Law Council
OF AUSTRALIA

Introducing mandatory guardrails for AI in high-risk settings: Proposals Paper

Department of Industry, Science and Resources

9 October 2024

Telephone +61 2 6246 3788
Email mail@lawcouncil.au
PO Box 5350, Braddon ACT 2612
Level 1, MODE3, 24 Lonsdale Street,
Braddon ACT 2612
Law Council of Australia Limited ABN 85 005 260 622
www.lawcouncil.au

Table of contents

About the Law Council of Australia **3**
Acknowledgements **4**
Introduction **5**
Intersection with other consultations **6**
Bans on high-risk AI **6**
Defining high-risk AI **8**
Principles versus a list-based approach 8
Intellectual property rights 9
Comments on specific principles 10
Principle (c): The risk of adverse legal effects, defamation or similarly significant effects on an individual 10
Principle (d): The risk of adverse impacts to groups of individuals or collective rights of cultural groups 10
Social utility 11
Biometrics 12
Administration of justice and democratic processes 12
Categories of use to be designated 'high-risk' 14
Options for introducing mandatory guardrails **14**
General comments 15
Views on the proposed options 15
Proposed mandatory guardrails 16
General comments 16
Intellectual property enforcement 17
Protections for domestic end-users 18
Specific comments 18
Other comments **20**
Environmental harm 20
Oversight 20
Use of AI to replicate personhood 21

About the Law Council of Australia

The Law Council of Australia represents the legal profession at the national level; speaks on behalf of its Constituent Bodies on federal, national, and international issues; promotes and defends the rule of law; and promotes the administration of justice, access to justice and general improvement of the law.

The Law Council advises governments, courts, and federal agencies on ways in which the law and the justice system can be improved for the benefit of the community. The Law Council also represents the Australian legal profession overseas, and maintains close relationships with legal professional bodies throughout the world. The Law Council was established in 1933, and represents its Constituent Bodies: 16 Australian State and Territory law societies and bar associations, and Law Firms Australia. The Law Council's Constituent Bodies are:

- Australian Capital Territory Bar Association
- Law Society of the Australian Capital Territory
- New South Wales Bar Association
- Law Society of New South Wales
- Northern Territory Bar Association
- Law Society Northern Territory
- Bar Association of Queensland
- Queensland Law Society
- South Australian Bar Association
- Law Society of South Australia
- Tasmanian Bar
- Law Society of Tasmania
- The Victorian Bar Incorporated
- Law Institute of Victoria
- Western Australian Bar Association
- Law Society of Western Australia
- Law Firms Australia

Through this representation, the Law Council acts on behalf of more than 104,000 Australian lawyers.

The Law Council is governed by a Board of 23 Directors: one from each of the Constituent Bodies, and six elected Executive members. The Directors meet quarterly to set objectives, policy, and priorities for the Law Council. Between Directors' meetings, responsibility for the policies and governance of the Law Council is exercised by the Executive members, led by the President who normally serves a one-year term. The Board of Directors elects the Executive members.

The members of the Law Council Executive for 2024 are:

- Mr Greg McIntyre SC, President
- Ms Juliana Warner, President-elect
- Ms Tania Wolff, Treasurer
- Ms Elizabeth Carroll, Executive Member
- Ms Elizabeth Shearer, Executive Member
- Mr Lachlan Molesworth, Executive Member

The Chief Executive Officer of the Law Council is Dr James Popple. The Secretariat serves the Law Council nationally and is based in Canberra.

The Law Council's website is www.lawcouncil.au.

Acknowledgements

This submission is informed by contributions from the following Constituent Bodies:

- the Law Institute of Victoria;
- the Law Society of New South Wales; and
- the Queensland Law Society.

The Law Council also acknowledges the guidance of its Futures Committee, and the contributions of its Business Law Section, specifically the Intellectual Property Committee and the Media and Communications Committee.

Introduction

1. The Law Council of Australia welcomes the opportunity to provide a submission to the **Department** of Industry, Science and Resources in response to its **Proposals Paper** on the introduction of mandatory guardrails for artificial intelligence (**AI**) in high-risk settings.

2. In our view, the Proposals Paper has been thoughtfully prepared. It is based on the premise that AI (in its various forms) has the potential to improve social and economic well-being, whilst also bringing both foreseeable and unforeseeable risks of harm that may not be adequately remediable under Australia's current regulatory framework.

