Submission to the Department of Industry, Science and Resources

regarding

Safe and Responsible AI

26 July 2023

Digital Rights Watch is a charity organisation founded in 2016 whose mission is to ensure that people in Australia are equipped, empowered and enabled to uphold their digital rights. We stand for Privacy, Democracy, Fairness & Freedom in a digital age. We believe that digital rights are human rights which see their expression online. We educate, campaign, and advocate for a digital environment where individuals have the power to maintain their human rights.[1]

[1] Learn more about our work on our website: https://digitalrightswatch.org.au/

Overview

Digital Rights Watch (DRW) welcomes the opportunity to submit comments to the Department of Industry, Science and Resources regarding the *Safe and Responsible AI in Australia Discussion Paper* (the Discussion Paper). We are pleased to note the Australian government's willingness to consider governance mechanisms to ensure AI is developed and used safely and responsibly in Australia, including the consideration of regulations, standards, tools, frameworks, principles and business practices.

As Australia's leading digital rights organisation, DRW is primarily concerned with the implications of AI and automated decision making (ADM) systems for the human rights, safety and wellbeing of individuals and communities. We actively participate in public consultations regarding the development of legislation and policy in relation to technology and human rights. Our recent submissions relevant to AI regulation and governance include:

- Submission to the Digital Technology Taskforce in response to 'Positioning Australia as a leader in digital economy regulation - Automated Decision Making and AI Regulation' Issues Paper[2]
- Submission to the Senate Economics Committee Inquiry into the influence of international digital platforms[3]

"AI" can be a slippery concept that has different meanings and purposes depending on who is using it and why. DRW believes the adoption of ISO definitions by the Department is sensible, however we do note that defining technology—especially AI technologies—can often be a point of contention, and may present drafting challenges in the regulatory context.

DRW has previously expressed concern about the priorities and framing of AI regulation in Australia. The Issues Paper published by the former Digital Technology Taskforce represented an approach that focused too heavily on reducing regulatory barriers for economic gain, without due regard to the ways that AI and ADM can create significant privacy and security risks, and lead to adverse impacts for individuals and communities. We are pleased to note that the current Discussion Paper has shifted focus to also include considerations regarding responsibility, governance, safeguards, trust and confidence.

[2] Digital Rights Watch Submission to the Digital Technology Taskforce on the Issues Paper 'Positioning Australia as a leader in digital economy regulation - Automated Decision Making and AI Regulation', 22 April 2022, Available at: https://digitalrightswatch.org.au/2022/04/22/submission-regulating-ai-and-automated-decision-making-in-australia/

[3] Digital Rights Watch Submission to the Senate Economics Committee inquiry into the influence of international digital platforms, 14 March 2023. Available at: https://digitalrightswatch.org.au/2023/04/26/democratising-digital-economies/

We also note that the Department is seeking system-wide feedback on actions that can be taken across the economy on AI regulation and governance. We welcome this big-picture approach, as the potential impacts and requisite regulatory levers cannot be contained to a single mechanism or government department. A systematic and coordinated approach is necessary in order for AI regulation to be meaningful and effective.

Placing human rights at the centre of AI regulation

DRW understands the genuine interest in possible economic and social benefits promised by AI and ADM. There are many areas in which these technologies may create immense public good, for example, in medical sciences and early detection of diseases. Robust regulation that places human rights and safety at the centre is not a threat to this kind of technological innovation.

However, these technologies also present significant challenges to privacy and digital security, and can result in biased, discriminatory or other harmful outcomes. This is of particular concern should an AI system result in individuals or groups being unable to access essential government, health or financial support and services, or where it is used in disciplinary, judicial or policing contexts.

