Submission to the Select Committee on Adopting Artificial Intelligence (AI)

regarding the inquiry into

The opportunities and impacts for Australia arising out of the uptake of AI technologies in Australia

17 May 2024

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 Select Committee on Adopting Artificial Intelligence (AI) as part of the inquiry into the opportunities and impacts for Australia arising out of the uptake of AI technologies in Australia.

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. As such, our focus in this submission is on the potential and existing harmful outcomes of further adopting AI in Australia, and the need to mitigate such consequences. In our view, the key way to do this is to establish and maintain focus on human rights.

Digital Rights Watch actively participates 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 Department of Industry, Science and Resources in response to the Safe and Responsible AI issues paper[2]
- 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[3]
- Submission to the Senate Economics Committee Inquiry into the influence of international digital platforms[4]

Given the broad scope of the terms of reference of this inquiry, we have endeavoured to provide high-level, thematic general comments and suggestions for further reading and inquiry.

Digital Rights Watch welcomes the opportunity to participate in public hearings, consultations and to provide comment and feedback on future specific proposals.

---

[2] Digital Rights Watch Submission to the Department of Industry, Science and Resources in response to the Safe and Responsible AI issues paper, 14 August 2023, Available at: https://digitalrightswatch.org.au/2023/08/14/safe-responsible-ai/

[3] 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/

[4] 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/

A. Terminology and hype

"AI" can be a slippery concept that has different meanings and purposes depending on who is using it and why. More often than not, "AI" is a marketing term. Given that the terms of reference do not list a working definition or scope of what the committee considers to be "AI", our first suggestion is that the committee establish what it is they are specifically seeking to consider. While "generative AI" systems such as large language models (LLMs) are currently attracting immense attention, it would be wise to consider whether other forms of AI are also within scope, such as machine and deep learning systems, in particular those that fuel automated decision making systems, computer vision, facial recognition and other biometric surveillance and analysis technologies, and so on. 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.

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 committee 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 committee not to become distracted by longtermist hype, fearmongering and speculative fiction.[5] 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

[5] 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/

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*.

B. Placing human rights at the centre

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.

These technologies 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 facial recognition used by law enforcement resulting in wrongful arrest,[6]
- **collective harm** affecting entire groups, for example as a result of racial profiling through predictive policing,[7]
- **harms of allocation** arising from discriminatory allotment of, or unequal access to, products or resources,[8]
- **harms of representation** which reinforce existing discrimination, disadvantage or stigma, often by using historical datasets which contain biased, incomplete or outdated data.[9]

---

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

[7] 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

[8] 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

[9] 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

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.[10]

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.

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.[11] For example, the European Union's 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.[12]

---

embeddings,' Microsoft, 2016. https://www.microsoft.com/en-us/research/publication/quantifying-reducing-stereotypes-word-embeddings/

[10] '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

[11] 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 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/

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

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

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

C. 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. This is also pertinent because the data used to train AI models is not in itself artificial, it is material in that it is derived from and about real people. Regulation that requires high quality data governance practices is essential if Australia is to foster a responsible AI industry in Australia, and encourage the adoption of AI that will benefit citizens.

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.[13] 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. This results in a serious lack of representation from affected communities,

[13] 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

as well as a missed opportunity for socially-responsible technology development that meets the genuine needs and local priorities of communities.[14]

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.[15] 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, there are many such uses that do raise privacy concerns. For instance, uses of AI and ADM that 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 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.[16] 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

[14] See, for example, the work of Connected by Data: https://connectedbydata.org/
[15] '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
[16] '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 Committee 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.[17] 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).[18] Formalising this is one of many proposed amendments to strengthen the Privacy Act.[19] 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.[20] For example, the use of post codes in machine learning systems has been shown to act as a proxy for race.[21] 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 important improvements that will have flow on effects with regard to how AI is able to be legally developed and deployed.

