



Microsoft Submission to the Australian Human Rights Commission Discussion Paper  
March 2020

Microsoft welcomes the opportunity to contribute to the consultation on the Australian Human Rights Commission (AHRC) *Human Rights and Technology Discussion Paper 2019* and commend the AHRC's work in advancing the discussion on human rights and technology. We believe approaches to the responsible development and use of Artificial Intelligence (AI) should be grounded on international human rights laws and principles and agree that AI needs to be human-centric. Our general thoughts and comments are structured under a few key themes, namely:

- [1. Role of legal and regulatory frameworks](#)
- [2. Transparency and explainability](#)
- [3. Pilots/testing/sandbox for human rights in practice](#)
- [4. Education and awareness raising](#)
- [5. AI Safety Commissioner](#)
- [6. Accessibility](#)
- [7. AI and human rights – beyond the technology](#)

We look forward to engaging in further discussions with the AHRC.

1. Role of legal and regulatory frameworks (Proposal 1)

We support a consistent strategy and approach to the regulation of new and emerging technologies such as AI, making sure to include the protection and advancement of human rights of all people. Microsoft has publicly stated its commitment to human rights in a [Global Human Rights Statement](#).

We agree with the AHRC that a multi-faceted regulatory framework is appropriate, involving the application of existing laws as well as 'soft law' co-regulation and self-regulation mechanisms such as international standards, industry best practices and guidelines, and internal governance procedures – to ensure the development and use of AI is responsible and aligned with human-centred values.

As Microsoft has called for in the past, including in its submission to the 2018 Australian Human Rights Commission process on AI, and the 2019 Data 61 discussion paper on AI Ethics, a more intensive and open dialogue between government, industry and other stakeholders is needed in Australia to assess where legislation already regulates the use of AI and addresses human rights and other concerns that may exist – as well as those areas where new regulation may be required.

We agree with the AHRC’s overall approach to legal reform – that is, applying existing laws and regulations as effectively as possible whilst looking out for gaps in the regulatory framework that warrant potential legal reform. In our view, the most effective approach to reform should start by identifying sensitive or consequential use cases of AI – applications of AI that have a significant impact on individuals or society (see below comment under ‘Risk-based approach to reform and definition of AI-informed decision making’). Given the broad range of applications of AI technologies, and the fact that the existing legal framework is likely to address many concerns with potential harms that involve the use of AI, this targeted approach will be essential. At the same time, given the time required to achieve legislative change, it will be important that other efforts proceed to prevent potential harms caused in specific use cases for AI – for example, developing standards and best practices, and building understanding of the features and limitations of specific technologies.

In this context, we would like to provide input on the following specific issues:

- [Ethical Frameworks \(Proposal 2\)](#)

We recognise the importance of ethical AI frameworks in guiding responsible AI use and development, while agreeing that these are not a substitute for regulation, standards, and other approaches that together are required to ensure the responsible use of AI. We have observed growing consensus in the region over such frameworks with the OECD and G20 having created amalgamated principles which have been agreed upon by members and some non-member nations. The Berkman Klein Center have also created a useful [visualisation](#) of various entities’ approaches to AI principles, which underlines the growing common ground between them. This is supported by an [exercise](#) undertaken by the Singapore Personal Data Protection Commission. In light of this, we question whether a new inquiry into ethical frameworks, as suggested in Proposal 2, is necessary.

As noted in the discussion paper, the Department of Industry, Science, Energy and Resources recently released 8 core AI Ethics Principles – one being focused on upholding human-centred values including respecting human rights. Microsoft, along with a number of other institutions, are piloting such principles and associated guidance to assess their application in practice. We suggest this pilot presents an opportunity to test the application of the principles, including as they relate to human rights, rather than the separate inquiry envisaged in Proposal 2.

- Risk-based approach to reform and definition of ‘AI-informed decision making’ (Question A)

We agree in substance with the AHRC’s approach to focus efforts on improving accountability measures on the most sensitive and consequential uses of AI. This is aligned with [guidance](#) on the application of Australia’s AI ethics principles provided by the Department to consider whether an AI-informed decision will have a “significant impact (positive or negative) on people (including marginalised groups), the environment or society”.