3. The tenor of the Proposals Paper is to minimise the risk of potential harms without unduly limiting innovation or driving it offshore. The Law Council considers it critical that any regulatory framework adopted in Australia should be flexible and scalable to future technological change, and so will allow for sufficient innovation, while maintaining appropriate guardrails. To this end, we agree with the Proposals Paper's focus on a risk-based approach with an emphasis on preventative measures.

4. In August 2023, we provided a submission to the Department as part of its '*Safe and Responsible AI in Australia'* consultation.[1] In that submission, we recommended that the Australian Government should adopt a risk-based approach that, in the short term, focuses on the regulation of high-risk AI technology and applications that are clearly egregious and pose significant risks to the privacy and way of life of Australians, as per the approach in the *European Union Artificial Intelligence Act* (**EU AI Act**).[2] In particular, we recommended enhanced regulation of the collection and use of biometric information (such as the use of automated facial recognition technology) and 'social scoring' practices, and options to reduce the risk of people being misled by AI-generated fakes and scams.[3]

5. The Proposals Paper canvasses various approaches to AI regulation across other, pro-active jurisdictions, including the European Union, Canada and the United States of America, and has, commendably, collated considerable research. The Proposals Paper poses important questions and offers pathways forward for legislating the use of AI in high-risk settings, and the use of general-purpose AI (**GPAI**) in Australia, underpinned by proposed mandatory guardrails that closely mirror the existing Voluntary AI Safety Standard.[4]

6. Identification of high-risk settings can facilitate further discussion on the additional guardrails and the extent of regulatory intervention required to reduce the likelihood of harm, whether through existing, technology-neutral frameworks (such as privacy laws, healthcare, anti-discrimination, and financial sector laws) or, in specific cases, through dedicated legislation.

7. Doing business in Australia means doing business in a multi-obligation environment that is already complex. For many, this complexity is a barrier to entry. There are consumer laws (dealing with misleading and deceptive conduct), privacy laws (governing use of sensitive personal information including biometric information), corporations laws (imposing directors' duties requiring due care and skill), online safety laws, intellectual property rights laws, anti-discrimination laws and sector-

[1] Law Council of Australia, Submission to Department of Industry, Science and Resources, *Safe and Responsible AI in Australia* (17 August 2023) 16-21.
[2] Ibid 16-21.
[3] Ibid.
[4] Commonwealth of Australia, *Voluntary AI Safety Standard* (August 2024).

specific laws. Many of these laws have the capacity to—and already—deal with data and technology use and misuse.

8. However, given the pace of technological change and the economy-wide embrace of AI (albeit in many cases still being tested), we agree that mandatory guardrails applying across the AI supply chain and throughout the AI lifecycle could be effective to reduce the likelihood of harms. There is merit in limiting specific AI-related laws to high risk and GPAI use cases, until a further, more meaningful review can be undertaken, based on a growing body of evidence and global AI use case examples.

Intersection with other consultations

9. As a general note, the Proposals Paper should not be viewed in isolation. There are several ongoing and recent consultations that will be relevant to discussions about the regulatory approaches to AI in Australia, including:

    - the Parliamentary Joint Committee on Human Rights' inquiry into Australian's Human Rights Framework, and its report of May 2024;[5]
    - the Privacy Act Review Report and the subsequent Government response, in addition to the first tranche of reforms to the *Privacy Act 1988* (Cth) introduced to Parliament in September 2024.[6] The Privacy and Other Legislation Amendment Bill is currently the subject of an inquiry by the Senate Legal and Constitutional Affairs Legislation Committee, due to report by 14 November 2024;[7] and
    - recommendations arising from the inquiry into the opportunities and impacts for Australia arising out of the uptake of AI technologies, currently being undertaken by the Select Committee on Adopting Artificial Intelligence, due to report by 26 November 2024.[8]

10. We also anticipate that a variety of action items and recommendations will arise from the Copyright and Artificial Intelligence Reference Group, established by the Commonwealth Attorney-General in December 2023.[9]

Bans on high-risk AI

11. In terms of whether the Government should consider banning any high-risk use cases, we consider that any regulatory response should carefully balance mitigating risk with enabling innovation in AI.

12. 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.

