Electronic Frontiers Australia Inc.
ABN: 35 050 159 188
W www.efa.org.au
E email@efa.org.au
@efa_oz

Supporting responsible AI
Department of Industry, Science and Resources: Technology Strategy Branch

13 August 2023

By web form

To the Department,

RE: Safe and Responsible AI

EFA welcomes the opportunity to comment on the Safe and Responsible AI consultation.

EFA's submission is contained in the following pages.

About EFA

Established in January 1994, EFA is a national, membership-based, not-for-profit organisation that promotes and protects human rights in a digital context.

EFA is independent of government and commerce, and is funded by membership subscriptions and donations from individuals and organisations with an altruistic interest in promoting civil liberties in the digital context.

EFA members and supporters come from all parts of Australia and from diverse backgrounds. Our major objectives are to protect and promote the civil liberties of users of digital communications systems (such as the Internet) and of those affected by their use and to educate the community at large about the social, political, and civil liberties issues involved in the use of digital communications systems.

Yours sincerely,

Justin Warren
Chair
Electronic Frontiers Australia

Kathryn Gledhill-Tucker
Vice-Chair
Electronic Frontiers Australia

Introduction

As we have consistently asserted in the past in response to other consultations, EFA considers that the most important aspect of responsible or ethical AI regulation is the introduction of a Federally enforceable human rights framework[1].

EFA suggests that there is likely no need for technology-specific legislation. Rather, there already exists a wealth of principles-based regulation of behaviour and harm that needs to be properly enforced. In addition to existing legislation, much of the proposed legislation in the Privacy Act Review would provide a strong foundation to protect the rights of individuals.

Summary of Recommendations

1. Enforce existing technology-neutral, principles-based legislation rather than rushing to create new, technology-specific legislation.
2. The Privacy Act should be amended to provide strong privacy protections for individuals and groups.
3. Individual and collective rights of action should be adopted as part of a graduated model of regulation that devolves and distributes power more widely.
4. The Federal government should coordinate with the various states and territories to provide a uniform and harmonised regulatory framework.
5. Private organisations that act for the government should be subject to all of the same regulations that bind the government.
6. The government should be required to compensate individuals and groups for redress of harms caused by its failure to implement automated systems safely.
7. Individuals harmed by government systems should be entitled to exemplary damages to incentivise the government to live up to its obligations.
8. Any risk-based framework must include a category of “unacceptable risk” that prohibits certain applications or practices.
9. Responsible AI must be mandated through regulation rather than voluntary principles.

[1] Electronic Frontiers Australia, ‘Submission on AI Ethical Framework Consultation’ <https://www.efa.org.au/wp-content/uploads/2019/06/310519_EFA-AI-Ethical-Framework-Submission.pdf>.

Outline of Submission

Definitions

1. Do you agree with the definitions in this discussion paper? If not, what definitions do you prefer and why?

We recognise that defining Artificial Intelligence (AI) is a contentious activity. There is an active and ongoing discussion where experts do not all agree on a single definition.

The EU AI Act proposes a definition of "artificial intelligence system":

> "a machine-based system that is designed to operate with varying levels of autonomy and that can, for explicit or implicit objectives, generate outputs such as predictions, recommendations, or decisions, that influence physical or virtual environments;"[2]

Researcher Kate Crawford asserts, "AI is neither artificial nor intelligent. It is made from natural resources and it is people who are performing the tasks to make the systems appear autonomous."[3] As we are defining AI and related technologies, it is important not to abstract these tools away from the human individuals responsible for their design and deployment. The responsibility of AI and consequences of their output should always sit with a person, not a tool.

