Business, Government and Law
Sturt Road, Bedford Park SA 5042
GPO Box 2100 Adelaide SA 5001
P: +61 8 8201 3300
BGL.Enquiries@flinders.edu.au

26 July 2023

Submission to Supporting responsible AI: discussion paper

Thank you for the opportunity to provide feedback on this paper.

The views expressed in this submission are matters of personal academic opinion and do not purport to represent any institutional position. We have chosen to respond to a selection of the questions put by the paper as set out below.

Potential gaps in approaches

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

There seems to be some inconsistency in the definitional discussion on pages 5 and 6: upfront AI is a "system that generates predictive outputs. or decisions for a given set of human-defined objectives or parameters without explicit programming", but at 6 a: "broad definition of AI is intended that includes any products or services using AI techniques… [t]hese techniques range from simple rules-based algorithms guided by human-defined parameters" (i.e. to machine learning / neural net based approaches). This seems to broaden beyond the original: in that the first excludes explicit programming but the second appears to re-admit it. At this stage of consultation we believe broader is better. Chasing semantic issues can lead to confusion and under or over inclusion, or unproductive boundary disputes at too early a stage.

There is a much more fundamental issue here that the affordances introduced by AI surfaces. Technology per se is not the issue: technology is a tool. Rather the more important issue is the underlying set of policy positions we have adopted in a range of complex interlocking areas. It is important not to allow the debate on technology and definitions obscure this. And if technology itself is not the issue, then it is the wrong target for regulatory reform.[1] To return briefly to the first paragraph, definitional issues would be more significant if technology alone was the issue.

Consider the example of the failures connected to the 737MAX. There may be arguments as to whether that software fits within a given definition of AI or ADM. But does that debate really matter in the context of risk and responsibility and regulatory responses? Arguably not: perhaps it is splitting hairs.
Let us then consider 'soft' automation in the broad: robodebt illustrates this. It was not a sophisticated 'AI' based implementation, but in the public imagination and debate it has been cast within the broader schema of AI, roboadvice etc. Really it was a policy decision implemented through a simple algorithm, deployed at scale and speed through software-based implementation, and developed, protected, and sustained by a policy position, failure of oversight and broader culture that was fundamentally human in origin and with dire human costs as well as the obvious financial and reputational impacts.

[1] See also Lyria Bennett Moses 'How to Think About Law, Regulation and Technology: Problems with 'Technology' as a Regulatory Target' https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2464750.

So perhaps the broader issue that it is hard to avoid one we start down the path of looking at AI and ADM is the issue of risk and responsibility around software in the broad. At this point we start to see the impracticability of special rules or a super regulator to deal with all 'software'. Once we have reached this point though we have not reduced the enquiry about responses to AI and ADM to absurdity. That would be stopping the analysis at too early a stage. Rather we have highlighted a more fundamental issue: the failure of our own cultures, processes, and systems of regulation.[2] These failures are not fundamentally driven by technology. Rather technology is a product of them, and now in the case of AI and ADM a sophisticated product and service that feeds back into them recursively – highlighting their faults. Technology is not just a tool: "technology is not neutral".[3] Rather: it embodies, and is situated within, culture.

There are already mechanisms to try to prevent harms and provide redress when preventative mechanisms fail. Failure of regulatory mechanisms needs to be expected, not treated as an outlier. There need to be regular oversight mechanisms that are learning from and adjusting to such issues. Humans need to be a core part of such mechanisms because we are at the centre of them. Australia has had a long and proud history of law reform bodies (both distinct with that as a driver and those with heavy industry participation, in all fields of endeavour).[4] However we have systematically dismantled many of these (under the guise of cost saving - but perhaps to centralise policy direction and control), or ignored their findings.[5] Such mechanisms do not provide an immediate fix to the issues raised by developments such as AI and ADM as we now find them, but they are part of a more considered systemic response to *all* issues. However, they will not have appropriate impact unless they themselves are well designed, include appropriate stakeholder/collaborators (neither constituting industry or regulatory capture nor entirely excluding it) listened to and acted upon. This is a challenging project that never ends and that has been de-prioritised over the last few decades.[6]

There are many perspectives that could be taken in response to the other questions posed in this consultation, after taking into consideration the definitional and frame level thoughts offered above. Our group has highlighted just a few, centred around the central issues of impact on people: workplaces, management, and data privacy.