Microsoft’s internal guidance for teams working on AI sets out considerations (along with illustrative examples) to assist in identifying whether a particular use of AI may be sensitive. This involves considering the potential for AI to be used in ways that infringe on human rights, whether it generates a risk of physical or emotional harm, or if it may result in the denial of consequential services. This guidance helps teams focus attention on evaluating potential impacts and designing mitigations for any negative impacts.

We wished to flag some possible confusion in the proposed use of the term ‘AI-informed decision making’ to refer to decisions that have a legal or similarly significant effect on individuals which are materially assisted by the use of AI.

We note the AHRC intends to use quite a broad term to describe AI-assisted decisions that have a particular level of significance of impact – noting that some decisions informed by AI may have an insignificant impact. We suggest the AHRC consider using a phrase like ‘AI-informed consequential decision making’ or ‘uses of AI with consequential or significant impacts’, which may provide greater clarity in terminology. This may be assisted by including illustrative examples such as the application of AI in credit risk scoring or in making hiring decisions.

- Legislating a rebuttable presumption of responsibility (Proposal 10)

We note accountability is a core principle reflected in the law where strong legal traditions, principles, and precedent already exist and apply regardless of whether software (AI or non-AI enabled) is being used, or not. For example, [Commonwealth legislation](#) provides in many cases that computer-made decisions are treated as a decision by the government officer – ensuring the officer remains accountable. In the private sector, Corporate disclosure rules and Director duties (among other obligations) continue to apply when AI is being used. We therefore urge caution in introducing new specific legislation relating to liability for AI, as suggested in Proposal 10, before a careful assessment of existing legal frameworks on liability and accountability.

We also note different liability regimes may be more appropriate in certain contexts compared to others and flag the potential conflict a rebuttable presumption of liability may create – for example, its interaction with the presumption of innocence in a criminal law context.

- Legal moratorium on use of facial recognition technology (Proposal 11)

Facial recognition technology (FRT) has tremendous potential to create benefits for society. In Australia, [FRT has been trialed in ATMs to improve financial efficiency and security](#). FRT has also been used to help [diagnose rare genetic diseases identifiable by subtle changes in face shape](#) that are difficult to detect by human eye. At the same time, Microsoft also recognises the potential harms such technology may create in certain use cases and which need to be appropriately regulated and/or otherwise addressed – for example the risk that FRT, combined with an increase in video surveillance technologies in public spaces, could result in ongoing surveillance of individuals by law enforcement without adequate legal protections.

To manage these concerns, we have advocated for [facial recognition principles](#) and identified [areas of legal significance and potential reform](#). FRT should not be used as the sole basis for making consequential decisions and we believe that a human must be involved in such processes. We do not however believe a legal moratorium on facial recognition technology is appropriate. We note Proposal 11 suggests a moratorium only for applications of FRT that have a legal or similar effect and we assume applications of FRT which do not have such an effect to not be included.

We believe any moratorium on FRT may do more harm than good. For one, the realisation of benefits such as those mentioned above and further developments in the technology will be greatly hindered. Australia, as a striving leader in innovation, may be left behind in developing expertise relating to FRT. To be clear, we support the view that technological innovation must not be pursued at the expense of safety and other responsible practices, however we fear a legal moratorium on FRT may also hinder the development and improvement of best practices for the responsible development and deployment of FRT.

It will be important for further and coordinated debate in Australia on whether existing regulatory frameworks adequately address the risks generated by the use of facial recognition technology in its current state of development, and whether new regulation is required. Such a step must be undertaken before a legal moratorium is considered.

As the AHRC report recognises, law enforcement uses for FRT come with significant challenges. Microsoft has also acknowledged these challenges, in the context of the importance of balancing potential public safety benefits of FRT with potential harms if the technology is misused. One example of a potential benefit is the strengthening of identity verification to reduce the risk of identity theft. At the same time, there are clearly risks involved, and we have identified publicly the need for special care in law enforcement use of facial recognition, as well as an effort to put appropriate legal protections in place to prevent harmful uses like ongoing, warrantless surveillance of individuals<sup>1</sup>. A thorough evaluation of the appropriateness of a specific facial recognition system for the purpose it is being used for is required, on a case by

---

<sup>1</sup> <https://blogs.microsoft.com/on-the-issues/2018/07/13/facial-recognition-technology-the-need-for-public-regulation-and-corporate-responsibility/> and <https://blogs.microsoft.com/on-the-issues/2018/12/06/facial-recognition-its-time-for-action/>

case basis. For these reasons, we do not believe that a moratorium for law enforcement use is appropriate.

