory model, coordinated by a proposed Technology Commissioner and including agencies such as the Office of the Australian Information Commissioner (OAIC), the Australian Competition and Consumer Commission (ACCC) and the Australian Securities and Investments Commission (ASIC). This will enable the leadership necessary to establish a constructive regulatory approach consistent with the principle of 'human rights by design and default'.

Beyond the legal and regulatory changes suggested above, we recommend considering ways in which businesses may be incentivised to drive continuous improvement in performance. The following interventions merit consideration.

Certification (The Turing Stamp)

The Issues Paper discussed the so-called 'Turing Stamp' advocated by Chief Scientist Alan Finkel, which would signal to consumers that a given product or service has been developed and designed in an ethical manner, with due consideration to human rights. This would be a voluntary measure; in theory, businesses that have secured the stamp for a given product would have an advantage over competitors that have not, entailing an incentive for businesses to better integrate human rights thinking into the design process.

While there is significant merit to this proposal, there are a number of issues that would need to be addressed. In one sense, a voluntary certification system for the compliance with human rights standards is conceptually problematic: companies are stringently obligated to comply with human rights standards, marking a difference with the 'Fair Trade' or 'Australian Made' certifications. Presumably, a trust mark would be available to businesses that go *beyond* their legal obligations, indicating a need to determine how the voluntary framework would interact with the obligations that all businesses are subject to.

Another issue with a voluntary framework such as this is the possibility that, for some products and services, none of the businesses operating in this area are concerned with meeting the standard that would secure the trust mark. This possibility is particularly prevalent where this is limited competition due to the market being dominated by a small number of large players.

Leveraging government procurement

The Federal and State Governments are major purchasers of digital goods and services, giving them the power to influence business behaviour. Government procurement may be leveraged to improve digital ethics standards, for example, by including ethics performance in the assessment of tenders for major Government contracts. If established, a trust mark may aid Government procurement decisions. In any event, Government procurement will require the skills to assess the use of AI-decision making and its human rights implications, and to assess the suitability of particular applications for government purposes.

A high-profile prize or award for innovation

A prize or award for innovation could help raise awareness of the human rights implications of new technology. The prize would be awarded to technology developed in Australia that innovatively addresses challenges linked to, for example, automated decision making. The prize could be administered by the Human Rights Commission, co-ordinating with other Government agencies e.g. the Chief Scientist, the Chair of Innovation Australia and representatives from industry and research.

The role of the university sector

Australia's university sector has a key role to play in the responsible development of new technology. Our universities will be central to ensuring that Australia labour force enjoys the high-level skills that are needed to drive ongoing innovation.

The importance of university research should also be emphasised. Publicly funded research is a key enabler of innovation in the broader economy and is essential to addressing the range of challenges that are associated with the emergence of new technologies. The rate of technological change heightens these challenges, making the target of public policy a moving one. The wealth of expertise that Australia's universities offer is needed to respond to these challenges as they arise. The research sector also offers rigour and integrity, supporting investigation of these matters in an independent and evidence-based manner. Universities bring a multi-disciplinary capability to education and research, which is essential to adequately dealing with the implications of new technologies.

Accessible technology

Question 8. What opportunities and challenges currently exist for people with disability accessing technology?

The University of Melbourne commends the attention given to the range of issues associated with new technology and persons with a disability. While it is appropriate to be especially concerned with potential transgression against the rights of persons with a disability, new technology entails benefits as well as challenges. We should hold both in view.

The principle of universal design demands that the perspectives of people with disabilities are made central to the design process, rather than treated as exceptions. People with disability are among the most digitally excluded groups in Australia, indicating that many experience serious barriers to access.[16] This has possible flow-on effects, for example where the voices and needs of persons with a disability are excluded from data-sets and design processes. AI systems built on narrow data sets intended to represent a hypothetical or paradigm citizen risk perpetuating bias and discrimination. Research is urgently required to address the gaps in our knowledge of these issues. Inclusive design is likely to improve the functionality of new technologies for all users and will inform more adaptive and innovative AI systems.

Similarly, the development and review of the regulatory framework should draw from the perspectives of persons with a disability, as required by the UN Convention on the Rights of Persons with Disabilities (Preamble para (o) and Article 33).