EFA considers that any engineered system reflects the intentions and biases of its designers. AI systems are "an organised connection of elements operating in order to produce [a] conduct or outcome."[4] As Elise Bant explains:

> "[C]orporations manifest their intentions through the systems of conduct that they adopt and operate, both in the sense that any system reveals the corporate intention and in the sense that it embodies or instantiates that intention."[5]

The use of AI, however defined, is therefore nothing more than a method for reducing the cost or effort required for humans to achieve an outcome that they would otherwise achieve in more laborious ways. AI technology is an "adopted system of

[2] European Parliament, 'Texts Adopted - Artificial Intelligence Act' (14 June 2023) <https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html>.

[3] Zoë Corbyn, 'Microsoft's Kate Crawford: "AI Is Neither Artificial nor Intelligent"', *The Observer* (online, 6 June 2021) <https://www.theguardian.com/technology/2021/jun/06/microsofts-kate-crawford-ai-is-neither-artificial-nor-intelligent>.

[4] Elise Bant, 'Catching the Corporate Conscience: A New Model of "Systems Intentionality"' [2020] (Part 3) *Lloyd's Maritime and Commercial Law Quarterly* 467, 472.

[5] Ibid.

conduct” that can be used to objectively characterise the associated intention of that system. To paraphrase Stafford Beer, “the purpose of a system is what it does”.

EFA considers that the regulation of behaviour and outcomes should not rest on any one definition of AI. A broader approach that rests on certain fundamental principles, rooted in a human rights framework, is less likely to be bypassed or become immediately obsolete as technology changes occur.

We consider the Robodebt scheme, which the Royal Commission branded an “extraordinary saga” of “venality, incompetence and cowardice”[6] should be taken as a cautionary tale. No *artificial* intelligence was involved, yet the manifest harms of the organised connection of elements that operated in order to produce the conduct and outcome of Robodebt were both technically unlawful and obviously harmful. Any regulation must surely capture the conduct and outcomes observed in the Robodebt saga to be of any value.

Potential gaps in approaches

2. What potential risks from AI are not covered by Australia’s existing regulatory approaches? Do you have suggestions for possible regulatory action to mitigate these risks?

All new legislation relating to technology in this country must be built on top of a Federally enforceable human rights framework. We challenge the suggestion that, “There are strong foundations for Australia to be a leader in responsible AI.” While our Privacy Act is still under review, with many fundamental improvements still to reach legislation, and while we do not have a federally enforceable Right to Privacy, the foundation on which new technological advancement is built in this country is fundamentally shaky.

Existing laws, such as Australian Consumer Law and the Privacy Act, provide some scaffolding to regulate technological advancements in AI. However, there is a lack of individual rights (such as a tort of serious breach of privacy) and an over-reliance on regulators. Regulators can fail to act, and are also not sufficiently resourced to chase every violation. A graduated model that allows for individual and collective action or redress frees up regulators to focus on systemic or egregious situations.

Recommendation: The Privacy Act should be amended to provide strong privacy protections for individuals and groups.

[6] Alexander Lewis and Ciara Jones, ‘Commissioner Brands Robodebt “Extraordinary Saga” of “Venality, Incompetence and Cowardice”’, *ABC News* (online, 7 July 2023) <https://www.abc.net.au/news/2023-07-07/robodebt-royal-commission-findings-revealed/102531450>.

Recommendation: Individual and collective rights of action should be adopted as part of a graduated model of regulation that devolves and distributes power more widely.

3. Are there any further non-regulatory initiatives the Australian Government could implement to support responsible AI practices in Australia? Please describe these and their benefits or impacts.
4. Do you have suggestions on coordination of AI governance across government? Please outline the goals that any coordination mechanisms could achieve and how they could influence the development and uptake of AI in Australia.

In Australia's federated system of government, a great deal of Australians' lives are governed not by Commonwealth legislation but by regulation at the state, territory, or local level. The federal government should seek to harmonise any legislation so that it functions similarly for all Australians.

A lack of harmonised legislation has been recognised as placing undue compliance burden on individuals and organisations where they operate in more than one jurisdiction. There is also a risk that unscrupulous operators could "jurisdiction shop" and seek to use more favourable laws in an alternate jurisdiction.