13. However, to ensure that innovation in AI is not simultaneously stifled, any ban of AI practices could potentially be implemented by way of subordinate legislation, to allow

---

[5] Parliamentary Joint Committee on Human Rights, *Inquiry into Australia's Human Rights Framework* (Report, May 2024)
[6] Attorney-General's Department, *Privacy Act Review Report* (February 2023); Commonwealth of Australia, *Government Response: Privacy Act Review Report* (September 2023); Privacy and Other Legislation Amendment Bill 2024 (Cth);
[7] Senate Legal and Constitutional Affairs Committee, *Privacy and Other Legislation Amendment Bill 2024 [Provisions]* (Web Page, 2024).
[8] Select Committee on Adopting Artificial Intelligence (Web Page, 2024).
[9] Commonwealth Attorney-General's Department, *Copyright and Artificial Intelligence Reference Group* (Web Page, 2024).

sufficient flexibility while AI continues to evolve. We suggest that any ban contain sufficient certainty in the definition and interpretation of the prohibited practice, and clarity about why the risk is unacceptable.

14. It may be instructive to refer to the AI practices that are prohibited under the EU AI Act for their incompatibility with individual and collective rights and fundamental values, such as the rule of law.[10]

15. AI practices that are prohibited under Article 5 of the EU AI Act include:

    (a) Subliminal techniques that 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 cause detrimental or unfavourable treatment that is disproportionate, or used in a way that is unrelated to the context in which the data was originally collected.

    (d) Risk assessment systems that assess the risk of a person committing a crime or re-offending (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' must mean live or near-live material, to avoid short recording delays circumventing this prohibition.

16. We note that some categories of prohibited AI practices—such as those in paragraphs (e) and (f), above—could leave the meanings of 'indiscriminate', 'untargeted', and 'emotional recognition' open for interpretation. However, if sufficient certainty is able to achieved, this would assist to provide clear expectations and safeguards, from the outset, about the necessity of protecting fundamental rights and values. This certainty could also remove ambiguity for organisations that might seek to interpret the principles in a way that compromises the rights of the end user.

[10] See Law Council of Australia, *Rule of Law Principles* (Policy Statement, March 2011).

Defining high-risk AI

17. The Proposals Paper sets out the following principles, designed to guide consistent assessment by organisations of whether the use of an AI system is high-risk, and therefore subject to mandatory guardrails shaping its development and use, based on intended and foreseeable uses:[11]

> *In designating an AI system as high-risk due to its use, regard must be given to:*
>
> *a. The risk of adverse impacts to an individual's rights recognised in Australian human rights law without justification, in addition to Australia's international human rights law obligations*
>
> *b. The risk of adverse impacts to an individual's physical or mental health or safety*
>
> *c. The risk of adverse legal effects, defamation or similarly significant effects on an individual*
>
> *d. The risk of adverse impacts to groups of individuals or collective rights of cultural groups*
>
> *e. The risk of adverse impacts to the broader Australian economy, society, environment and rule of law*
>
> *f. The severity and extent of those adverse impacts outlined in principles (a) to (e) above.*

Principles versus a list-based approach

18. We recognise that the nature of risk in the development and use of AI is a sliding scale, and risk frameworks are likely to evolve over time. In this respect, AI 'risk' is contextual, with risks arising from its use by individuals and organisations, as well as from the technology itself.[12]

19. There are also different kinds of AI systems—that is, expert, scientific and generative systems—some of which may be used every day, which are not differentiated in the Proposals Paper. This distinction is important when considering risks associated with AI systems, and therefore any accompanying regulatory responses.

20. We acknowledge the advantages of a principles-based approach, in that it avoids inadvertently capturing some uses of AI that are, in fact, low-risk. However, while this approach seeks to respond to the contextual basis of risk of harm that may be presented by an AI tool, it may create challenges in terms of providing sufficient clarity and certainty for organisations. For example:

    - It is currently unclear whether proposed principles (c) and (d) are intended to be applied to people who are not users of the AI system, such as third parties, or groups whose rights could be impacted; and
    - proposed principle (e), in particular, lacks specificity.

[11] Proposals Paper, 19.
[12] Lisa M. Given, 18 January 2024, *Australia plans to regulate 'high-risk' AI. Here's how to do that successfully*, The Conversation (Online, 18 January 2024).

21. We consider that, as a general matter, it is appropriate for the creator of an AI system to consider all third parties and groups whose legal rights are—or could be—affected by the creation, operation or use of the AI system.