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?

---

[2] "The issues that AI governance is often truly about are not technical but deeply normative and distributive—which actors make decisions in society, who bears risks and errors, and what justice should look like procedurally and substantively (Balayn & Gürses 2021)" from Veale, Matus & Gorwa 'AI and Global Governance: Modalities, Rationales, Tensions' Annual Review of Law and Social Science, vol 19, 2023 17.

[3] Melvin Kranzberg, 'Kranzberg's Laws' (July 1986) 27(3) *Technology and Culture* 544.

[4] For example: Australian Human Rights Commission (AHRC), Human Rights and Technology Final Report (2021) <https://humanrights.gov.au/our-work/rights-and-freedoms/publications/human-rights-and-technology-final-report-2021>

[5] This extends to a large array of consultative mechanisms that operated during the 1980s and 1990s. Consider for instance those bodies that previously existed to provide ongoing consultative input to intellectual property reforms, given their relevance to the information economy and AI related issues. We observe also many contributions over decades advocating for recognition of indigenous and cultural intellectual property (ICIP). While noting with interest the Federal Government's recent intent to provide substantive reform in ICIP and commend this, we observe that many other IP reform suggestions remain unaddressed.

[6] See for example the discussion of Emeritus Professor Ricketson's analysis of these issues in Chalmers, 'Burning platforms: Harnessing crisis to reform media regulation' (2022) 24 MALR 178.

Whilst much of the discussion about the risks potentially or provenly associated with AI has focussed on its generic characteristics and uses, there are very specific risk that emerge when AI tools are used in *workplaces*.[7] Conventional technology, from conveyor belts to robots, were designed to work under the guidance of humans. The purpose of (installing) AI, in contrast, is to guide humans, thus reversing the hierarchical relationship between human and machine. This has profound impacts on workplaces and creates new risk and stresses for people working 'in the loop' with automation – let alone for the subjects of decisions produced by fully or partly automated systems.[8]

The surveillance function of AI, intended to guide workplace behaviour through monitoring and sanctioning, has already been well documented. Other everyday workplace risks include but are not exclusive to:

- accelerated work processes, which has ripple effect in workplace areas and functions beyond the one in which AI is deployed (e.g., accelerated production also requires accelerated procurement or sales);
- new physical accident and health risks owing to the spatial integration of AI machine and human, and the use of AI instruments affecting the human body (e.g., AR and VR tools);
- modified supervisory and relational arrangements when human-human reporting lines are replaced or mediated by machine-human interaction with reduced reciprocity where AI decisions cannot or are difficult to overwrite/question;
- challenge to seniority status principles and/or experience-based decision making autonomy affecting the job content and status for individual employees and the balance of employee task profiles, responsibilities and accountabilities across an organisation.

Besides obvious implications for physical health, each of these risks has the potential to impact psychosocial well-being (as well as privacy).

Current workplace health and safety (WHS) regulations have a strong focus on the promotion of physical safety in workplaces. Whilst this remains relevant to AI applications in workplaces, additional psychosocial risks are insufficiently addressed by current regulation. The recent Australian WHS conducted by the NSW Centre for Work Health and Safety[9] showed that whilst 48% of respondents agreed with the statement that "WHS is a priority when new technology is introduced", 26% did not. Australian business are slow adopters of new technology and especially AI. If this changes, however gradually, the penetration of frontier technologies will affect an increasing number and share of employees – and do so much more radically than conventional, human-controlled technology. Regulation must prepare for this future, and Australia is not alone in being currently under prepared.[10]

---

[7] https://www.centreforwhs.nsw.gov.au/research/ethical-use-of-artifical-intelligence-in-the-workplace; Andreas Cebulla, Zygmunt Szpak, Catherine Howell, Genevieve Knight & Sazzad Hussain 'Applying ethics to AI in the workplace: the design of a scorecard for Australian workplace health and safety' <https://doi.org/10.1007/s00146-022-01460-9>

[8] See e.g. some of the very interesting discussion on this issue in "Government use of Artificial Intelligence in New Zealand" (2019) - Final Report on Phase 1 of the New Zealand Law Foundation's Artificial Intelligence and Law in New Zealand Project <https://www.cs.otago.ac.nz/research/ai/AI-Law/NZLF%20report.pdf> (cited in Chalmers, Human Rights and Technology Project Discussion Paper submission (2020)).