---

[17] 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

[18] '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

[19] Specifically, by updating the definition of 'personal information', proposal 4.3 in the privacy act review report.

[20] '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

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

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.[22] 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.[23]

Recommendations

2. Comprehensively reform the *Privacy Act 1988*.
3. 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

D. Key risks and harms arising from the adoption of AI technologies

Below we provide a very high level overview of some of the key areas of harm related to AI, with examples and suggestions for further reading.

[22] '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
[23] 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

Facial recognition presents specific forms of harm, especially in contexts where it is perceived as highly effective. This is because biometric data is inherent (unlike a password, it cannot be changed in the event the information falls into the wrong hands). It is also because of the immense scope for surveillance presented particularly by real time facial recognition. Further, there is scope for bias (a failure to recognise faces from different racial backgrounds) as well as discrimination, particularly in law enforcement settings (invasive use of the technology on already over policed communities). There are also serious questions about its accuracy, with accounts from the UK of only one in seven (15%) of live matches used by the police being correct.[24] An illuminating example comes from Georgetown Law School, which documented the use of facial recognition to identify a suspect of a crime. The NYPD officers obtained no useful match when an image of the offender was submitted to the system, and then proceeded to use a picture of Woody Harrelson because they thought he looked like the offender. The officers arrested a person who presented as a match.[25] This is a highly alarming use of a dangerous technology.

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. We note with alarm that while the OAIC in fact made findings against ClearView AI in respect of the data provenance of their facial recognition algorithm, these do not appear to have been complied with.[26] This suggests a recklessness within the industry which should not be tolerated, and should be the basis for a prohibition on the use of this product in Australia.

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."[27]

---

[24] This is contested by the police, but with an absence of detailed data, it is hard to accept their claims. See Big Brother Watch, 'Understanding Live Facial Recognition Statistic,' May 2023 https://bigbrotherwatch.org.uk/blog/understanding-live-facial-recognition-statistics/

[25] See 'Garbage in, garbage out, face recognition on flawed data' Georgetown Law School https://www.flawedfacedata.com/ May 2019.

[26] 'Clearview AI breached Australians' privacy,' OAIC, 3 Novemebr 2021 https://www.oaic.gov.au/newsroom/clearview-ai-breached-australians-privacy

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

The Committee 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.[28]

Online safety

The use of AI in relation to online safety is complex and currently lacks clear guidelines and regulation in Australia. AI and Machine Learning are often touted as potential solutions to or remedies for online abuse. During the Women's World Cup hosted in Australia in 2023, FIFA engaged 'Threat Matrix' to monitor the social media of players, coaches and staff for online abuse and harassment using AI.[29] However, the effectiveness of these automated 'detect and delete' systems has been called into question[30] as they are often opaque and lack a nuanced understanding of communicative norms. These approaches also only monitor publically available comments and posts, meaning that direct messages and emails can only be monitored by allowing a third party to seriously invade the privacy of a particular user's social media and online accounts. Without robust privacy protections, such invasions of privacy can then lead to issues around data provenance raised earlier in this submission. We encourage the committee to consider the flow-on effects of a lack of proper regulation of AI, even when the potential benefits seem obvious.

Relatedly, and as discussed in our submission to the Online Safety Draft Industry Standards, "[h]igh quality, automated detection of [Child Sexual Abuse Material and "pro terror" content] is extremely difficult, especially when the scope of target content is broad, not strictly defined, or difficult to assess without additional context." The AI and Machine Learning software currently used to automatically scan and detect content is often over- and under-effective, making it unreliable and risky for a range of human rights abuses.[31]

At the same time, we are increasingly concerned with the rising use of generative AI to create "deepfakes". Deepfakes are images, videos and audio files that look and sound real but have been either manipulated or fabricated using AI. Recent high-profile examples have included Joe Biden and Taylor Swift, but the risks for