There is a wealth of recent examples documenting ways that AI and ADM technologies can result in significant harm, such as:

- **individual harm**, for example by facial recognition used by law enforcement resulting in wrongful arrest,[4]
- **collective harm** affecting entire groups, for example as a result of racial profiling through predictive policing,[5]
- **harms of allocation** arising from discriminatory allotment of, or unequal access to, products or resources,[6]

[4] This has occurred at least three times, see for example: https://www.nytimes.com/2020/12/29/technology/facial-recognition-misidentify-jail.html

[5] See, for example: 'Technology can't predict crime, it can only weaponise proximity to policing,' Electronic Frontiers Foundation, September 2020, https://www.eff.org/deeplinks/2020/09/technology-cant-predict-crime-it-can-only-weaponize-proximity-policing

[6] For example, in 2019 Apple was accused of discrimination after offering a lower credit limit to a woman compared to a man with a similar credit rating. See: 'Apple Card Investigated after gender discrimination complaints,' The New York Times, November 2019. See: https://www.nytimes.com/2019/11/10/business/Apple-credit-card-investigation.html

- **harms of representation** which reinforce existing discrimination, disadvantage or stigma, often by using historical datasets which contain biased, incomplete or outdated data.[7]

Perhaps one of the most egregious and well-known examples of harm caused by an algorithmic system is the commercially-available risk assessment tool called COMPAS, used in the US to predict recidivism in applications for parole and to assess a criminal defendant's future likelihood of committing a crime. This tool was found to be racially biased—inaccurately predicting that Black defendants were twice as likely to reoffend than white defendants—and notably, no more accurate or fair than predictions made by people with little to no criminal justice experience.[8]

If Australia is seeking to be a leader in digital economy regulation and earn public trust and confidence regarding the use of AI—especially in the public sector—it is essential that we learn from these and other examples of AI and ADM creating or exacerbating harm here and around the world.

In addition to regulatory mechanisms specific to AI, Digital Rights Watch strongly suggests that the Australian government prioritise the creation and enactment of a federal Human Rights Act. Doing so would:

- assist in the creation of a rights-respecting culture in Australia,
- ensure that human rights are proactively considered in any new legislation related to AI,
- create a powerful tool to challenge injustice, including where facilitated by AI and ADM technologies, and
- provide opportunities for people to take action and seek justice where their rights have been violated.

We also support the creation of a separate but complementary Charter of Digital Rights and Principles, which could specifically focus on the application of human rights to existing and emerging technologies.[9] For example, the European Union's

[7] For example, Microsoft found gender bias arose in models based on data that contained gendered stereotypes. See 'Man is to computer programmer as woman is to homemaker? Debiasing word embeddings,' Microsoft, 2016. https://www.microsoft.com/en-us/research/publication/quantifying-reducing-stereotypes-word-embeddings/

[8] 'The accuracy, fairness, and limits of predicting recidivism,' Julia Dressel and Hany Farid, Science Advances, Volume 4, Number 1, January 2018. https://advances.sciencemag.org/content/4/1/eaao5580; 'Machine Bias,' ProPublica, May 2016. See https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[9] We suggest that this could be modelled on the European Union's Declaration of Digital Rights and Principles. For more detail see: Digital Rights Watch Submission to The Parliamentary Joint

Declaration on Digital Rights and Principles was designed to complement existing rights, and to provide guidance for the European Union and its member states as they pursue "human centric" and "sustainable" digital transformation.[10]

Given that Australia is embarking upon our own digital transformation (not limited to AI and ADM, but also the reinvigoration of myGov and the establishment of a national digital identity), a guiding set of overarching digital rights and principles would be useful to ensure that Australia's digital future is grounded in human rights, safety and dignity of all Australians.

Recommendation

Enact a comprehensive federal Human Rights Act, as well as a separate but complementary Charter of Digital Rights and Principles.

AI Governance

We need not look to far-future hypothetical scenarios to understand the ways in which AI can cause harm: it is already happening. There is over a decade of case studies from around the world, research, analysis and recommendations to draw from. More than ever before, Australia is in a position to move from identifying problems and toward taking steps to remediate and mitigate them. Digital Rights Watch urges the Department to take this task seriously, and to recognise that there is nothing about AI that is inevitable. The government can—and should—intervene.

Since the unveiling of ChatGPT, a new wave of AI hype has started that shows no sign of abating. Over this period, the media has been awash with reporting of high profile figures in the AI industry sounding the alarm on the so-called "existential risk" of AI.