Recommendation: The Federal government should coordinate with the various states and territories to provide a uniform and harmonised regulatory framework.

Responses suitable for Australia

5. Are there any governance measures being taken or considered by other countries (including any not discussed in this paper) that are relevant, adaptable and desirable for Australia?

Consistency with international governance measures provides interoperability for Australian organisations and citizens. Organisations frequently have a global presence and are already required to adhere to international standards. EFA recommends bringing legislation in line with international legislation such as the EU's General Data Protection Regulation (GDPR) and AI Act.

While benchmarking against international measures provides us with a high watermark for legislative standards, all governance measures ought to be principles-based and grounded in a human rights framework.

Target areas

6. Should different approaches apply to public and private sector use of AI technologies? If so, how should the approaches differ?

When it comes to regulating technology, there is not always a clear distinction between the public and private sector. Governments frequently compel private organisations to collect information for law enforcement purposes that those organisations would otherwise not collect, and which it would be impractical for the government to collect by itself.[7] Governments also purchase surveillance data that it would otherwise be impractical or unlawful to collect directly.[8] In many cases, private sector organisations act — sometimes voluntarily, sometimes under duress — as extensions of the state. Modern governments are tightly enmeshed with private corporations that act on their behalf.[9] Many government services are outsourced to private providers that act, for all intents and purposes, as if they *are* the government.

Governments are in a unique position, however, given their monopoly on the use of force, to deprive individuals of their liberty and possessions. Governments have extraordinary power over others, and must therefore live up to a higher standard of behaviour, and be subject to extraordinary scrutiny. What may be permissible in the private sector is frequently not permissible by governments, and for good reasons.

Private organisations that act for the government should be subject to the same regulations as the government itself. This should be true whether that conduct is voluntary or when governments compel private organisations to act as extensions of the government. Governments should not be able to sidestep constraints on their conduct by hiring mercenaries to do their dirty work for them.

Recommendation: Private organisations that act for the government should be subject to all of the same regulations that bind the government.

7. How can the Australian Government further support responsible AI practices in its own agencies?

[7] See e.g. AG, 'Telecommunications and Other Legislation Amendment (Assistance and Access) Act 2018' <https://www.legislation.gov.au/Details/C2018A00148/Html/Text, http://www.legislation.gov.au/Details/C2018A00148>.
[8] Joseph Cox, 'How the U.S. Military Buys Location Data from Ordinary Apps', *Motherboard* (17 November 2020) <https://www.vice.com/en/article/jgqm5x/us-military-location-data-xmode-locate-x>.
[9] Though not always, see e.g. Henry Belot, 'Deloitte Admits Misuse of Government Information as Scandal Engulfing PwC Widens', *The Guardian* (online, 14 July 2023) <https://www.theguardian.com/business/2023/jul/14/deloitte-misuse-of-government-information-senate-inquiry-pwc-scandal>; Henry Belot, 'PwC Admits to Another Conflict of Interest Breach', *The Guardian* (online, 12 July 2023) <https://www.theguardian.com/business/2023/jul/12/pwc-scandal-second-conflict-of-interest-breach-government-information>.

Australia has a dark history of implementing automated decision making systems with resultant real world harm to vulnerable individuals. The recent report from the Royal Commission into Robodebt detailed the serious harms possible when the pursuit of cost cutting is prioritised over the welfare of human beings.

Too often, the government makes "risk weighted" decisions where the upside risks accrue to itself or a small collection of private interests while the downside risks are borne by ordinary Australians. What the government believes is in its own interests does not always align with the public interest. Outsized, fanciful benefits in the far future are used to justify taking risks that are not borne by those taking them. Risky bets are made with other people's money and when the frequently predictable, and predicted, downside risks manifest themselves, no redress for harm is offered.