We believe there is a need for more intensive public debate, including government agencies and law enforcement, civil society groups, and other stakeholders, on the regulation of FRT. The general principles that Microsoft assesses should be addressed through regulation of facial recognition technology include:

- Requiring companies to provide both technical and non-technical documentation explaining the capabilities and limitations of the technology;
- Enabling testing and comparison, by requiring that providers of commercial facial recognition enable reasonable testing of systems for accuracy and unfair bias by independent, third parties;
- Requiring meaningful human review in consequential use cases;
- Requiring that facial recognition is not used as a basis for unlawful discrimination;
- Requiring the use of techniques for providing notice and ensuring consent appropriate to the circumstances.

## 2. Transparency and explainability

We note that several of the AHRC's proposals relate to ensuring transparency and accountability in consequential uses of AI. We agree that transparency and accountability are foundational principles, and think it is important to consider the following in the context of some of the AHRC proposals:

- Requirement to inform when AI is materially used in a decision that has a legal, or similarly significant, effect on an individual's rights (Proposal 5)

We support the general principle on providing transparency when AI is used in consequential use cases (Proposal 5). Thorough analysis would be required to identify whether there is a gap in Australia's legal framework in this regard. For example, obligations to provide notice and obtain consent exist (in relation to personal information) in the Privacy Act 1988. It would also be important to consider whether exceptions would be needed.

- Government AI pre-deployment procedures (Proposal 6 and 17)

We agree that Government (as with any entity) should undertake the necessary checks and assessments before deploying AI in consequential use cases. We would encourage an approach that looks to build on procurement and deployment practices currently used by Government. For example, exercises like conducting a cost-benefit analysis are likely to be a requirement (in different forms) of all government procurement. Rather than creating new, AI-specific review processes, our view is that improving information available to decision-makers in procurement processes involving AI is more likely to be effective. A key goal of providing such information should be to ensure that decision-makers in government are asking the right questions about whether a specific AI-enabled solution is appropriate for the particular use case being considered, including what potential human rights impacts the use of the solution may entail, and what mitigations should be considered and implemented.

A comprehensive review as outlined in Proposal 17 would be a major undertaking. The variety of use cases for AI deployment across government, many of which would not have consequential impacts, would make it impractical to achieve some of the goals of the proposed review (for example, a cost-benefit analysis of all uses of AI, as 17(b) appears to suggest). Our view is that a more practical step forward would be to build awareness across agencies and levels of governments of the range of potential use cases for AI, along with the opportunities and risks of AI deployments, and tools available for more informed decision-making on AI deployment. A focus should be in building capacity to assess whether there are salient risks (e.g. consequential impacts) in deploying an AI system and if so, conduct a more in-depth human rights impact assessment ascertaining the risk impacts, procedures to mitigate potential harms, and ensure remedy mechanisms are in place. The review undertaken by the [New Zealand Government](#) is a good example of an exercise to build awareness in government use of AI.

- [Legislating for Explainability in consequential use cases \(Proposal 7\)](#)

We agree with the principle of providing an explanation in consequential uses of AI but flag that detailed consideration will be needed to identify whether new legislation is required. Overly prescriptive legislation may do more harm than good especially as it may not be feasible to foresee all the specific developments and nuances in evolving use cases and technology. It will be important to identify whether there are potential use cases (involving software/AI or not) where existing mechanisms for transparency would not be sufficient.

As a more immediate measure, we recommend first considering whether the same goal of transparency can be achieved through less prescriptive methods such as internal review processes, explanatory notes, and guidelines.<sup>2</sup> This relates to the point made earlier in our submission that the appropriate mix of legislative and non-legislative approaches is needed for the responsible use of AI, especially given the length of time required to pass legislation.

- [No deployment without a reasonable explanation \(Proposal 8 and Questions B, C\)](#)

In relation to Proposal 8, we agree in principle that AI should not be deployed in consequential use cases without understanding how an output of AI was arrived at. We wish to note however that the ‘reasonableness’ of an explanation may vary according to context including the type of audience seeking an explanation, the use case of the AI, relevant security/confidentiality measures, as well as the technical limits of providing a complete explanation in certain AI models. We would seek clarification and to engage in further discussion on what is meant by a ‘reasonable explanation’.