It is essential that we address AI's role and impact, not as a philosophical futurist exercise, but as something that is being used to shape the world around us right *now*. We urge the Department not to become distracted by longtermist hype,

Committee on Human Rights regarding the Inquiry into Australia's Human Rights Framework, 29 June 2023. Available at: https://digitalrightswatch.org.au/2023/07/11/efa-sub-human-rights/

[10] European Digital Rights and Principles, *European Commission*. Available at: https://digital-strategy.ec.europa.eu/en/policies/digital-principles/

fearmongering and speculative fiction.[11] Critically, much of the AI hype—both negative and positive—serves the interests of companies who stand to profit the most from widespread adoption of their products in a low regulation environment. We should not allow our laws and policy to be shaped by AI Industry leaders for their own purposes, especially given that those leaders are generally not based in Australia, and represent a different set of values that do not always apply well in the Australian context.

To that end, the following sections explore considerations regarding approaches to regulating AI and its harms, *now*.

Mechanisms for AI regulation

In order to meaningfully address AI—the technology itself, the data it generates and uses, and the power and influence of the broader AI industry—we will need a combination of regulatory measures.

Digital Rights Watch supports the creation and enactment of a standalone AI-specific law, *in addition to* improvements to other regulatory levers. Some harms to individuals caused by AI are related to a breach of existing bodies of law (e.g. privacy, copyright, consumer, anti-discrimination, labour and defamation[12] laws), and we support bolstering these laws where there are gaps.[13] This is especially apparent in the *Privacy Act 1988*. We specifically address data governance and regulation of AI through privacy law in the section below on data governance.

However, this will not go far enough to comprehensively address the scale and scope of risks presented by AI, which would be better addressed by a standalone AI-specific law.

[11] For analysis on the longtermist ideology underpinning many tech leaders' concerns regarding AI, see: ‘The Wide Angle: Understanding TESCREAL — the Weird Ideologies Behind Silicon Valley’s Rightward Turn,’ *The Washington Post,* 1 May 2023. Available at: https://washingtonspectator.org/understanding-tescreal-silicon-valleys-rightward-turn/

For analysis on the so-called existential risk of AI as a distraction from harms already occurring, see: ‘AI Doesn’t Pose an Existential Risk—but Silicon Valley Does,’ *The Nation,* 7 June 2023. Available at: https://www.thenation.com/article/economy/artificial-intelligence-silicon-valley/

[12] The intersection of generative AI technology and Australian defamation law has already come up at least once since the advent of ChatGPT. See ‘Australian mayor prepares world’s first defamation lawsuit over ChatGPT content,’ *The Guardian,* 6 April 2023. Available at: https://www.theguardian.com/technology/2023/apr/06/australian-mayor-prepares-worlds-first-defamation-lawsuit-over-chatgpt-content

[13] For a breakdown of common harms to individuals caused by AI and existing laws that may apply, see Table 3, ‘The State of AI Governance in Australia,’ *Human Technology Institute,* 31 May 2023. Page 34. Available at: https://www.uts.edu.au/human-technology-institute/projects/future-ai-regulation-australia

Considerations for AI-specific regulation

We note that there are already a significant number of 'ethical' AI principles and frameworks in Australia and globally. Many of these sets of principles, frameworks and codes have conflicting requirements and considerations, making it challenging for organisations to understand their obligations (if any), and how to implement best practices.

Australia's AI Ethics Framework, containing eight voluntary AI Ethics Principles are functionally ineffective. While we appreciate that they align with the OECD's Principles on AI, they do not function as useful guardrails for organisations when implementing AI, as they lack enough detail about what to do to achieve a principle. Similar to many voluntary tech ethics frameworks, they provide a collection of nice statements that organisations and executives are encouraged to think about, but not much else. We suggest that any further development of AI principles should adopt the approach from NIST's AI Risk Management Framework (RMF), which proposes a series of technical measures as suggested ways to implement each principle.[14]

While ethics and principle-based approaches may be able to provide a framework to understand the intention, design and establish shared values, they are not sufficient as an AI governance strategy alone. For example, empirical research has shown that AI principles and codes of ethics have minimal impact on the behaviour of engineers developing AI systems.[15] The Human Technology Institute notes that "such principles may induce a false sense of security that the problems have been managed."[16] This does not mean that codifying AI principles is pointless, but that they must be supported by practical strategies, policies, structures and enforcement, in order for them to be effective.

