THE LAW SOCIETY
OF NEW SOUTH WALES

Our ref: PDL:CBgl170723

17 July 2023

Dr James Popple
Chief Executive Officer
Law Council of Australia
PO Box 5350
Braddon ACT 2612

By email: john.farrell@lawcouncil.asn.au

Dear Dr Popple,

Safe and responsible AI in Australia

Thank you for the opportunity to contribute to the Law Council's submission to the Department of Industry, Science and Resources in response to the Safe and responsible AI in Australia Discussion Paper. The Law Society's Privacy and Data Law Committee has contributed to this submission.

General comments

In our view, the development of safe and responsible AI in Australia requires an interoperable framework that will enable Australian organisations to innovate while being carefully balanced with sufficient safeguards. The framework should be flexible, scalable, and future proof.

Flexibility
The framework should build upon, and be adapted to, existing processes that Australian organisations have in place; for example, enterprise risk frameworks and methodologies, software and other technology project assessment and management frameworks and methodologies, privacy and security by design and default, and privacy risk assessment. It must also be cognisant of existing laws; for example, privacy, data security, product safety, consumer protection and human rights-based laws such as anti-discrimination statutes.

Scalable
Since data and provision of cloud-based services have no geographic boundaries, the framework must be scalable. As different regulatory models in diversely regulated jurisdictions apply at various points in a data-driven service supply chain, AI regulatory initiatives should be determined with reference to evolving regulation in other jurisdictions. These models will impact links in the AI supply chain and Australia's assessment of the extent of the impact and effectiveness of that regulation to effect safe and responsible AI at the Australian end of that supply chain. The framework needs to take into consideration and leverage international initiatives that can facilitate responsible and accountable flows of data, and cross-border business models that enable Australian businesses to cost-effectively expand and compete globally.

Future proof
A futureproof framework will enable Australian organisations, not only as adopters of AI, but also as producers of AI.

Generally, we support adaptability through the adoption of principles-based legislation, providing for legal responsibility and accountability of entities across the AI service supply chain. This should include measures to ensure that those entities have appropriate incentives to adopt risk of harms assessments, mitigation and management of residual risks, supported by a risk management framework. The risk management framework should broadly align to existing risk frameworks, as to some degree, the fundamental risks remain the same, only these risks are amplified.

Australia's place in the global economy

Australia's regulatory approach to AI should be consistent with Australia's unique economic conditions, and its place in the global economy. At present, Australian businesses and other organisations are net deployers and users of third-party AI services, rather than developers and suppliers of AI services. There is value for Australian businesses in aligning the Australian approach, so far as is reasonably practicable and consistent with protection of Australian citizens, to the approaches followed in key regulating jurisdictions including the United Kingdom (UK), the European Union (EU), the United States of America (US), Canada and Singapore. If Australia adopted a bespoke approach, that does not take into consideration global and international negotiations underway on AI, this disconnect may have a detrimental economic impact and limit our ability to harness the benefits of AI and adequately safeguard against the risks.

While it is appropriate to seek a degree of alignment with countries with whom we share similar values such as the UK and the EU, particularly in the short term, it is critical that the approach adopted does not entrench Australia as a net user of AI, but rather provides for the longer term, by enabling Australian organisations to be developers and creators of AI. If Australia remains a net user, this is likely to have adverse long term economic repercussions.

The fundamental policy question

In considering the future responsible use of AI in Australia, the fundamental policy question is how we should proceed. One approach is to utilise existing regulatory frameworks and enhance their application to AI. The development of a standalone new framework without adequate consideration of existing frameworks is not supported. We note that international standards can be a useful policy tool to enable interoperability across regulatory initiatives and enable scalability of the AI services and products countries produce. We further note the current efforts of the UK, EU and US in leveraging international standards in this manner.

There is merit in reviewing the international frameworks that are being developed and considering whether Australia's approach should be aligned with those frameworks. However, if Australia monitors global developments for too long before participating in these policy initiatives and does not leverage its participation in existing international initiatives, it risks lagging behind, losing the benefits of shaping AI policy internationally and locally, and being exposed to these risks.