The government should lead by example, holding itself to the highest standard of ethical conduct and ensuring that the consequences for its own failures are both swift and proportional to the harm it has caused. Those harmed by government failures should be entitled to both redress for that harm and also exemplary damages, representing the severity of the government's failure to live up to its obligations to the society that has granted it such extraordinary power over our lives. They should not be required to mount costly and time consuming legal challenges in order to compel their own government to act honestly and ethically.[10]

Recommendation: The government should be required to compensate individuals and groups for redress of harms caused by its failure to implement automated systems safely.

Recommendation: Individuals harmed by government systems should be entitled to exemplary damages to incentivise the government to live up to its obligations.

8. In what circumstances are generic solutions to the risks of AI most valuable? And in what circumstances are technology-specific solutions better? Please provide some examples.

The majority of the risks of AI are the same as any other human decision-making system. The major difference is the speed, scope, and scale at which automated systems can act. It is therefore less a technology-specific question than a speed, scope, and scale question. The same base principles regulating conduct and outcome can be applied, modified by a general principle of proportionality: if conduct at small scale is bad, then the same conduct at large scale is worse.

[10] Darren O'Donovan, 'Let's Be Clear. Robodebt Was Ended by Welfare Recipients with Their Suffering', *The Guardian* (online, 8 July 2023) <https://www.theguardian.com/commentisfree/2023/jul/09/lets-be-clear-robodebt-was-ended-by-welfare-recipients-with-their-suffering>.

However, there comes a point at which the speed, scope, or scale of harm becomes *qualitatively* different. This should be the focus, as the principle of a qualitative step change holds true across a number of domains. For example, going slightly over the speed limit might be acceptable when overtaking another vehicle on the road. Exceeding the speed limit significantly becomes an offence, and justifying it becomes more challenging; consistently driving at speeds that potentially endanger others is considered more serious again. The regulations addressing speeding are rightly concerned less with the particular technology employed, than with the potential risks and outcomes involved.

9. Given the importance of transparency across the AI lifecycle, please share your thoughts on:
    a. where and when transparency will be most critical and valuable to mitigate potential AI risks and to improve public trust and confidence in AI?
    b. mandating transparency requirements across the private and public sectors, including how these requirements could be implemented.

AI and associated algorithms can often be opaque, referred to as a "black box", leading to an inability to determine how the machine has come to a conclusion. This is of particular concern when the output of an algorithm impacts an individual; if there is a flaw or bias in the system that harms a person, it can be difficult to trace back the source of that harm if the system itself cannot be easily understood.

The inscrutability of the algorithm should be taken into account when assessing the risk of harm. The more opaque, the lower the threshold for acceptable risk of any AI.

Taking guidance from the ACM U.S. Public Policy Council & ACM Europe Policy Committee and their Principles for Algorithmic Transparency and Accountability, we recommend establishing some guidelines around "data provenance":

> "A description of the way in which the training data was collected should be maintained by the builders of the algorithms, accompanied by an exploration of the potential biases induced by the human or algorithmic data-gathering process."[11]

These recommendations may seem onerous to organisations that are used to deploying analytics and artificial intelligence with wild abandon, but handling the data of individuals and developing algorithms that impact human beings is a highly sensitive exercise and must be treated with appropriate caution. No amount of

[11] ACM U.S. Public Policy Council & ACM Europe Policy Committee, 'Joint Statement on Algorithmic Transparency and Accountability' (2017) <https://www.acm.org/binaries/content/assets/public-policy/2017_joint_statement_algorithms.pdf>

profit-seeking should supersede the need to protect people from algorithmically-driven harm or exploitation.

10. Do you have suggestions for:
    a. Whether any high-risk AI applications or technologies should be banned completely?
    b. Criteria or requirements to identify AI applications or technologies that should be banned, and in which contexts?