Consistency with international frameworks

When developing Australia's approach to AI regulation, it is important that there is a level of functional consistency with other leading international instruments. This will help to position Australia well within the international community, attract trade and investment, and promote compliance.

Currently, different jurisdictions and standards developers are starting to approach AI governance in different ways. Developers of AI technology and deployers of AI (in this instance we mean specifically multinational organisations

[14] 'AI Risk Management Framework,' *National Institute of Standards and Technology* (NIST). Available at: https://www.nist.gov/itl/ai-risk-management-framework

[15] Thilo Hagendorff, 'The Ethics of AI Ethics: An Evaluation of Guidelines' (2020) 30 Minds and Machines 99.

[16] 'The State of AI Governance in Australia,' *Human Technology Institute*, 31 May 2023. Available at: https://www.uts.edu.au/human-technology-institute/projects/future-ai-regulation-australia

wishing to employ AI for commercial practices) consequently have the task of identifying the different rules and laws associated with AI governance and then attempting to apply them.

This has several consequences. First, the plethora of different approaches often means that identification of the applicable rule or law is missed. Second, even when identified appropriately, attempting to apply all of the relevant rules or laws becomes difficult or they may conflict with one another. Lastly, in an effort to address the former two issues, many developers and deployers select one law or standard as their 'baseline', accepting the risk that relying on their baseline may mean that they are non-compliant with all requirements across all jurisdictions.

Having a consistent approach to AI regulation with other leading international instruments would resolve many of these issues. At a minimum, DRW recommends consideration from:

- the EU's General Data Protection Regulation (GDPR),
- the US's Children's Online Privacy Protection Act (COPPA),
- the EU's new AI Act,
- and standards like ISO 23894:2023 and NIST's AI RMF.

Types of regulation

In some instances, voluntary mechanisms may be effective where organisations and companies wish to demonstrate that they use AI in a trustworthy way, to build trust and credibility with their stakeholders. However, voluntary mechanisms fall short where there is no regulatory 'stick' or consequence for a lack of compliance. In Australia, we can observe some of the challenges of voluntary mechanisms in areas of complex tech policy. For example, the voluntary misinformation code has been criticised for not effectively dealing with the issue of mis- and dis- information.[17]

We also note that the co-regulation approach currently favoured in Australia also has its own set of challenges. As the development of the Industry Codes under the *Online Safety Act* has shown:

- the process can take a very long time;

[17] For example, see 'Digital code of conduct fails to stop all harms of misinformation, Acma warns,' *The Guardian,* 21 March 2022. Available at: https://www.theguardian.com/media/2022/mar/21/digital-code-of-conduct-fails-to-stop-all-harms-of-misinformation-acma-warns

See also: 'The DIGI code review - a missed opportunity?' *University of Technology Sydney,* 15 June 2022. Available at: https://www.uts.edu.au/research/centre-media-transition/news/digi-code-review-missed-opportunity

- the process devolves quite a lot of power to industry and the regulator to determine best approaches;
- it can result in requirements becoming established that go beyond the scope of the original legislation, established with little oversight or scrutiny, opportunity for public debate, and without processes such as assessing human rights compatibility that are associated with the creation of legislation through parliament.

We suggest the Department consider ways to account for these factors when deliberating over co- or self- regulation approaches.

Digital Rights Watch suggests that the Department consider adopting an Australian version of the approach set out in the EU AI Act, including a sophisticated risk-based approach which we discuss in further detail below. We also recommend that corresponding technical recommendations, measures or standards are developed with regard to specific uses or types of AI, such that organisations are equipped to implement principles in practice.

A risk-based approach to AI regulation

In general, Digital Rights Watch supports a risk-based approach to AI regulation. When implemented well, a risk-based approach can be a sensible way to apply appropriately graduated safeguards. However, we are concerned that the presented risk management table in the Discussion Paper (Box 4) does not have the level of sophistication required to adequately identify and assess risks of AI.