22. The proposed principles for categorising AI as 'high risk' are broadly framed, and this will likely make them difficult to apply in practice. For example, the Proposals Paper considers what the characteristics of high risk *might* mean, but it does not seek to define the circumstances under which an AI system *will* be high risk or produce a high-impact result.

23. On the other hand, while a list-based approach has the advantage of promoting clarity and certainty, we are concerned that a static list may not evolve quickly enough to capture novel use cases.

24. Accordingly, we support a balanced approach to regulating AI, combining both a prescriptive and principles-based approach to high-risk settings.

25. Given the narrow scope of the Proposals Paper (focusing only on AI in high-risk settings) there would be benefit in being more prescriptive by giving illustrative examples of what constitutes high-risk scenarios. We consider that more clarity and certainty will result from combining a principles approach with non-exhaustive lists of examples under each of the principles, similar to the approach taken to the guidance provided under the *Security of Critical Infrastructure Act 2018* (Cth).

26. We acknowledge that AI systems are likely to develop more rapidly than any accompanying guidance materials. As such, if a principles-based approach is adopted, it is important that the principles are workable, and can be readily applied by the AI system owner to determine whether a particular AI system is high-risk.

Intellectual property rights

27. It is a question of policy as to whether an AI system is to be designated as high-risk (and thereby attracting the *ex ante* regulatory obligations of such systems) if it may involve—in its creation or use—the infringement of intellectual property rights.

28. If this is not the intended policy outcome, then some clarification to the definition of 'high-risk' may be required. Clarification would also be required if it is intended that the risk of infringement of only certain types of intellectual property rights—such as Indigenous cultural and intellectual property rights—are to be taken into account in the 'high-risk' determination.

29. On the other hand, if it is intended that the risk of infringement of intellectual property rights generally may make an AI system 'high-risk', then it is important that intellectual property law reform in relation to AI systems keeps pace with the proposed *ex ante* regulatory regime. This will ensure that there is sufficient clarity concerning, for example, questions of infringement.

30. We note the law reform activity currently being conducted by the Attorney-General's Department in relation to copyright law, in consultation with the Copyright and Artificial Intelligence Reference Group,[13] and to the proposed development of stand-alone legislation in respect of Indigenous cultural and intellectual property.[14]

[13] Commonwealth Attorney-General's Department, *Copyright and Artificial Intelligence Reference Group* (Web Page, 2024).
[14] See Commonwealth Department of Infrastructure, Transport, Regional Development, Communications and the Arts, *New stand-alone legislation to protect ICIP – Topics for discussion* (Web Page, 2024).

31. We consider that any implementation of the proposed mandatory guardrails may lend a degree of urgency to those related reforms, especially if the mandatory guardrails create unintended policy settings.

Comments on specific principles

Principle (c): The risk of adverse legal effects, defamation or similarly significant effects on an individual

32. From an intellectual property perspective, proposed principle (c) may be difficult to apply in circumstances where (as is currently the case) the copyright infringement risks of inputs and outputs of AI systems have not been clarified, unless it is intended that principle (c) be interpreted very narrowly. This means that principle (f) ("the severity and extent of those adverse impacts") may also be difficult to apply in practice.

33. For instance, the discussion on principle (c) in the Proposals Paper focuses on the effects on an individual who engages with the AI system,[15] rather than posing the broader question of whether the AI system affects the legal rights of any individual. In an intellectual property context, this includes:

    (a) copyright owners whose rights are infringed by the training of the AI system, the use of accessing of information by the AI system in response to a request or instruction, and/or the AI system's creation of a work in response to queries or instructions; and/or

    (b) individuals whose ability to make use of works generated by AI, with minimal human input, might be curtailed by the inability to obtain licences, or, at least, have clarity regarding copyright status and ownership.

34. As such, it is not clear whether principle (c) is intended to require the AI system owner to assess only the impact of the system on the legal rights of individual end users who are actually engaging with that AI system.

Principle (d): The risk of adverse impacts to groups of individuals or collective rights of cultural groups

35. As with principle (c), we consider that principle (d) could be difficult for an AI system owner to assess, particularly if the assessment is based on the potential of the AI system to cause such an adverse impact.

