THE LAW SOCIETY
OF NEW SOUTH WALES

Our ref: PDL:BMml270924

27 September 2024

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

By email: Nathan.MacDonald@lawcouncil.au; John.Farrell@lawcouncil.au

Dear Dr Popple,

Mandatory guardrails for AI in high-risk settings

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 Proposals Paper for introducing mandatory guardrails for AI in high-risk settings. The Law Society's Privacy and Data Law Committee contributed to this submission.

Defining high-risk AI

In our view, the principles-based approach to determining high-risk AI allows for flexibility and adaptability. This is consistent with our previous support for principles-based legislation that would allow for flexibility, adaptability and a futureproof framework for AI.[1]

However, from a compliance point of view, and to assist in the design of AI tools, we believe developers and deployers of AI would benefit from a list of practical examples under each of the proposed principles. Given the narrow scope of the Proposals Paper, focusing only on AI in high-risk settings, we believe there is benefit in being more prescriptive in giving examples of what constitutes high-risk. This is informed by the consideration that the risk appetite and the use of AI in decision-making across different industries can be substantially different, and the measure for 'adverse impact' may also differ.

Therefore, we believe more clarity and certainty will result in combining a principles approach with non-exhaustive lists of examples under each of the principles, similar to the guidance provided for the *Security of Critical Infrastructure Act 2018*.

In terms of whether Government should consider banning any high-risk use cases, we believe any regulatory response should carefully balance mitigating risk with enabling innovation in AI. In our view, there is merit in banning AI practices that have an unacceptable level of risk, that is, where the risk cannot be mitigated, or the consequences of the practice pose unacceptable and irremediable harm to individuals and communities. To balance this with not

[1] Law Society of New South Wales, Safe and responsible AI in Australia (Submission 17 July 2023).

stifling innovation in AI, any ban of AI practices could potentially be implemented by way of subordinate legislation, to allow sufficient flexibility while AI continues to evolve. We suggest any ban contain sufficient certainty in the definition and interpretation of the prohibited practice, and clarity about why the risk is unacceptable.

It may be instructive to refer to the AI practices that are prohibited under the European Union's Artificial Intelligence Act (EU AI Act), for their incompatibility with individual and collective rights and fundamental values, such as the rule of law.

AI practices prohibited under Article 5 of the EU AI Act include:[2]

(a) Subliminal techniques which can materially distort a person's behaviour by impairing their ability to make an informed decision in a way that causes, or is reasonably likely to cause, them significant harm.
(b) Exploiting the vulnerabilities of a person or specific groups of people (for example, due to their age, disability or economic situation) which can materially distort their behaviour in a way that causes, or is reasonably likely to cause, them significant harm.
(c) Social scoring systems based on known, inferred, or predicted personality characteristics which causes detrimental or unfavourable treatment that is disproportionate, or used in a context unrelated to the context in which the data was originally collected.
(d) Risk assessment systems which assess the risk of a person to commit a crime or re-offend (except in support of a human assessment based on verifiable facts).
(e) Indiscriminate or untargeted web-scraping for the purposes of creating or enhancing facial recognition databases.
(f) Emotion recognition systems in the workplace or educational institutions (except for medical or safety reasons).
(g) Biometric categorisation systems used to infer characteristics, such as race, political opinions or religion.
(h) Real-time, remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement except (subject to safeguards and within narrow exclusions) searching for victims of abduction, preservation of life, and finding suspects of certain criminal activities. Real time means live or near-live material, to avoid short recording delays circumventing the prohibition.

To our previous point about providing certainty to the definition of prohibited practices, we believe categories such as (e) and (f) might leave open for interpretation the meanings of 'indiscriminate' or 'untargeted', and 'emotional recognition'. If sufficient certainty is achieved, this could provide clear expectations and safeguards from the outset about the necessity of protecting fundamental rights and values, and remove ambiguity for organisations who might seek to interpret the principles in a way that compromises the rights of the end user.

Proposed mandatory guardrails

We acknowledge that 9 of the 10 proposed mandatory guardrails are identical to the Voluntary AI Safety Standard. We note the Voluntary AI Safety Standard has been mapped against Australia's AI Ethics Principles and broadly aligns with existing frameworks (such as the ISO 42001:2023 and the US National Institute of Standards and Technology's Artificial Intelligence Risk Management Framework). Our comments on each of the guardrails are below.