The EU AI Act provides sound recommendations for AI applications or technologies that should be placed in a category of "unacceptable risk". These practices are considered to be such a clear threat to people's safety, livelihood, and rights that their use should be prohibited. These practices include:

> "AI systems that deploy harmful manipulative 'subliminal techniques';
> AI systems that exploit specific vulnerable groups (physical or mental disability);
> AI systems used by public authorities, or on their behalf, for social scoring purposes;
> 'Real-time' remote biometric identification systems in publicly accessible spaces for law enforcement purposes, except in a limited number of cases.[12]"

We note at the time of this consultation there is a severe lack of legislation protecting individuals from the harms of biometric surveillance. Without adequate regulation, there is a risk of creating a culture of normalising surveillance, and going past a point of no return when deploying technology that is capable of capturing sensitive and immutable details of an individual. In the event of a data breach, an individual cannot change their face.

Australia's lack of a fundamental Bill of Rights creates challenges for determining if any practices should be banned in Australian society. Unlike the EU, Australia has not yet wrestled with the thorny problem of defining the fundamental principles on which its liberal democracy should be based. We decry certain actions of foreign governments that are viewed as authoritarian or anti-democratic, and yet when those same actions are performed by Australian governments, the conduct is somehow rendered acceptable. The rule of law requires that everyone should be subject to the same standards of behaviour; "it's okay when we do it" should not be the basis for our regulatory frameworks.

Some conduct should be prohibited because it violates the fundamental principles on which our nation is based. Australia's challenge is that we are unable to articulate those fundamental principles with any degree of coherence.

[12] https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf

Notably, the risk management framework proposed in the Safe and Responsible AI in Australia report has no “unacceptable risk” rating. There are some practices that pose such a serious threat that their use should be prohibited. We support the introduction of prohibited practices such as those categorised as “unacceptable risk” in the EU legislation.

As noted in the report, algorithmic bias is a legitimate concern. AI will replicate the bias of a system on which it is trained. Unwanted bias, such as racial discrimination, cannot be remedied by the introduction of more data if the system (i.e. the underlying dataset) itself is biassed. For example, Aboriginal and Torres Strait Islander peoples make up 3% of the general population of Australia, but closer to one third of the imprisoned population[13]. Any AI applied to the system of incarceration or law enforcement risks further entrenching this overrepresentation of Aboriginal and Torres Strait Islander peoples in our prison system. We therefore suggest that the existence of extreme bias in a system be taken into account when assessing the risk of AI technologies.

Recommendation: Any risk-based framework must include a category of “unacceptable risk” that prohibits certain applications or practices.

11. What initiatives or government action can increase public trust in AI deployment to encourage more people to use AI?

The framing of this question presupposes that more AI is inherently a good thing. The case has not yet been made. Not all innovation is useful. Not all change is worthwhile. The onus is on those who wish to deploy new technologies to demonstrate their value and safety, and to accept the consequences if they are wrong. The reckless deployment of unproven technologies onto the public at large should be met with scepticism. Breathless techno-utopian claims should not be accepted at face value by any government that claims to value evidence-based decision making.

To assist with determining if there is real value to a technology, and that its value outweighs any costs to individuals and society collectively, the government could encourage small-scale trials under tightly controlled conditions. This will minimise the risks to Australians while helping to establish a robust evidence base that would justify further support. Such trials should require the publication of detailed findings, both positive and negative. Australians would then be able to better inform themselves of the value of technologies such as AI, and either encourage or discourage further use of public funds to support their development.

[13] *Prisoners in Australia, 2022 | Australian Bureau of Statistics*. (2023, October 5). https://www.abs.gov.au/statistics/people/crime-and-justice/prisoners-australia/latest-release

Direct support is not the only action the government could take. A robust regulatory environment that supports a just and equitable society would encourage the development of technologies that further improve Australians' quality of life. Active steps to redistribute power and wealth would make Australia a more equal society, less prone to abuses of power by self-interested cliques.

Technological systems are socially constructed. Government choices about which technologies should be used and which will not shape the environment in which technologies develop. The choices the government makes should be based on fundamental principles about the kind of society Australia wants to be. All of its choices reflect those principles. When it chooses to favour the interests of multinational corporations and business groups over those of ordinary citizens, it is telegraphing the kind of society it thinks Australia should be.