As noted in Microsoft’s 2019 response to the Data 61 discussion paper on AI ethics, in terms of transparency, it is important to pursue contextual explanations geared to the needs of particular stakeholders. This may require explanation of one or more of a number of elements

---

<sup>2</sup> One example is the Microsoft [Transparency Note on Face API](#), which provides an explanation of the features; limitations and indicative uses of Face API, complementing the technical documentation available.

of a system, not solely focused on the code/algorithm itself. For example, in some cases, a stakeholder may require an explanation of why the resulting model produced a particular output or prediction, or the ways in which the broader system used that output.

In this regard, there are several ways to explain AI behaviour and there is much research being done to improve these methods and tools. These include data documentation and examination procedures, tools to approximate and test complex model behaviour, and design practices that opt for appropriately observable algorithms and models.<sup>3</sup> In light of such efforts in producing guidelines, toolkits and technical resources, we suggest Australian law reform as suggested in Question C is not appropriate at this stage.

It is also important to acknowledge that in certain instances, a complete explanation may not be appropriate or possible. There may be competing laws and interests such as protecting confidentiality, intellectual property and trade secrets etc, as well as deciding the level of explainability for AI-assisted decisions regarding topics of national security or other security-critical contexts (especially for Government uses of AI in decision making). Despite technical efforts referenced above, complete explainability may also not be technologically feasible. It is for these reasons that a rebuttable presumption of unlawfulness (in the event of an inability to present a reasonable explanation) as considered in Question B, may be problematic.

### [3. Pilots/testing/sandbox for human rights in practice \(Proposals 13, 14, 15\)](#)

We support the concept of developing and implementing practical initiatives such as sandboxes, human rights impact assessments and pilots to promote the protection of human rights. In forwarding the operationalising of responsible AI principles, Microsoft has participated in a number of pilots and projects with government agencies, industry and other stakeholders to great effect; generating lessons that are relevant in the Australian context.

We highly encourage governments to work collaboratively with industry on piloting and providing feedback on principles (including the protection of human rights). This should be done in a 'sandbox' way so that there can be open sharing of lessons learned.

Microsoft is participating in the Department of Industry, Science, Energy and Resources exercise to pilot Australia's AI Ethics Principles, launched in 2019. This is an appropriate way of testing approaches to incorporating human rights in the design process, in particular because respecting human-centred values (and protecting human rights) is one of Australia's 8 core principles. One important consideration will be how to achieve a pragmatic approach to assessing human rights impacts in the design (or deployment) of AI, and incorporate potential mitigations. The pilot provides an important opportunity for this.

---

<sup>3</sup> Examples include work by Microsoft and the Partnership on AI on producing 'Datasheets for datasets' increasing transparency in data used to train AI models, Microsoft's [Interpret ML toolkit](#) available open-source on Github assisting interpretability of machine learning models, and procedures for documentation of decision-making reasoning for auditing purposes.

#### 4. Education and awareness raising (Proposals 16, 25-28)

We agree that education and awareness raising is important (relating to various proposals, including 16, and 25-28). One important priority for awareness raising is in sharing practical examples of putting responsible AI principles into practice, including how the development and deployment of AI technologies can reflect human rights considerations. As noted above, the Department's pilot trials with industry to test the application of AI principles can provide an important source of information in this regard. We note that the pilot of Australia's AI ethics principles will be one of a number of initiatives, as well as examples being generated by multi-stakeholder efforts like the Partnership on AI. It is important that awareness-raising initiatives in Australia draw on examples from across the world.

One other consideration is that in Microsoft's experience, a variety of roles within organisations are relevant in implementing responsible approaches to AI. For example, developers working on software are important and need information and tools to evaluate the impact of decisions they might make on the data used for training an AI model, or the type of model used for a specific application. However, the issues should not be left to developers alone, and a range of roles, including legal, compliance, sales, communications, and training, as well as an organisation's leadership, all have a role to play. Education and awareness-raising efforts need to be tailored appropriately. For example, Microsoft's AI Business School is a free online curriculum on AI aimed at business decision makers across industry sectors, and includes content on responsible AI and AI governance targeted at their level. For developers working on AI, there are various technical trainings available that focus on specific issues like bias detection and mitigation in the use of datasets for AI models. To be most effective, it is likely that educational and awareness-raising content is pitched appropriately at these roles. Another issue is the need to ensure education and training includes building capacity to understand how emerging technologies impact and interact with different stakeholders including vulnerable groups. Ensuring civil society is included in the discussion, and ensuring advocacy groups know the right questions to ask will be an important part of ensuring human rights are not impacted.