[2] Thomson Reuters UK, Practice Note: EU AI Act, https://uk.practicallaw.thomsonreuters.com/w-042-3394?transitionType=Default&contextData=(sc.Default)

Proposed mandatory guardrail	Law Society comments
1. Establish, implement and publish an accountability process including governance, internal capability and a strategy for regulatory compliance	We believe that the process for accountability needs to be a best practice process.
2. Establish and implement a risk management process to identify and mitigate risks	We believe specific examples of risk management processes or frameworks are necessary, particularly for an organisation that is the final end user of an AI system. We note the Proposals Paper suggests that risk should be assessed on a use case basis. However, this can be difficult for AI systems with multiple use cases, which supports our suggestion for examples of risk management processes to be provided.
3. Protect AI systems, and implement data governance measures to manage data quality and provenance	We believe developers and deployers would benefit from guidance that establishes the benchmark for data quality, data provenance, and data security.
4. Test AI models and systems to evaluate model performance and monitor the system once deployed	We suggest that the testing of AI models and AI systems on an ongoing basis needs to be performed by persons with accreditation in the subject matter. We also recommend for the testing to go beyond pure monitoring, such that the results are reported to an accountable authority within the organisation.
5. Enable human control or intervention in an AI system to achieve meaningful human oversight	We support this guardrail.
6. Inform end-users regarding AI-enabled decisions, interactions with AI and AI-generated content	We support this guardrail. Given the introduction of the Privacy and Other Legislation Amendment Bill 2024, we believe it would be beneficial for further clarity to be provided on the relationship between this guardrail and the Bill's proposed requirement for disclosure in privacy policies if personal information is used in automated decision-making, which could involve AI systems.
7. Establish processes for people impacted by AI systems to challenge use or outcomes	We suggest the processes include clear guidelines for any challenge of the use or outcomes of AI systems to be dealt with in a timely manner, with clear communication mechanisms.
8. Be transparent with other organisations across the AI supply chain about data, models and systems to help them effectively address risks	We have consistently supported the need for transparency, on the basis that transparency brings robustness to AI regulation.
9. Keep and maintain records to allow third parties to assess compliance with guardrails	We suggest that the record-keeping processes should be consistent with external standards, with reference to

	examples such as the *Archives Act 1983* (Cth) or the *State Records Act 1998* (NSW).
10. Undertake conformity assessments to demonstrate and certify compliance with the guardrails	We understand that a 'conformity assessment' is intended to be an accountability and quality assurance mechanism that relies on documentation captured pursuant to Guardrail 1 and 9. We suggest that conformity assessments be carried out by an accredited third party, or by government entities or regulators, rather than the organisations themselves, and that the accountability and quality assurance metrics be standardised.

Regulatory options

As a general comment, we have long supported an interoperable AI regulatory framework, and advocated for consistency with related legislation, including privacy, data security, product safety, consumer protection, intellectual property, defamation, and human rights law.[3] We support the ability of consistent definitions across regulatory frameworks to facilitate greater certainty and compliance among industries.

We believe the principles-based approach to the guardrails would be assisted by enforcement through a legislative framework, which may allow for flexibility, while incentivising compliance through its legislative basis. An analogous model is the Australian Privacy Principles, which is given legal force by the *Privacy Act 1988*, but remains flexible, as the principles are interpreted by guidance.

In our view, option 2 appears to adopt a middle-ground approach between options 1 and 3. While option 1 appears to be beneficial for reducing the risk of inconsistency or duplication across regulatory frameworks, it is likely to be a more time-consuming project, given the need to review and amend existing laws to embed the guardrails.

The benefit of option 2 appears to be its ability to adapt existing regulatory frameworks and expand the powers of existing regulators to account for high-risk AI. However, given its reliance on amending existing laws, it is unlikely to capture AI developers, or other areas that existing laws do not apply to.

Option 3 might allow for Australia to be more consistent with the broader global scheme, given the recent implementation of the EU AI Act, and the UK's plan to introduce a standalone AI Act. This comment is also informed by the consideration that the guardrails are intended to only apply to AI in high-risk settings, thereby limiting the scope of application to a limited set of AI uses and outcomes. However, this approach will most likely result in overlap in existing laws and may require frequent amendment for currency to keep up with the fast-evolving pace of AI development.

[3] Law Society of New South Wales, Submission No 30 to the NSW Legislative Council, *Parliamentary Inquiry in Artificial Intelligence (AI) in New South Wales*, 20 October 2023, 7-9.

If you have any questions in relation to this submission, please contact Mimi Lee, Policy Lawyer, by phone (02) 9926 0174 or by email to mimi.lee@lawsociety.com.au.

Yours sincerely,

Brett McGrath
President