Aligning to the framework of a common law system, such as the UK will have inherent advantages and may be an appropriate approach for Australia to follow in the short term. The UK approach is principles based and implementation is through existing regulators, centrally co-ordinated for consistency. This model has the advantage that through industry specific regulators, the application of the principles can be appropriately nuanced to that industry. We also note the role of the UK's AI Standards Hub,[1] managed by the Alan Turing Institute, and its work in AI standards development, assessment and use.

A risk-based framework, as being developed by the EU, also has merit. It has the attraction that entities and corporations are already dealing with enterprise risk, and in our view, a risk-based framework for AI builds on the foundations of the same risks, but by new pathways. For example, enterprise risk could be expanded to address algorithm risk where an enterprise deploys AI. However, we are concerned that the EU approach may not provide adequate flexibility for a technology such as AI that is fast evolving.

We suggest that the concept of product stewardship is important in the approach adopted, meaning that each party in the supply chain, from the original developer to the final supplier, is accountable and responsible in relation to the use of AI. This requires appropriate standards of transparency and disclosure at each stage in the supply chain. Transparency between organisations, as well as transparency or visibility to the regulator is essential. Organisations must be able to understand how the AI product they deploy functions. Regulators need to be able to verify that there is appropriate risk management in place and that it is functioning as intended. Without transparency, the notion of Responsible AI is not, in our view, achievable.

We therefore support a risk-based approach, with a strong focus on transparency, accountability and responsibility. As mentioned, this should be developed cognisant of global approaches so as not to create unwanted difficulties or disconnects with the global economy. Leveraging international frameworks, such as the one being developed by the International Organization for Standardization (ISO)[2] across more than 50 countries (including the EU, UK, and US), can assist in ensuring that Australia has a regime that will enable it to consume AI products from the US, EU and other suppliers, as well as develop its own products for export.

Private sector and public sector uses of AI

The regulatory framework should support organisations, whether public or private, as users or implementers of AI. One of the challenges in developing the responsible use of AI, specifically for Generative AI, is that access to many AI tools is unrestricted, and it is quite feasible that in any given organisation, unknown and unsanctioned uses of AI are already occurring. From a perspective of ensuring there are adequate systems for monitoring the use of AI in an organisation, there may be a role for an AI gatekeeper who is responsible for how and when AI is being deployed in the organisation.

Where an organisation uses AI developed outside Australia, subject to a different regulatory framework, the organisation will need guidance to ensure the responsible use of AI. Whether the organisation is a government agency, private sector or a not-for-profit organisation, controls around the implementation of AI are needed through an appropriate risk assessment and management framework. That said, there are arguments to differentiate between private sector and public sector uses of AI and the associated governance obligations. For example, obligations to give reasons in the public sector must be supported by strong explainability and transparency practices.

We see merit in government acting as a role model, leading by example, in the adoption of ethical AI and responsible technology practices. In our view, the public sector should be held to a higher standard of responsible use of AI. The government should be a model user of AI,

[1] AI Standards Hub, accessed at https://aistandardshub.org/the-ai-standards-hub/#:~:text=The%20AI%20Standards%20Hub%20is%20led%20by%20The%20Alan%20Turing,and%20the%20Office%20for%20AI.

[2] For example, ISO/IEC JTC 1/SC 42 Artificial intelligence, International Organization for Standardization, accessed at: https://www.iso.org/committee/6794475.html.

assisting the creation of appropriate behaviours and standards which can then be applied more broadly to the private sector's use of AI. We suggest, on that basis, that there is merit in considering the NSW Artificial Intelligence Assurance Framework[3]. This framework was developed in NSW to assist NSW agencies design, build and use AI-enabled products and solutions, and to help agencies identify risks that may be associated with their projects. Consideration could be given to implementing a similar framework nationally.

Monitoring and review

Given the rapid development of AI, it is important that any framework, regulatory or otherwise, is subject to shorter, accelerated review cycles than would ordinarily apply. For example, the common statutory review period of five years would be inappropriate in our view. We do not yet know all the tasks to which AI will be deployed, making it difficult to build adequate safeguards, suggesting myriad approaches may be necessary. The recent high profile data breaches that have occurred in Australia are, in our view, an example of mature legislation, with the *Privacy Act 1988*, and established cybersecurity requirements, proving to be insufficient safeguards against bad actors. The exponential rate of change in this environment demands continuous monitoring.