In particular, we are concerned that the chart flattens out risks of the technologies themselves and their use cases. For example, when it comes to considering risk and safety of motor vehicles, we have risk mitigation strategies in place for both the technology (safety standards, requirements for certain features etc), as well as its use (speed limits, laws against driving while intoxicated or underage). The current proposed tiers of risk do not adequately accommodate for this distinction.

We are also concerned that the risk approach presented is predicated upon the ability to understand and anticipate the future impact in order to designate which tier it would sit within. Yet AI systems can deliver harm at immense speed and scale, in ways that are not always predictable and easily anticipated. Risks from AI systems can also manifest across their lifecycle and these systems can be dynamic, for instance their behaviour can change with new inputs and data, or when integrated into a different environment. A robust risk-based approach needs to account for such changes, for instance by integrating monitoring or re-visitations of risk assessments.

We suggest that in addition to anticipating the *impact* of AI systems, the risk-based approach should also factor in considerations of the context in which they are to be deployed. For example, given the potential severity of consequence, use of AI systems in law enforcement ought to be in a higher risk category by default. The use of AI or ADM systems in critical sectors such as healthcare, justice, housing and access to government services or support should also be considered higher risk.

DRW also suggests that the type of information that is to be used to build, train, test and as input into AI systems should be considered as a factor for determining level of risk. For instance, AI systems that use personal or sensitive information (as defined in the Privacy Act) or protected attributes (under anti-discrimination law), ought to carry with them a higher risk status.

The current risk management approach does not account for the possibility of cumulative impacts of "low risk" AI systems, such as longer term changes in behaviour (for example, spending behaviour or harmful applications like online gambling), or the potential increase for privacy risk as some "low risk" systems gather more and more personal information about a user.

The Ada Lovelace Institute has highlighted four ways to think about risks from AI systems, which we suggest the Departement consider as it further develops its risk-based approach:

- **Risks of particular harms** - such as representational harms, harms to equality, informational harms, physical or emotional harms, human rights infringements, etc

- **Risks associated with scenarios of AI systems** - such as best- or worst-case scenarios, system failure, malicious use or misuse, or when the context around the AI systems change

- **Risks associated with particular AI technologies** - where some particular models or forms of AI have commonly associated risks, such as facial recognition technology

- **Risks associated with specific domains of application** - context specific risks such as those specific to healthcare, law enforcement, or risks related to the economy or the environment[18]

Finally, we note that the EU AI Act also implements a risk-based approach, which includes consideration of 'unacceptable risk' AI systems. Such systems are

[18] 'AI Risk: Ensuring effective assessment and mitigation across the AI lifecycle,' *Ada Lovelace Institute,* 18 July 2023. Available at: https://www.adalovelaceinstitute.org/report/risks-ai-systems/

considered to be a threat to people and to be banned.[19] This includes, with some exceptions:

- Cognitive behavioural manipulation of people or specific vulnerable groups
- Social scoring: classifying people based on behaviour, socio-economic status or personal characteristics, and
- Real-time and remote biometric identification systems, such as facial recognition.

We strongly suggest that the Department consider adding a similar 'unacceptable risk' category to its matrix, to define prohibited AI practices.

Improving political coordination on digital and technology policy

How digital technologies—the internet, digital platforms, and of course, AI—are governed impacts everyone, and touches every part of the government's work. Too often, digital and technology policy are treated as an afterthought, or as a standalone issue. Digital Rights Watch is concerned that policy fragmentation in tech is a major roadblock to Australia's capacity to respond to emerging risks (as well as opportunities), despite being home to some of the world's leading academics, technologists, and entrepreneurs in this field.

We suggest the establishment of a new Joint Standing Committee on Digital Affairs and be a driving force to better allocate time, resources and expertise to develop a more sophisticated approach to digital and technology policy broadly, but especially as AI and ADM become more prevalent. By establishing a new standing committee on digital and technology policy and appointing a Minister for Digital Capabilities, we can start the work of building and improving tech policy expertise within Parliament and the public sector.