[9] NSW, 'Australian WHS Survey' <https://www.centreforwhs.nsw.gov.au/research/national-whs-radar/australian-whs-survey> .

[10] Simon Jack, 'AI: Workers need more protection, says TUC' <https://www.bbc.com/news/business-66248125> .

Many businesses and government departments are already using AI as a screening tool to deal with job applications, so the impact of AI on work can operate also to exclude people from work opportunities at the outset. This means that AI alone may screen an applicant out of a process. We question the appropriateness of this approach: how is it possible for an AI to make an effective evaluation of a candidate's suitability for a complex role based on some form of automated review of isolated features of a video interview or other material? We are aware of one outstanding candidate that was screened out of consideration for a Federal government role recently by exactly such a process: these risks are not theoretical, they are not future, they are here and now and they have been introduced without any public debate, nor necessarily any independent screening of the utility of validity of the tools employed (to say nothing of potential bias of those systems).

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.

The findings from the 2022 Responsible AI Global Executive Study and Research project, reported that "[i]n response to the heightened stakes around AI adoption and impending regulations, organizations worldwide are affirming the need for RAI, but many are falling short when it comes to operationalizing RAI in practice".[11] Accordingly, strong human leadership is required to mentor and monitor responsible design and use of AI so that responsible AI (RAI) frameworks are operationalised effectively in practice. One of the non-regulatory Government initiatives, therefore, is to empower and build human leadership capacity through Responsible AI (RAI) leadership development programs. The focus of the RAI leadership development program should be to equip leaders with skills that would empower them to be actively involved in RAI practices at all levels. Following are the four types of involvement in responsible AI (RAI) practices that the RAI leadership development program can focus on, in no particular order of importance.

Human relations involvement

Human relations involvement signifies that leaders can play the roles of *mentor* and *facilitator* while being involved in operationalising RAI frameworks in practice. Leaders need training in *facilitating* an inclusive work culture and providing a human touch to embed trust and shared meaning regarding ethical AI practices among all stakeholders. Additionally, leaders ought to bridge the functional separation that exists between technical and non-technical experts, listen to ethical AI concerns, and provide empathic *mentoring* to clearly communicate human rights laws and responsibilities related to responsible AI design and use.[12]

Open Systems involvement

The open systems involvement is represented by the roles of *broker* and *innovator*. Being *innovative*, the leaders are required to be proactive visionaries, develop creative foresight and hyperawareness to be able to flexibly scan internal and external environments and identify opportunities and threats to responsible AI. Moreover, multiple perspectives, values and contributions are likely to complicate and challenge the framing of AI problems and responsible AI solutions within organisations. Therefore, leaders ought to be

[11] Elizabeth Renieris, David Kiron, Steven Mills 'To Be a Responsible AI Leader, Focus on Being Responsible' <https://sloanreview.mit.edu/projects/to-be-a-responsible-ai-leader-focus-on-being-responsible/>.
[12] Ben Shneiderman, 'Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy' <https://doi.org/10.1080/10447318.2020.1741118>; Sarah-Louise Richter & Dörte Resch, 'Leadership in the Age of Artificial Intelligence—Exploring Links and Implications in Internationally Operating Insurance Companies' <https://link.springer.com/chapter/10.1007/978-3-030-48332-6_21>.

brokers with good negotiation skills to present and persuade all key stakeholders to commit collectively to RAI practices.[13]