A risk management approach where the operation of AI is subject to extensive monitoring and detection systems has merit. This aligns with the approach to cyber security principles adopted in the Australian Government's Signals Directorate Information Security Manual[4] of "Govern, Protect, Detect and Respond" and also aligns with the NIST AI Risk Management Framework[5] of "Govern, Map, Measure, Manage" and the ISO risk management approach to AI. Under these frameworks, the focus is on ensuring adequate systems for monitoring and detecting issues are in place such that when an adverse consequence does arise, it may be responded to quickly. This more flexible approach rather than a more prescriptive and prohibitive approach has merit in our view and can be used to inform a governance framework that assesses risk and provides redress.

In our view, many AI risks can be addressed by establishing appropriate detect and respond incentives, and therefore an upfront restriction or prohibition is not required or justified. It is also very difficult to design appropriate and futureproof upfront restrictions or prohibitions for AI applications, given the rapidly evolving, changing and unpredictably diverse ways in which AI is already being used to assist humans in performance of myriad tasks.

We suggest that upfront restriction or prohibition is not required or justified if:

- prompt detection of a significant harm to human or the environment is likely,
- financial recompense to affected persons that have suffered that harm is appropriate to redress their loss/damage,
- penalties are appropriately substantial, and
- recovery of damages or penalties is sufficiently likely that entities are incentivised to properly mitigate risks of relevant harms.

Each regulated entity should in this circumstance apply risk of harms assessment appropriate to mitigate risks of relevant harms. Transparency requirements may be particularly important to ensure that prompt detection of a significant harm to humans or the environment, and

[3] NSW Artificial Intelligence Assurance Framework, accessed at https://www.digital.nsw.gov.au/sites/default/files/2022-09/nsw-government-assurance-framework.pdf.

[4] Australian Government's Signals Directorate Information Security Manual, published 2 March 2023, accessed at https://www.cyber.gov.au/sites/default/files/2023-03/02.%20ISM%20-%20Cyber%20Security%20Principles%20%28March%202023%29.pdf.

[5] National Institute of Standards and Technology AI Risk Management Framework accessed at https://airc.nist.gov/AI_RMF_Knowledge_Base/Playbook.

attribution of that harm to a particular AI activity conducted by a regulated entity, is sufficiently likely. Requirements as to transparency of the fact that an AI risk assessment has occurred, although not necessarily the content of the AI assessment, may also assist in creating incentives to ensure that risks have been appropriately mitigated by regulated entities.

Human rights and AI

The Discussion Paper identifies on page 8 that 'Algorithmic bias is often raised as one of the biggest risks or dangers of AI' and further notes the major focus on this issue in the Australian Human Rights Commission's *Human Rights and Technology Report* in 2021.

Data-driven AI enables intentionally or unintentionally differentiated treatment of individuals and groups in Australian society. This differentiation may affect bias or other errors. Data-driven AI outputs may be based upon, create or amplify misinformation or disinformation, or produce outputs that are otherwise unreliable or unsafe for the reliance that humans place upon those outputs. Data used to produce those outputs may reveal information about an individual person's characteristics, interests, attributes and activities in both public and private spaces. Both regulated personal information, and other non-identifying information, may be used in ways that are beneficial, or in ways that are unreliable, unsafe or otherwise cause harms to those persons, impacting their human rights and legitimate expectations to be informed of that use. The question of when uses of AI are reasonable, appropriately transparent and justified, is broader than legal assurance of protection of the right to privacy and other human rights. That noted, it is crucial that the human rights impact, including the privacy impact of the operation of AI should be part of the key considerations in determining Australia's framework.