#### 5. AI Safety Commissioner (Proposal 19)

Creating a standalone statutory position to "focus on the prevention of individual and community harm, and the promotion and protection of human rights" for AI seems to miss the fact that AI is a set of technologies not a single technology, and that they are used across a wide range of contexts. Importantly, the way in which AI technologies are used in the context of an overall system that includes human decision-makers, and other non-technology factors, is what needs to be considered when determining whether AI is used responsibly.

In this context, overseeing the use of AI technologies and whether they are used in line with Australian legal obligations, including with regard to human rights, is already the responsibility of various specific regulators and subject to different laws. For example, the use of AI technologies in the context of medical devices, the use of AI technologies in a financial services



context, or the use of AI technologies in motor vehicles, are all going to be under the purview of existing regulators.

Implementation of Australia's AI Ethics Principles is also a relevant consideration. The Commonwealth Government has issued AI Ethics Principles, which include respect for human rights as a core principle. There is also a clear ongoing set of priorities for Government to promote implementation of the principles, including pilots and sharing of best practices. This means that establishing a new statutory position with a very similar mandate would create an overlap in responsibilities and result in confusion. Rather than creating a new agency with the same mandate, it would be more constructive to focus on practical ways in which tools, practices and resources on implementing the AI Ethics Principles can be shared, including those relating to human rights.

## 6. Accessibility (Proposal 20, 21)

Microsoft supports the emphasis placed by the AHRC report on the potential for AI to improve accessibility. We agree that it is essential to ensure that people with a disability have the tools and technology that they need to be productive and fulfilled members of the workforce and society.

Microsoft supports the use of recognised global and local accessibility standards by Government (Proposal 20). Australia adopted the EN301549 standard: *Accessibility requirements suitable for public procurement of ICT products and services* in 2016, and while a number of governments have referenced the standard as part of procurement activities, the NSW Government remains the only jurisdiction to have formally adopted the standard as a [mandatory requirement in procurement rules](#).

Microsoft also supports the development of policies that improve the accessibility of Government Service Delivery. Most Australian governments have strong guidance and policies around developing services for accessibility and inclusivity – such as Criteria number 9 of the Australian Government's Digital Services Standard, [Make it Accessible](#) and the NSW Government's Digital Design Standard [Design with users, for users](#). Ensuring that agencies comply with these policies should be a key focus for all levels of government.

Microsoft would welcome incentives that improved broader compliance with accessibility standards (Proposal 21). Any inquiry undertaken in this area should also consider how to encourage senior executive awareness and board level support for accessibility and growing the support available to organisations seeking to implement good practice – such as accessibility design and development expertise – which is in relatively short supply in Australia.

## 7. AI and human rights – beyond the technology

As a general point, we think it is important to recognise that ensuring the responsible use of AI in line with human rights objectives requires attention not just to the technical features of a

particular AI model, but to a range of other factors that make up the 'AI decision-making system'.

This includes ensuring adequate protections are implemented within the selection of data used as inputs for the training and use of the model; the appropriateness of the model used for any given use case; and various non-technology factors like the role and level of human review; the human use of AI outputs and the capacity of a human reviewer or decision-maker to understand AI reasoning; whether there are opportunities for recourse/an ability to challenge a decision and other factors.

We therefore suggest the wording in, for example, Proposal 18 (by requiring a 'system' include adequate human rights protections) may be too simple and therefore misleading – without an appropriate explanation that protections surrounding an AI-informed decision making system are to relate to the technical AI system as well as important non-technical factors. Without such an explanation, there is a risk deployers of AI may be misdirected from identifying their relevant risks and putting appropriate mitigations in place.

### Conclusion

Microsoft welcomes ongoing engagement with the AHRC, and other interested parties on the best way to ensure responsible use of AI which is grounded in human rights.