Internal Processes involvement

The ability to *monitor* and *coordinate* ethical and responsible AI practices effectively is one of the key roles of responsible AI leadership. The *Monitoring* function is key to ensuring proper implementation of RAI practices such as impact assessment, auditing trails, bias testing, compliance procedures and accountability traces. Developing close supervision skills would result in risk mitigation from the design to deployment stages of AI. In addition, miscommunication may limit the ability of all stakeholders to support RAI development in a unified manner within organisations. Hence, leaders' *coordination* skills are vital to engage in collaborative relationships and effectively manage multiple teams towards the successful implementation of responsible AI practices.[14]

Rational Goal involvement

The rational goal leadership involvement is focused on productivity and is represented by the roles of *producer* and *director*. The leadership skills development in *direction* entails responsible judgement, goal clarification, goal attainment and the ability to evaluate employees' needs and inspire them towards implementation of responsible AI practices. As a task-oriented *producer* responsible for ethical AI outputs, leaders ought to have the ability to clearly define ethical protocols, create ethical AI roadmap, update policies, define system boundaries and implement correct parameters to evaluate AI outputs.[15]

By promoting and supporting the RAI leadership development program as one of the key initiatives across organisations nationwide, the Australian Government can ensure the operationalization of AI regulatory frameworks in practice at all levels.

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?

We commend the recommendations of the AHRC in relation to the importance of the role of education in enabling our society to understand and respond to the use of AI and ADM.[16]

---

[13] Bogdana Rakova, Jingying Yang, Henriette Cramer, Rumman Chowdhury 'Where Responsible AI meets Reality: Practitioner Perspectives on Enablers for Shifting Organizational Practices' <https://doi.org/10.1145/3449081>; Mathieu d'Aquin, Pinelopi Troullinou, Noel E. O'Connor, Aindrias Cullen, Gráinne Faller, Louise Holden 'Towards an "Ethics by Design" Methodology for AI Research Projects' <https://doi.org/10.1145/3278721.3278765>.

[14] Keng Siau, Weiyu Wang 'Artificial Intelligence (AI) Ethics: Ethics of AI and Ethical AI' https://www.igi-global.com/article/artificial-intelligence-ai-ethics/249172; Daniel Schiff, Bogdana Rakova, Aladdin Ayesh, Anat Fanti, Michael Lennon, 'Principles to Practices for Responsible AI: Closing the Gap' <https://doi.org/10.48550/arXiv.2006.04707>.

[15] Keng Siau, Weiyu Wang 'Artificial Intelligence (AI) Ethics: Ethics of AI and Ethical AI' https://www.igi-global.com/article/artificial-intelligence-ai-ethics/249172; Kolbjørnsrud, Vegard; Amico, Richard & Thomas, Robert J. 'How AI will redefine management' <https://enterprisersproject.com/sites/default/files/how_artificial_intelligence_will_redefine_management.pdf>.

[16] AHRC, *Human Rights and Technology Final Report* (2021) <https://humanrights.gov.au/our-work/rights-and-freedoms/publications/human-rights-and-technology-final-report-2021>.

More specifically in relation to some of the issues discussed above in relation to workers' rights, consider the royal decree-law that updated Spain's Ley del Estatuto de los Trabajadores (Worker's Statute Law) in 2021 to include a provision requiring companies to inform employees of the parameters, rules and instructions of algorithms or artificial intelligence as they affect decision-making, working conditions, access to and maintenance of employment, including profiling.[17]

Target areas

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?

As a matter of principle, there is no point at which transparency ought not to be an option. Transparency may not equate to understanding but is essential for enabling critical review and reflection by users as well as producers of AI. In a workplace context, the national and international evidence is clear: open consultation, information, debate, and discussion across an organisation are key in enabling the safe introduction of AI technology. They also facilitate the collective monitoring of AI impacts on workplaces over time. This is important as those impacts are changeable and vary with the AI implementation stages, use and reach (across part of the organisation).[18]

b. mandating transparency requirements across the private and public sectors, including how these requirements could be implemented.