36. In response to Consultation Question 2, relating to First Nations peoples, we make the following preliminary comments for the Department's consideration in respect of proposed principle (d):

    - In Australia, the consequences of algorithmic bias have the potential to be especially concerning with respect to First Nations peoples. Algorithmic bias must, therefore, be front of mind when introducing Australia's AI regulatory regime.
    - It is especially important that AI regulations strive to ensure that inputs, throughputs (being material or items passing through a system or process),[16] and outputs are free from bias. The risk of inadequate, or unacceptably

[15] Proposals Paper, 22.
[16] *Cambridge Dictionary* (Online at 25 September 2024) 'throughput'.

limited, datasets, being used to train AI technologies must be addressed through suitable risk management procedures, including greater transparency.

- Further, concerns have been raised within our membership about the utilisation of ICIP, including literary and artistic works, in the training of AI. We recently provided a submission to the Department of Infrastructure, Transport, Regional Development, Communication, and the Arts in response to its consultation on a new legislative framework to address and protect Indigenous cultural and intellectual property rights,[17] and look forward to continuing to engage with the Government as its consideration of that framework progresses.

37. Beyond this, we submit that the Department should conduct meaningful consultation with First Nations peoples and communities to ensure that the proposed principles (particularly principle (d)) will sufficiently capture high-risk AI that may inappropriately use Indigenous cultural and intellectual property, or may, more broadly, utilise data about individuals within First Nations communities.

Social utility

38. Feedback received by the Law Council suggests that what may be missing from both the principle and list-based approaches is a focus on the social utility of the use case from which the high or low risks emanate. If a graduated regulatory approach is adopted—which would allow more time for use cases to be developed and risks (particularly the 'unknown unknowns') to emerge for a more informed legislative approach (if needed)—then it may be helpful to move beyond a focus on use case 'type'.

39. In addition to use case 'type', further consideration could be given to the context or the 'purpose' of the use case and its social utility. This might require a re-shaping of the current risk-based approach to one that is both risk-based and purpose-driven. This may also be helpful in the absence of a formal bill of rights in Australia, which exist in the other jurisdictions that are currently undertaking pro-active AI regulatory reforms.

40. One source of guidance in this regard is Annex III of the EU AI Act, which sets out a list of high-risk AI systems, referred to in Article 6.2. However, Article 6.3 sets out exceptions to the blanket approach in Article 6.2:

> *… an AI system referred to in Annex III shall not be considered to be high-risk where it does not pose a significant risk of harm to the health, safety or fundamental rights of natural persons, including by not materially influencing the outcome of decision making.*[18]

41. Article 6.3 provides the approach required to meet the exception. Article 6.4 requires that a provider seeking to use the exception shall "document its assessment before that system is placed on the market or put into service".

42. Examples of how such an approach could apply against identified risk domains are outlined below.

[17] Law Council of Australia, Submission to the Department of Infrastructure, Transport, Regional Development, Communication, and the Arts, *A new legislative framework to address and protect Indigenous cultural and intellectual property rights* (5 July 2024).
[18] EU AI Act, Article 6.2.

Biometrics

43. In the Proposals Paper, biometrics is listed as a domain.[19] Its high-risk use case description is an AI system "used to identify or categorise individuals, assess behaviour or mental state, or monitor and influence emotions".[20]

44. To help inform the overall assessment of 'risk' in the biometrics domain, identifying the outcome to be achieved, and, by extension, its social utility might also be highly beneficial. The question could be asked: is the purpose behind using AI systems processing higher-risk biometric information to diagnose medical conditions, to increase retail sales, to predict criminality, to screen for employment prospects, or to assess health insurance eligibility?

45. Articulating the purpose (or the 'end game') sought to be achieved could be useful in establishing not only whether—and to what extent—any proposed mandatory guardrails apply, but whether the ultimate purpose justifies those higher risks, notwithstanding compliance with the guardrails.

46. For example, there may be limited social utility in a high-risk biometrics use case where the outcome is to socially 'score' individuals for employment or for eligibility to rent or purchase property in certain areas, or to qualify for organ donation. A purpose-driven approach might require consistency with, or be complementary to, express and implied rights in the Commonwealth Constitution and other Australian laws—and representative democracy.

47. We note that the exception provided in Article 6.3 of the EU AI Act would exempt some applications of biometrics, despite biometric AI systems being defined as high risk. For example, a medical diagnosis would potentially fall into the exemption by meeting Article 6.3(a): "the AI system is intended to perform a narrow procedural task".