If the government wants the public to trust it, it must first demonstrate that it is trustworthy. Recent evidence suggests it has a great deal of work to do before that will be true.

Implications and infrastructure

12. How would banning high-risk activities (like social scoring or facial recognition technology in certain circumstances) impact Australia's tech sector and our trade and exports with other countries?

This question highlights the shortfalling of the proposed risk assessment model in this report. Activities such as social scoring or facial recognition are more severe than "high-risk"; they ought to fall into a category of "unacceptable risk" as they pose such a threat to individuals that they should be outright banned. There is no amount of guardrails or supervision that can make an activity such as social scoring compatible with a liberal democracy.

This question highlights the extent to which the government is prepared to take a cold and amoral approach when discussing the rights of its citizens. It is akin to asking "would banning the mass imprisonment of politicians impact Australia's tech sector and our trade and exports with other countries?" without flinching. The disappointment of those keen to profit from prison expansions would not be seriously balanced against the desire of politicians to remain at large. Why, then, is the government prepared to contemplate such fundamental alterations to the nature of our society as social scoring as being somehow related to *trade and exports*?

Why not investigate the potential for an over-70's *Logan's Run* regime to save on the aged pension? Perhaps the local tech sector could be given a boost building miniature

GPS trackers to inject into the neck of every public servant? These are obviously ludicrous suggestions, and so is the question posed here.

13. What changes (if any) to Australian conformity infrastructure might be required to support assurance processes to mitigate against potential AI risks?

We have no specific notes on this question.

Risk-based approaches

14. Do you support a risk-based approach for addressing potential AI risks? If not, is there a better approach?

We recognise the EU Artificial Intelligence Act proposes a risk-based approach to AI legislation that is technology-neutral. A risk-based approach is challenging to implement when there is a lack of historical data from past incidents to inform risk assessments. Guesses are not evidence. The simple novelty of AI technology renders any risk-based approach fundamentally flawed as there is no basis — beyond mere speculation — on which to base a risk assessment. “She’ll be right” should not form the basis of government regulation.

Automation of a well-known process with a lengthy history of evidence supporting known-good practices is less likely to go wrong in unexpected ways. Automation of a new process with no history or evidentiary base for safety assessments would be reckless. We suggest that the latter form of automation should fail a “due diligence, expertise, and skill” probity test.

15. What do you see as the main benefits or limitations of a risk-based approach? How can any limitations be overcome?
16. Is a risk-based approach better suited to some sectors, AI applications or organisations than others based on organisation size, AI maturity and resources?
17. What elements should be in a risk-based approach for addressing potential AI risks? Do you support the elements presented in Attachment C?
18. How can an AI risk-based approach be incorporated into existing assessment frameworks (like privacy) or risk management processes to streamline and reduce potential duplication?
19. How might a risk-based approach apply to general purpose AI systems, such as large language models (LLMs) or multimodal foundation models (MFMs)?
20. Should a risk-based approach for responsible AI be a voluntary or self-regulation tool or be mandated through regulation? And should it apply to:
    a. public or private organisations or both?
    b. developers or deployers or both?

Voluntary self-regulation grants too much discretion and undue faith in the hands of technology developers and deployers. Frameworks such as the Australian AI Ethics Principles are admirable, but fundamentally ineffective and unenforceable. Organisations in both the public and private sector cannot be trusted to act in the best interests of individuals, especially in a capitalist system that prioritises the pursuit of profit.

This is not to say that a new suite of legislation is required to effectively moderate the development and deployment of AI. Rather, we ought to reflect on the effectiveness of existing legislation, and ensure regulators (such as the OAIC) are sufficiently funded and empowered to enforce such legislation. If current legislation is not effective at protecting individuals from the real world harms that are occurring today, we ought to understand why and remedy these failures as a priority.

Recommendation: responsible AI must be mandated through regulation rather than voluntary principles.