The verdict on the value of mandating transparency may yet be open. However, the empirical evidence is clear: AI is bias prone; AI producers and users do not always share the same understanding of the purpose, utility and functionality of AI tools; AI producers cut corners to sell products that may not be suited to the task they are intended for; AI users are not fully aware of how and when their AI tools operate beyond their intended scope (boundary creep). All these risks require monitoring, which in turn requires transparency.

As a minimum, transparency mandates disclosure and documentation of algorithm design and programming, and independent review prior to point of sale.

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

This question begs a question as to whether there is a need to increase public trust. We would challenge this inbuilt assumption. Our current environment is pervasively influenced by AI, and there is much hand wringing - often from the tech industry and advisers (or critics) about the current or potential future risks of AI (and the people, perspectives and practices that support its development).

[17] Royal Decree-Law 9/2021, which modifies the consolidated text of the Workers' Statute Law (Royal Legislative Decree 2/2015) to guarantee the labor rights of people dedicated to delivery in the field of digital platforms <https://boe.es/diario_boe/txt.php?id=BOE-A-2021-7840>; Carmen Villarroel Luque 'Workers vs Algorithms' <https://verfassungsblog.de/workers-vs-ai/>.

[18] Andreas Cebulla, Zygmunt Szpak, Genevieve Knight 'Preparing to work with artificial intelligence: assessing WHS when using AI in the workplace' <https://doi.org/10.1108/IJWHM-09-2022-0141>.

The consultation paper cites *Trust in Artificial Intelligence: A Global Study*[19], in support of the notion of the primacy of surveyed perceptions around AI and the potential for greater uplift. However, we might critically analyze whether behaviour follows what people say in response to a survey, given that we are not seeing a widespread grassroots public campaign or behaviours that might be expected if there was a big trust issue. People may say one thing and do another. Convenience and cost issues are often larger drivers than trust, and people can continue to interact with systems that are novel, popular or appear to give advantage. We question the extent to which human behaviour reflects the rather simplistic flow chart calculus depicted in Figure 44 of the consultation paper.[20]

The intense activity that is occurring in many academic and business circles around AI ethics and governance, and that has been for many years, is operating in a bubble removed from the understanding of the vast majority of the population, most of whom would be unaware of the debates underway. This is not to downplay the relevance of some of the thinking in those debates, but it is raised to break the assumed correlation between stated trust and adoption. In any event, most people are not involved in adoption of AI. They are consumers of, and subject to the influence of, AI tools designed by others. They are almost certainly not aware of most of the AI systems they are interacting with (actively or passively; directly or as data subjects). In this respect we draw a contrast to the strong public reaction to certain types of biotechnology and agricultural practice that ran through the 1990s and resulted in regulation such as the Gene Technology Act 2000 (while noting that general levels of concern on those matters now seem to have abated).

Government action should not be to "increase public trust" (or "improve public trust" as under 9a), which is a passive construct of acceptance. Government ought to ensure that the way AI technology is used deserves public trust, that is, trust is earned and provenly warranted because the trustworthiness of AI technology is demonstrated.

Government ought to promote a critical understanding of AI risks amongst the public in general and workforces specifically. Initiatives should aim to build 'critical AI resilience', whilst legislation ought to create the venues for the application of that critical resilience, e.g., by mandating employee consultation processes.

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?

Whilst the term 'artificial intelligence' implies a degree of independence or autonomy, all AI based systems depend on data. Machine learning systems depend on large amounts of data which can be then used to make predictions about other data sets. The datasets used to train machine learning models frequently contain sensitive personal information. This sensitive information could include health status, sexual and gender identity, political allegiance such as union membership or criminal history. Even where machine learning models are not trained using sensitive personal information, they may still be used to infer this sensitive information about individuals.[21] Initially machine learning models were targeted at generating inferences about individuals for targeted advertisements but can be used to draw inferences in almost any

[19] N Gillespie, S Lockey, C Curtis, J Pool and A Akbari, 'Trust in Artificial Intelligence: A Global Study, The University of Queensland and KPMG Australia', p 14, 2023
[20] Model of the key drivers of trust and acceptance of AI systems (63)
[21] Sandra Wachter, 'Data Protection in the Age of Big Data' (2019) 2(1) Nature Electronics 6.