Administration of justice and democratic processes

48. A further example is in the domain of the administration of justice and democratic processes.[21] In the Proposals Paper, its high-risk use case description is as follows:

> *AI systems used for making a determination about an individual in a court or administrative tribunal, such as systems used for evaluating facts, evidence and submission to proceedings."*[22]

49. Concerns arise in this domain partly because the law is not reliant on approximation or a synthesis of multiple sources.[23] As such, in common law jurisdictions, AI's information-synthesis approach to case law may not identify accurately legal precedent or a case's *ratio decidendi* for the purposes of applying precedent.

50. Accuracy is also critical to the democratic process and the ability to exercise democratic voting rights based on accurate information. Those democratic rights and processes are clearly compromised by misinformation and disinformation.

---

[19] Proposals Paper, 26 61.
[20] Ibid 26.
[21] Ibid 26, 63.
[22] Ibid 26.
[23] Research itself has traditionally been an important 'tool of trade' for lawyers. It is also notable that lawyers' use of 'search' engines, like Google, in their research has provided multiple possible answers to assess rather than an attempt at 'the' answer, resulting from a culmination of sources that can be unknown and undiscoverable.

51. Concerns further arise in the law and, by extension, the legal profession, because laws are both ancient and living. It would be difficult to use tools that have 'cut-off dates' for training data and expect them to deliver 'justice'. It is critical that any reliance on GPAI in assisting to dispense the law is appropriate and proportionate.

52. Other important aspects in the context of the administration of justice use cases are that the law:

    - is language-based;
    - often contains a mental or cognitive element in order to apply it; and
    - is often open to interpretation or nuance.

53. Rather than being codifiable or reduced to logical 'if this, then that' rules, the Australian legal system recognises the need for flexibility when applying the law to the facts of a specific case. We acknowledge, however, that this is not the case in all matters: for example, where Parliaments have passed strict liability laws.

54. Nonetheless, in most cases, the law is applied by practitioners and judges using a process of interpretation and the exercise of professional judgment. The application of the law is then relied on by clients, parties in legal proceedings and, ultimately, the broader community. If the application of the law is incorrect, then lawyers and judges are answerable to the participants in the legal system.

55. In 1881, Oliver Wendell Holmes observed that 'the life of the law has not been logic: it has been experience'.[24] Not every experience, even today, is captured or represented in our digital lives by data or a digital footprint. Another way to see the law is 'open-textured', as Professor Simon Chesterman notes; it is why a rule such as 'no vehicles in the park'—an example first offered by H.L.A. Hart[25]—is open-textured enough to extend to cars, trucks, bicycles, and even drones, but perhaps not toy trucks or toy trains.[26] This demonstrates how the law often requires interpretation and experience.

56. Defining high-risk use cases in the administration of justice as those making determinations about an individual may benefit from the further assessment of the purpose or social utility underlying the use case. For instance, if the purpose (or outcome) is to ease the burden on courts and increase efficiency in administering the law, does the margin for error or false convictions that using AI may bring, justify its use? Or should a social utility dimension also require that the efficient administration of justice must not result in denying any person justice?

57. We consider that the advantage of a risk-based, purpose-driven approach could help ensure that high-risk uses are limited to those having social utility (by reference to consistency with, or being complementary to existing Australian law). This approach could further minimise any unintended, unforeseeable harms or outcomes that are inconsistent with the rule of law and our democratic processes.

[24] Anonymous (Oliver Holmes), 'Book Notices' (1880) 14 *American Law Review* 233, 234.
[25] H.L.A. Hart, *Positivism and the Separation of Law and Morals*, 71 HARV. L. REV. 593, 607 (1958), cited in Frederick Schauer, "A critical guide to vehicles in the park" New York University Law Review, 83:1109, 2008.
[26] Simon Chesterman, 'We, the Robots: Regulating Artificial Intelligence and the Limits of the Law' (NUS Law Working Paper No 2022/015, Cambridge University Press, 2021).

Categories of use to be designated 'high-risk'

58. If a general principles-based approach is to be retained, we suggest that certain categories of use should also be designed 'high-risk'. Such designation would have the effect of imposing a requirement that such uses comply with the proposed mandatory guardrails.

59. This approach would ensure greater clarity for organisations developing and deploying AI, and enhanced transparency for users, whilst ensuring that novel applications are appropriately regulated.