Recommendations

1. Develop an AI-specific law that adopts an approach in line with the EU's new AI Act. This should happen *in addition* to strengthening other indirect regulatory levers, such as privacy, copyright, IP, defamation and consumer protection laws.
2. Revisit the draft risk management approach for managing AI risks (Box 4 in the Discussion Paper) to establish a more sophisticated

[19] The EU AI Act, https://artificialintelligenceact.eu/the-act/

framework for understanding, assessing and managing risks associated with AI systems.

In particular:

- Consider the distinction between assessing and managing risks inherent to AI technologies as separate to the risks associated with the use of such technologies.
- Establish an ‘unacceptable risk’ or ‘no-go zone’ category.
- Consider other factors for assessing risk, such as whether the AI system involves a **critical sector** (such as healthcare, welfare, housing, insurance, employment, education, political processes, the legal system and law enforcement), whether it uses **data inputs** containing protected attributes or sensitive information, and if it is designed to make decisions or predictions about or for **vulnerable or marginalised groups**.

3. Establish a Joint Standing Committee on Digital Affairs.

Data Governance

Given that AI techniques require the use of data to be trained, tested and as inputs to function, an essential component of AI regulation and governance must also consider the regulation and governance of data itself.

Many of the most powerful and established tech corporations in the AI Industry that stand to profit most from the current AI boom are those who have generated, accumulated and monetised huge amounts of personal information for years—often benefitting from relatively light-touch regulation.

These companies have a significant *data advantage,* after years of accumulating vast amounts of data, including personal information.[20] They have also established a social norm and expectation that companies will collect and monetise our personal information for their own benefit—in order to challenge the risks and harms of AI, we must also challenge this norm.

Too often, those who are subject to data extractivism, that is, those who have their data harvested, have no say on how and for what purpose their data will be used.

[20] For analysis on the data advantage of big tech companies within the AI industry, see ‘2023 Landscape: Confronting Tech Power’, *AI Now,* 2023. Available at: https://ainowinstitute.org/wp-content/uploads/2023/04/AI-Now-2023-Landscape-Report-FINAL.pdf

This represents a serious lack of representation from affected communities, as well as a missed opportunity for socially-responsible technology development that meets the genuine needs and local priorities of communities.[21]

In many ways, the AI industry benefits from the past decade of privacy-invasive data harvesting practices under the ideology of surveillance capitalism. Privacy and data protection regulations cannot address *all* of the issues created or exacerbated by AI, but by regulating the data, we regulate the fuel upon which the AI industry is being built.

For example, in May 2023 the US Federal Trade Commission ruled that Amazon Ring illegally surveilled customers and proposed an order that would prohibit Ring from profiting from unlawfully collected data.[22] Preventing companies from profiting from unlawful data collection is a great example of how privacy and data protection legislation can, in turn, regulate the development and implementation of AI by limiting the on-use of that data.

Regulation of AI through privacy law

Not all uses of AI and ADM will handle personal information or raise obvious privacy issues. However, those that do process personal information, make predictions or decisions based on personal information as data inputs, or are designed to interact with or have an impact upon individuals and communities will likely raise some privacy and data protection considerations.

Many of the harms that arise from AI and ADM stem from inappropriate collection and use of personal information. As such, robust privacy regulation can go a long way toward mitigating privacy-related harms caused by AI.

One important factor to consider is **data provenance**, that is, how the data was originally collected, by whom, and for what purpose. It is not uncommon for data inputs to algorithmic systems to be a *secondary use* of personal information which was originally collected for a different primary purpose, or for that data to have been collected in ways that may not have been ethical, or not reasonably expected or understood by individuals.[23] In many cases, data used as inputs to build, train, test and otherwise run AI models has been obtained from the data broker industry, which collates, aggregates and sells data. Tightening restrictions on secondary use of personal information and limiting the on-sharing and selling

[21] See, for example, the work of Connected by Data: https://connectedbydata.org/

[22] 'FTC Says Ring Employees Illegally Surveilled Customers, Failed to Stop Hackers from Taking Control of Users' Cameras,' *Federal Trade Commission,* 31 May 2023. Available at: https://www.ftc.gov/news-events/news/press-releases/2023/05/ftc-says-ring-employees-illegally-surveilled-customers-failed-stop-hackers-taking-control-users

[23] 'Algorithms, AI, and Automated Decision — A guide for privacy professionals,' Salinger Privacy, 2021. Available at: https://www.salingerprivacy.com.au/downloads/algorithms-guide/

of personal information in the data broker industry would, in turn, mitigate some inappropriate or legally-ambiguous uses of data in the AI ecosystem.