[28] 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
[29] See https://inside.fifa.com/social-impact/campaigns/no-discrimination/fifa-social-media-protection-service/fifa-womens-world-cup-australia-new-zealand-2023
[30] See https://journals.sagepub.com/doi/10.1177/2053951719897945
[31] For further information, please see our submission to the Online Safety Draft Industry Standards https://digitalrightswatch.org.au/2024/01/08/submission-online-safety-standards/

everyday Australians remains high when deepfake pornography is an increasing social problem in Australia.[32] Further, deepfakes pose a risk to democracy and democratic processes in how images and audio can be manipulated and shared through easily-accessible online tools.[33] Our recommendation is that the Committee take seriously the threats posed by deepfakes while also keeping in mind Australians' rights to privacy and the limitations of legislation and technology in addressing the risks and harms of deepfakes.

The Committee may like to look to the EU's AI Act for an international example of how AI is regulated. The EU Act outlines a range of legislative and regulatory measures for different kinds of AI systems.[34] The Act has not come about without criticism (discussed further below). The responses to address online safety and AI must be robust and long-term, tackling the social and cultural dimensions of inequity that leads to harmful, dangerous and inappropriate conduct using AI, and not simply relying on legal frameworks to address long-standing social issues.

AI in the workplace; automation of work

A key risk of adopting AI technology is the potential to disenfranchise workers significantly, especially those in lower paid jobs. Lauren Kelly, a scholar examining workplace automation, has highlighted that "*Across digital labour platforms as well as conventional employment settings…automation often functions as a tool of work intensification, not elimination*."[35] Research is beginning to reveal that many advancements in automation do not reduce or supplement work, but rather augment the role of management.[36] They also often result in the increase of pervasive surveillance and monitoring of workers under the guise of productivity.

Workers are at serious risk of their rights being encoded away into decision making systems that fail to account for their personal circumstances and respect their individual dignity. Advances in the adoption of AI, both as used by workers

---

[32] See https://pursuit.unimelb.edu.au/articles/picture-to-burn-the-law-probably-won-t-protect-taylor-or-other-women-from-deepfakes

[33] See https://www.unswlawjournal.unsw.edu.au/article/disinformation-deepfakes-and-democracies-the-need-for-legislative-reform

[34] See https://artificialintelligenceact.eu/

Kelly, Lauren Kate. "Automation is changing work—not erasing it". *State of Digital Rights Report: A 2021 Retrospective*. Melbourne: Digital Rights Watch.

See, for example: Mateescu, Alexandra and Nguyen, Aiha. "Algorithmic Management in the Workplace". *Data & Society*. 6 February 2019, 1-15; Moore, Phoebe, Upchurch, Martin and Whittaker, Xanthe. *Humans and Machines at Work: Monitoring, Surveillance and Automation in Contemporary Capitalism*. London: Palgrave Macmillan, 2018; Banks, David. "Automatic for the Bosses: Workers may be more affected by robots taking their bosses' jobs than their own." *Real Life*. July 9, 2020; and Kelly, Lauren Kate. *Technology and power: Understanding issues of insecure work and technological change in Australian workplaces*. Melbourne: United Workers Union, 2020.

and used to manage them, must be accompanied by a strengthening of their rights. This could include, for example, improved rights of review of automated decisions that affect workplace conditions, consultation and participation in system design.

E. Bias, discrimination and error, and the importance of accountability

Bias, discrimination and error are significant potential harms that arise in the context of AI. There is a view that such harms are already provided for in regulation, via discrimination, tort law and consumer protections, among others. We understand this perspective, and while it is accurate it is also incomplete. Regulations must be enforceable and enforced to be effective. Relying on individuals to be able to bring claims in circumstances where the harm is unknowable to them, or difficult to identify and attribute to a single company or government agency is unrealistic. A lack of enforceability fosters an industry that is irresponsible, where compliance is considered at best costly and inconvenient and at worst unnecessary, rather than a baseline social licence that all companies must create and maintain.