The benefits of new technology

Notwithstanding the ongoing challenges, new technology is proving to be a powerful enabler for persons with a disability, delivering significant benefits. It is expected that, out of roughly $22b in

[16] Thomas, J, Barraket, J, Wilson, CK, Cook, K, Louie, YM & Holcombe-James, I, Ewing, S, MacDonald, T, (2018) "Measuring Australia's Digital Divide: The Australian Digital Inclusion Index 2018", RMIT University, Melbourne (see p.6).

NDIS expenditure, $1b will be spent on assistive technology.[17] Since this spending is driven by participant choice (rather than being controlled by Government), this serves as an indicator that participants themselves discern the value of new technology in helping overcome the barriers they face.

It is also worth noting that the tech industry is already embracing universal design. This is largely due to the market advantages of doing so; people with disability demand accessible design, and people without disability benefit from accessible features, e.g. being able to use smart devices when they cannot see them, touch them or hear them. In the age of personalisation, a device which cannot be used in every environment is not competitive. As a result, universal design is becoming core to technology companies, not a "nice to have", as evidenced by the accessibility features which are being added to Apple and Microsoft devices and software continuously.

This is promoting digital access for people with a disability, because these smart devices are getting cheaper (due to Moore's Law) and are being improved faster than specialist technology designed just for people with disabilities. There are further potential benefits relating to digitally enabled improvements in the accessibility of communication materials for those with cognitive impairments.

As well as the 'stick' of regulation, we should consider 'carrots'. A key question is how technology companies can be further incentivised to embrace universal design. Our view is that there is a need for much more co-design, where persons with a disability are included in the design process. By putting people with disability at the centre of design, we design for all and therefore embrace 'the edges' described by Manisha Amin (Centre for Inclusive Design).[18]

Designing for Disability = Designing for All

University of Melbourne researchers are in the earlier stages of developing a research program to address some of the major gaps in global research on disability support.[19] The 'Designing for Disability = Designing for All' research program will target two key areas: the benefits and challenges concerning consumer-directed human services; and the establishment of a longitudinal database that captures information on persons with a disability, e.g. details of diagnosis and functioning, the use of services and supports, and so on. The data platform will allow for the identification of new innovations as they arise, and will allow for the outcomes of technological and service innovations to be properly evaluated.

[17] Centre for Digital Business Pty Ltd, "Submission to Joint Standing Committee Inquiry on the NDIS ICT Systems", p.13.
https://www.aph.gov.au/Parliamentary_Business/Committees/Joint/National_Disability_Insurance_Scheme/NDISICTSystems/Submissions

[18] See "Technology – friend or foe of people with a disability?" (2018), *Big Ideas*, ABC Radio National.
http://www.abc.net.au/radionational/programs/bigideas/technology-%25E2%2580%2593-friend-or-foe-for-people-with-a-disability/10056698

[19] The 'Designing for Disability = Designing for All' program is being developed within the Melbourne Academic Centre for Health (MACH), of which the University of Melbourne is a member.

Contributors to this submission

Dr Greg Adamson, Enterprise Fellow, Melbourne School of Engineering

Professor Uwe Aickelin, Head, School of Computing and Information Systems

Dr Paul Barry, Adviser, Policy and Government Relations

Professor Bruce Bonyhady, Executive Chair and Director, Melbourne Disability Institute

Dr Suelette Dreyfus, Lecturer, Department of Computing and Information Systems

Mr Assyl Haidar, Director, Digital and Data

Dr Yvette Maker, Senior Research Associate, Melbourne Social Equity Institute

Professor Bernadette McSherry, Foundation Director, Melbourne Social Equity Institute

Associate Professor Tim Miller, Academic, School of Computing and Information Systems

Associate Professor Carsten Murawski, Academic, Department of Finance

Professor Ampalavanapillai (Thas) Nirmalathas, Director, Melbourne Networked Society Institute

Associate Professor Jeannie Paterson, Academic, Melbourne Law School

Professor Megan Richardson, Academic, Melbourne Law School

Associate Professor Ben Rubinstein, Senior Lecturer, School of Computing and Information Systems

Professor Liz Sonenberg, Pro Vice-Chancellor (Digital & Data)

Associate Professor Mark Taylor, Deputy Director, Centre for Health, Law and Emerging Technologies (HeLEX @Melbourne)

Professor Frank Vetere, Director Microsoft Research Centre for Social Natural User Interfaces

Professor Monica Whitty, Professor of Human Factors in Cyber Security, School of Culture and Communication

Professor Tony Wirth, Professor, School of Computing and Information Systems