60. We acknowledge work that is already underway at the Commonwealth level to reform the Privacy Act, combat misinformation and misinformation, review the *Online Safety Act 2021* (Cth), and develop a consistent legal framework for automated decision-making by government. In light of this work, we submit that consideration should be given to designating the following categories as high-risk:

    (a) Uses of AI systems to make decisions impacting people's health, safety and human rights, or which have an impact on an individual's legal rights or entitlements, or a similarly significant effect on an individual.

    (b) Uses of AI involving the input of sensitive data in medical and health environments (and in other industries involving sensitive information processing).

    (c) Uses of AI systems to generate or replicate personhood or aspects of personhood (i.e., appearances, voices, etc.) for any identifiable individuals, especially minors.

    (d) Uses of AI-generated content in political campaigns.

61. Regardless, we suggest that the Department seek expert guidance from independent and appropriately qualified individuals as to the risk profile of categories of AI use.

Options for introducing mandatory guardrails

62. The Proposal Paper poses three options for implementing the proposed mandatory guardrails, namely:

- **Option 1:** Adapting existing regulatory frameworks to include the proposed mandatory guardrails;[27]
- **Option 2:** Introducing framework legislation, with associated amendments to existing legislation;[28] or
- **Option 3:** A whole-of-economy approach through the introduction of a new AI-specific Act.[29]

[27] Proposals Paper, 46-48.
[28] Ibid 48-49.
[29] Ibid 49-51.

General comments

63. The Law Council has previously indicated a preference for a 'multi-faceted' approach to the regulation of AI based on the following elements:

    - an initial focus on the regulation of high-risk AI technology and applications;
    - a strengthened and precautionary approach to AI regulation, where there is evidence that existing laws and regulations are insufficient to address the issues and harms arising;
    - comprehensive regulatory reform to ensure that the use of automated decision making, including by the Australian Government, is transparent, capable of review, and consistent with administrative law principles; and
    - consideration of the regulatory models adopted by other jurisdictions to determine an optimal and bespoke approach for Australia that reflects the nuances of Australia's pre-existing constitutional and regulatory framework, and unique local market environment.[30]

64. We have previously expressed support for the expansion of current legislation and, where necessary, new targeted legislation, rather than seeking to explicitly regulate AI via a comprehensive 'AI Act'.

65. The majority of responses we received in response to this consultation favour an approach that minimises the potential for conflict—or duplication—with existing laws.

Views on the proposed options

66. When consulting on the Proposals Paper, it was apparent that there are differing views amongst the legal profession as to the appropriateness of each option, particularly the merits of a new AI-specific Act, as proposed in Option 3.

67. For example, the Law Institute of Victoria and Queensland Law Society have expressed a preference for Option 3, supporting the approach taken by the EU AI Act, which presents a clear articulation of what uses of AI pose the greatest risk, with a consequent regulatory focus in those areas. These bodies submit that Option 3 would best address the use of AI in high-risk settings, and would allow for additional reforms to be undertaken, by way of legislative amendment, as AI development and use evolves over time.

68. We acknowledge that Option 3 may 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.

69. Other contributors, however, have raised concerns that the Option 3 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. We also recognise that the implementation of this option may not be feasible in the immediate short-term.

[30] See, e.g, Law Council of Australia, Submission to Department of Industry, Science and Resources, *Safe and Responsible AI in Australia* (17 August 2023) 11; Submission to Senate Committee on Adopting Artificial intelligence, *Inquiry into the opportunities and impacts of the uptake of artificial intelligence technologies in Australia* (20 May 2024).

70. On the other hand, Option 1 appears to be beneficial for reducing the risk of inconsistency or duplication across regulatory frameworks. However, it is likely to be a more time-consuming project, given the need to review and amend existing laws to embed the guardrails. Further, this domain-specific approach could lead to inconsistent regulatory outcomes and enforcement mechanisms.

71. We observe that Option 2 provides an appropriate middle-ground approach between Options 1 and 3, by creating a new framework legislation that mitigates risks, while allowing AI development. In addition, this option could potentially be implemented as an interim measure until new legislation prescribing specific uniform requirements is introduced.

72. On balance, we do not consider a comprehensive AI-specific Act is required at this early stage. However, the gradual introduction of regulations that address high-risk AI and applications, having social utility, will be helpful in minimising risks, both known and unknown.

73. We consider that the areas in which regulation is likely to be introduced earliest are the areas regulated by the:

    - Australian Competition and Consumer Commission;
    - Australian Communications and Media Authority;
    - eSafety Commissioner; and
    - Office of the Australian Information Commissioner.