Virginia Eubanks, in her book *Automating Inequality* sets out numerous significant examples of this. The use of algorithmic technology to assess risk for the purposes of child removal is one such example, which may have discriminatory outcomes at a population level but are nonetheless virtually impossible to litigate as an individual. Eubanks documents how in a range of settings, data points are accumulated by individuals who are in contact with state agencies and these are then used against them to create a 'digital poorhouse'. The purpose of this is to manage poverty, not alleviate it. Middle class families, in contrast, have fewer reasons to be in contact with the state and can escape such surveillance and algorithmic management.

Robodebt is the obvious example of this in Australia, with the well-documented harms generated by this costly system. Such programs of digital transformation should never occur again, and this should be a key objective of regulation of AI.

Our suggestion is that the committee consider recommending the creation of a bespoke regulator with powers to obtain information and supervise the use of AI systems as required. Such a regulator should also have the capacity to impose fines and remedies, and prohibitions on the use of AI systems where they do not meet safety standards (Clearview AI being one such example). We also suggest that the committee consider whether powers ought to be available to order the retraining of algorithmic technology where there have been identified problems with data provenance, and more general, new powers to restrain the use of AI which has given rise to documented and directly unfair outcomes.

We also believe that individuals should have the right to redress, which might include compensation for harm, but also more generally, standard sums or penalties for unfair treatment. This is especially important where it may be very difficult and costly for an individual to identify harm that is causally linked to the AI. We think the committee could consider, for example, a new coursese of action arising from unfair treatment by an AI system, which seeks to remedy the indignity of being subjected to these systems, even if harm is challenging to quantify.

Recommendations

4. Create a new, bespoke regulator with powers to obtain information, supervise AI systems, require algorithmic retraining and restrain the use of applications in certain settings.
5. Introduce a right to redress for individuals, as well as a prohibition on unfair treatment by an AI system.

F. Mixed results: the approach to mitigating AI risks in the EU

The European Union has taken a risk based approach to regulating AI. There are benefits to this approach, as it serves to effectively ban the use and sale of very high risk AI applications. We think there is utility in considering outright bans on certain kinds of AI tools and welcome this framing in this respect.

However, there are also limitations. Generative (foundational) models can be used in very harmful ways but are not classified as high risk because of the many low risk uses of these systems. The approach eventually adopted by the EU also provided for self-regulation via giving industry the capacity to identify the risk level of the application, which creates problematic incentives. We have also been highly concerned by the wholesale exclusion of the use of AI in national security settings. In our view, a better model is to regulate through the prism of risk, as well as human rights.[37]

The public has a right to know about how AI is being developed and used in all settings, and have a discussion about its safety and the appropriate limits that

[37] We concur with our European civil society colleagues on these matters, see https://protectnotsurveil.eu/

ought to be imposed by regulation. While the EU regulation has some significant benefits, we should also ensure we do not replicate its mistakes.

Recommendations

6. Consider the list of high risk technologies in the EU regulation as a non-exhaustive list of applications that should be banned or highly restricted in their use.
7. Adopt a regulatory model that is built on Human Rights principles that is led by Government and civil society, with minor input from the AI industry.
8. Do not adopt self-regulatory models that rely on industry to identify the level of risk (and therefore regulation) that applies to their AI applications.
9. Do not adopt wholesale exclusions of regulation for national security uses.

G. 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 committee 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 exploration of bias and discrimination in AI and automated systems we suggest *Automating Inequality* by Virginia Eubanks, *Race after Technology* by Ruha Benjamin
- Kelly, Lauren Kate. *Technology and power: Understanding issues of insecure work and technological change in Australian workplaces*. Melbourne: United Workers Union, 2020.
- 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 Committee may wish to consider the Data Trust Initiative of Cambridge University and research conducted by the Ada Lovelace Institute.

Contact

Samantha Floreani | Head of Policy | [redacted]