The Department may also wish to consider developing requirements for AI providers to declare the data sources for their foundational models as well as any subsequent data or instruction sets given in training and fine-tuning the model into “alignment”. Doing so would assist the public and/or regulatory bodies and researchers to understand and evaluate the data and directives upon which AI systems have been built.

In some cases, machine learning models are able to make inferences about individuals based on other, seemingly benign data points. For example, Facebook has long been able to predict sensitive information such as sexuality and political beliefs based on behavioural data.[24] The Office of the Australian Privacy Commissioner (OAIC) has referred to this practice as “**collection via creation**” and has issued advice that the generation or inference of personal or sensitive information, based on other data, is considered to be “collection” under the Australian Privacy Principles (APPs).[25] Formalising this is one of many proposed amendments to strengthen the Privacy Act.[26] Doing so would ensure that entities using AI to infer personal or sensitive details about individuals would need to meet the requirements of the APPs, and offer people in Australia an added level of protection.

It is also possible, and common, for AI models to discriminate based on “proxy variables” which may not readily appear to be personal information, but can still lead to unfair or discriminatory outcomes.[27] For example, the use of post codes in machine learning systems has been shown to act as a proxy for race.[28] Addressing the shortcomings in the Privacy Act such as the definition of personal information, the exploitation of so-called de-identified data, the issue of individuation, and improving transparency and explanation mechanisms are

[24] For example, see ‘Facebook users unwittingly revealing intimate secrets, study finds,’ *The Guardian,* 12 Mary 2013, available at: https://www.theguardian.com/technology/2013/mar/11/facebook-users-reveal-intimate-secrets

[25] ‘Guide to Data Analytics and the Australian Privacy Principles,’ *Office of the Australian Information Commissioner (OAIC),* March 2018. Available at: https://www.oaic.gov.au/privacy/guidance-and-advice?a=3086

[26] Specifically, by updating the definition of ‘personal information’, proposal 4.3 in the privacy act review report.

[27] ‘Using artificial intelligence to make decisions: Addressing the problem of algorithmic bias,’ Australian Human Rights Commission, 2020, page 11. Available here: https://humanrights.gov.au/our-work/rights-and-freedoms/publications/using-artificial-intelligence-make-decisions-addressing

[28] ‘Calculating Race: Racial Discrimination in Risk Assessment’, Benjamin Wiggins. Oxford University Press (2020)

important improvements that will have flow on effects with regard to how AI is able to be legally developed and deployed.

Adoption of AI models also brings the possibility of privacy threats via overfitting: when the outputs of an AI model conform too closely to the data on which the model was originally trained. This issue presents privacy risks when the training data contains sensitive information which may subsequently be leaked in the model's output by accident or through malicious attempts to retrieve it.[29] With increased use of such models, the Privacy Act must account for issues such as overfitting risks to ensure that this type of privacy breach is adequately safeguarded against by AI providers before their products reach the public, and that redress for breaches of this nature are directly addressed rather than left in a legal grey area.

While robust privacy regulation cannot solve or mitigate *all* risks raised by AI technologies, strengthening the Privacy Act will play a fundamental role in upholding people's right to privacy and the protection of their personal information as AI becomes increasingly commonplace.[30]

Recommendations

1. Comprehensively reform the *Privacy Act 1988*.
2. Introduce a range of individual rights with regard to the use of AI and ADM, including:
    a. A right to object to and opt-out of ADM
    b. A right to review and appeal a decision made wholly or partly by automated means
    c. A right to an accessible, plain language explanation about how the AI/ADM system works, how a decision has been made, and the personal information used