Algorithmic bias, particularly its potential impact on vulnerable people, should be addressed through appropriate risk management processes. The responsible use of AI must address questions of fairness dictated by the required context. It must include safeguards to manage both data quality and human bias.[6]

When Australian consumers or entities use AI, the ultimate user may be exposed to AI that has been developed by technical experts with limited or no training in ethics or human rights, and without appropriate oversight. The challenge for consumers is that they do not have visibility of any incorrect decisions made about them using AI and data. Without further detail, merely notifying consumers that an AI system is being used is likely to provide insufficient redress when a consumer is adversely impacted. Similar issues arise in respect of the limitation of privacy notices in providing adequate consumer protection. Careful consideration of appropriate remedies is required. Liability should be considered in light of existing product laws and developments, such as the EU Artificial Intelligence Liability Directive.[7]

Closely aligned with an approach based on human rights is a harm minimisation approach which considers what are the potential harms to humans and then regulates accordingly. The approach adopted in Canada is anchored to its human rights regime but is articulated in terms of reducing the risks and harms associated with AI. Given the absence of a federal Bill of Rights, it may be more appropriate, in our view, to frame considerations through the lens of harm minimisation.

[6] See ISO/IEC CD TS 12791 Information technology — Artificial intelligence — Treatment of unwanted bias in classification and regression machine learning tasks, accessed at https://www.iso.org/standard/84110.html?browse=tc.

[7] Briefing – EU Legislation in Progress, Artificial intelligence liability directive, published by the European Parliamentary Research Service, February 2023, accessed at https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/739342/EPRS_BRI(2023)739342_EN.pdf.

Automated decision making (ADM)

Transparency is critical for the responsible use of ADM by Australian organisations, both in the public sector and private sector. Citizens should know when and how ADM is being used in any way which significantly affects their human rights, their legitimate expectations to be informed of how and why they are being singled out for differentiated treatment, and their legitimate expectation that an automated decision is reasonable having regard to the circumstances in which it is made and the impact that this automated decision might reasonably be expected to have on affected humans and the environment.

Where an ADM tool is used to make a discretionary decision, we suggest that the tool produce an output report which can be provided to an affected citizen in such a way that they are able to interrogate the results and identify errors (rather than, for example, incomprehensible 'raw data' which would require expert knowledge to interpret). This is consistent with the Australian Human Rights Commission's recommendation that individuals have a right to reasons for automated decisions affecting them.[8]

In our view, people affected by fully or partially automated decisions should not be limited from accessing administrative law review and accountability mechanisms, such as the Commonwealth Ombudsman, merits review and freedom of information applications. We note that access to these forms of review was restricted by Centrelink throughout the operation of the "Robodebt" program. Independent review could have improved the operation of that service and reduced the magnitude of its harm. The impact of restricting access to administrative law forums is evident in the eroding of public trust that resulted from that program. The costs associated with implementing accountability will ultimately result in better systems and processes.

Infrastructure and AI

In our view, AI is likely to be used extensively in the operation of infrastructure. This is an area that requires particularly wide-reaching safeguards, especially in relation to critical infrastructure. Current legislation in relation to critical infrastructure does not, in our view, sufficiently address the operation of AI.

Privacy

In considering the role for Australia's privacy law and its possible further development to address AI risks of harms, as compared to European laws including the European General Data Protection Regulation (GDPR), it is important to note that Australia does not have a federal Bill of Rights to support the jurisprudence that underlies how the GDPR is interpreted and applied in European courts. The GDPR is given a more extensive and protective application than Australia's privacy laws because European courts give effect to human rights jurisprudence when interpreting the GDPR. Without similar rights-based jurisprudence in Australia, it is particularly important that AI trustworthiness is legally assured without principal reliance upon human rights law.

In our view, risk management and any assessment developed as part of the proposed framework should take into account existing requirements and processes under the *Privacy Act 1988*, such as privacy impact assessments. Such considerations are important from the perspective of a co-ordinated and holistic regulatory approach and will assist in limiting the compliance burden on organisations. A sensible approach to AI regulation is to ask whether rules that restrict or prohibit particular uses of AI, or that mandate application of a particular

[8] Australian Human Rights Commission, *Human Rights and Technology* (Final Report, 2021) 62.

risk assessment framework or methodology, are justified, or whether detect and respond incentives are sufficient to cause appropriate mitigation of risks by regulated entities.

If you have any questions in relation to this submission, please contact Gabrielle Lea, Senior Policy Lawyer, by phone (02) 9926 0375 or by email to gabrielle.lea@lawsociety.com.au.

Yours sincerely,

CBanks

Cassandra Banks
President