domain.[22] For instance, a machine learning model to predict the likelihood of a person being diagnosed with a disease could be used by health insurers to offer discriminatory pricing. Further, machine learning models have significant potential for 'dual use'.[23] These dual use capabilities have the potential to risk fundamental human rights or lead to unintended consequences. For example, a machine learning model trained to identify sexual identity could be used in a country where homosexuality remains a criminal offence.[24]

The use of machine learning to derive inferences represents a fundamental challenge to both privacy law and a risk assessment-based approach to regulating AI. Under the notice and consent model underpinning Australian privacy law, most regulatory activity focuses on how data is collected rather than how it is used. Once the data collector has obtained valid consent from an individual to use their information, they face limited restrictions on how they use this information. As others have written about extensively, the notice and consent model does not anticipate inferences being drawn using machine learning. Although an individual can access information which has been collected about them, or amend this information if incorrect, exercising this right depends on the individual knowing about this information. If individuals themselves are not aware of what these inferences are, they will not be able to exercise these rights.[25] Therefore, the notice and consent model offer limited tools for regulating the use of machine learning tools. Further, consent does not offer a guarantee against the use of machine learning tools for dual use purposes which the individual may not have anticipated.

There are two legal requirements which should be integrated into Australian privacy law to respond to the risks of big data exceptionalism and machine learning. These legal requirements can exist alongside a risk-based approach to regulating artificial intelligence and machine learning. First, privacy law should mandate that any organisation or entity using personal information to train a machine learning model must follow a 'privacy by design' approach. This approach would require the entity training the model to ensure appropriate technical and organisational measures exist to guarantee the security of personal information. These requirements would need to be implemented prior to the processing of any personal information, including training machine learning models. Implementing this requirement as a pre-requisite would help to ameliorate some of the weaknesses of the notice and consent model with respect to big data research. This privacy by design approach should also require consideration of any potential dual use risks that might arise with that machine learning model in the future.

Second, privacy legislation should mandate that entities implement security measures to protect any data used for training or processed with machine learning systems. There is a risk that even without releasing training dataset, inversion attacks can be conducted on a machine learning model to retrieve data.[26] Although specific technical measures should not be mandated in legislation, advanced privacy enhancing technologies could include homomorphic encryption, differential privacy, and secure multiparty

[22] 'Generating User Information for Use in Targeted Advertising' *United States US20050131762A1*, filed on 31 December 2003 (Issued on 16 June 2005) <https://patents.google.com/patent/US20050131762A1/en>.
[23] Anna Jobin, Marcello Ienca and Effy Vayena, 'The Global Landscape of AI Ethics Guidelines' (2019) 1(9) *Nature Machine Intelligence* 389.
[24] Marcello Ienca and Effy Vayena, 'Dual Use in the 21st Century: Emerging Risks and Global Governance' (2018) 148(4748) *Swiss Medical Weekly* w14688
[25] Helen Nissenbaum, 'Deregulating Collection: Must Privacy Give Way to Use Regulation?' *SSRN* (Working Paper, 2017 < https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3092282>.
[26] Michael Veale, Reuben Binns and Lilian Edwards, 'Algorithms That Remember: Model Inversion Attacks and Data Protection Law' (2018) 376(2133) *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 20180083.

computation. These technical measures should be complemented with appropriate organisational solutions such as separating data custodians and data processors.[27] This combined approach recognises that guarding against privacy threats is contextual and requires continual revision. We note that the Federal government has recently amended some of the privacy law framework and more changes are pending.

Regards,

Dr Andreas Cebulla, Associate Professor in The Future of Work
Robert Chalmers, Senior Lecturer
Dr Rajesh Johnsam, Senior Lecturer
Professor Tania Leiman, Professor and Dean of Law
Dr James Scheibner, Lecturer

[27] James Scheibner et al, 'Revolutionizing Medical Data Sharing Using Advanced Privacy-Enhancing Technologies: Technical, Legal, and Ethical Synthesis' (2021) 23(2) *Journal of Medical Internet Research* e25120.