[29] 'Overfitting, robustness, and malicious algorithms: A study of potential causes of privacy risk in machine learning', Samuel Yeom et al., *Journal of Computer Security*, 4 February 2020. Available at: https://content.iospress.com/articles/journal-of-computer-security/jcs191362

[30] For further detail on the ways the Privacy Act should be reformed, see Digital Rights Watch Submission to the Attorney-General's Department on the 2022 Report regarding the review of the *Privacy Act 1988*, 31 March 2023. Available at: https://digitalrightswatch.org.au/2023/04/03/submission-privacy-act-review-report/

Biometric data and surveillance

The use of biometric information is not currently specifically regulated in Australia despite the significant risks to privacy and security should it be misused. Rather, it is included under "sensitive information" in the Privacy Act. Similarly, there are currently no specific limitations on the use of facial recognition technology—one of the most invasive and controversial applications of AI.

A 2021 report produced by the Australian Human Rights Commissioner (AHRC), which is referenced in the discussion paper, suggests "a moratorium on the use of facial recognition and other biometric technology in decision making that has a legal, or similarly significant, effect for individuals, or where there is a high risk to human rights, such as in policing and law enforcement."[31]

The Department may wish to look to international precedents when considering the regulation of facial recognition, including the bans on its use in San Francisco and Maine, as well as significant limitations on it in Virginia, Massachusetts and Washington.[32]

With regard to restricting the use of biometric information by the private sector, the Department may consider the Biometric Information Privacy Act (BIPA) passed unanimously in 2008 in Illinois.[33] The BIPA imposes obligations and prohibitions on how private companies can handle biometric information, as well as a private right of action to allow any person aggrieved by a violation to bring an action in court. A specific right of action for misuse of biometric information would be a useful and significant inclusion in any regulation of biometric information.

Recommendations

1. Strictly limit the use of one-to-one facial recognition systems and other biometric surveillance technologies.
2. Ban the use of one-to-many facial recognition systems which enable real-time and/or mass surveillance.

[31] See https://tech.humanrights.gov.au/artificial-intelligence/facial-recognition-biometric-tech

[32] See, for example, 'Main passes the strongest state facial recognition ban yet,' The Verge, 2021. Available here: https://www.theverge.com/2021/6/30/22557516/maine-facial-recognition-ban-state-law

[33] Illinois General Assembly, Biometric Information Privacy Act https://www.ilga.gov/legislation/ilcs/ilcs3.asp?ActID=3004&ChapterID=57

Additional resources

In addition to the references included throughout this submission, we wish to raise attention to a few other areas of work relevant to the regulation of AI and ADM which we suggest the Department consider as they progress:

- For an examination of the power of 'Big Tech' and the AI Industry, we suggest AI Now's 2023 'Confronting Tech Power' Report.

- For an explanation on the need for public and collective forms of data governance, we suggest 'A Relational Theory of Data Governance' by Salomé Viljoen, and 'Everyone should decide how their digital data are used—not just tech companies' by Jathan Sadowski, Salomé Viljoen and Meredith Whittaker.

- For a deep dive into the relationship between information privacy law, harm, and risk assessments in relation to AI and ADM, we suggest 'Algorithms, AI, and Automated Decisions — A guide for privacy professionals' from Salinger Privacy.

- For an analysis of approaches to risk-based assessments for AI, we suggest the Ada Lovelace Institute report: 'AI risk: Ensuring effective assessment and mitigation across the AI lifecycle'.

- For an introduction into the growing field of work being done in data sovereignty and Indigenous AI Protocols we recommend the work being conducted by Australian National University, Old Ways, New, and the Goethe Institute in their Indigenous Protocols // AI Laboratory project as well as 'Out of the Black Box: Indigenous protocols for AI' by Angie Abdilla, Megan Kelleher, Rich Shaw and Tyson Yunkaporta, as well as the Position Paper developed by the Indigenous Protocol and Artificial Intelligence Working Group.

- For a look at a public data trust framework and data stewardship, the Department may wish to consider the Data Trust Initiative of Cambridge University and research conducted by the Ada Lovelace Institute.

Contact

Samantha Floreani | Program Lead | samantha@digitalrightswatch.org.au