74. These four regulators have a mechanism for collaborating on digital platform regulations called the Digital Platform Regulators Forum (**DP-REG**). The DP-REG is welcome initiative that is likely to facilitate the effective introduction of domain-specific solutions.

Proposed mandatory guardrails

General comments

75. The Proposals Paper sets out proposed mandatory guardrails for AI systems used in high-risk settings, as well as high-risk GPAI systems.[31] The guardrails are preventative measures that would require developers and deployers of AI used in high-risk settings to take specific steps across the AI lifecycle.[32]

76. It is acknowledged that nine of the 10 proposed mandatory guardrails is identical to the Voluntary AI Safety Standard. The Voluntary AI Safety Standard has been mapped against Australia's AI Ethics Principles and broadly aligns with relevant existing frameworks[33] (such as the ISO 42001:2023 and the US National Institute of Standards and Technology's Artificial Intelligence Risk Management Framework).

77. We agree that the guardrails should apply to all GPAI models in Australia. As GPAI models may exceed even what are currently perceived as their potential uses, there should be a willingness on the part of the government to modify the regulatory approach as the technology becomes more sophisticated.

---

[31] Proposals Paper, 35-42.
[32] Ibid 19.
[33] See, e.g., the ISO/IEC 42001:2023 (Information Technology – Artificial Intelligence – Management system); US National Institute of Standards and Technology, *Artificial Intelligence Risk Management Framework* (January 2023).

78. As the mandatory guardrails are broadly framed, we suggest it would be beneficial to develop a more detailed set of minimum standards, derived from these guardrails. Organisations developing or deploying AI systems should subsequently be required to adopt these more detailed standards, unless they establish their own standards that are compliant with the overarching guardrails.

79. For example, in respect of guardrail 3 ("protect AI systems, and implement data governance measures to manage data quality and provenance"), we recommend that developers and deployers should be more specifically obliged to establish and implement measures that sufficiently protect the data that is involved in the development of AI systems, including personally identifiable information of any individuals. We consider that it would be preferable if, unless an organisation developed its own standards to comply with guardrail 3, that organisation would be required to comply with the government-developed minimum standards, including one requiring the establishment and development of measures to protect personal data.

Intellectual property enforcement

80. From an intellectual property enforcement perspective, proposed transparency in respect of the use and operation of AI (guardrail 7) will be of critical importance, as will establishing processes for people impacted (guardrail 8). The nature and accessibility of these processes will also be important to ensure that people or groups who are impacted can access information and exercise their rights.

81. Specific issues arise in relation to transparency in the context of intellectual property rights. For example, there is a question whether a transparency obligation may require the disclosure of information concerning whether—and the extent to which—copyright protected materials may have been used in the creation of an AI system (for example, in training it). That is a question requiring detailed consideration of policy and practical matters, and we note that it is currently being considered by the Copyright and Artificial Intelligence Reference Group.

82. Further, questions arise as to:

    - how any transparency obligation will avoid the risk of inappropriately requiring the disclosure of valuable trade secrets; and
    - whether intellectual property rights holders should have a means of enforcing the transparency obligation.

83. Consideration could particularly be given to whether it is necessary to have an external means of reviewing the outcome of such a process. Currently, in order to enforce legal rights, individuals would most likely be required to resort to litigation, which is unlikely to be an efficient or affordable option.

84. These matters illustrate the necessity for the proposed *ex ante* guardrails to be accompanied by reforms to intellectual property law reform and other forms of specific guidance.

85. From a copyright perspective, we suggest that consideration be given to what remedies might be appropriate for an individual or group whose legal rights are impacted. For example, where a copyright work has been used as an input to train AI, or accessed in the process of answering a query, what remedy should be available to the copyright owner? We note that these questions are currently being explored in litigation in other jurisdictions. However, those individual matters will not provide clear answers for Australian copyright owners.

86. We also note the ongoing wider discourse about whether copyright infringement issues associated with AI systems might be addressed, or mitigated, by some form of licensing framework, statutory exceptions or defences to infringement, or some form of *sui generis* legislation.

Protections for domestic end-users

87. Recent commentary has observed that, even if the mandatory guardrails have extraterritorial effect—recognising that many AI developers are based outside of Australia—their efficacy in protecting Australian deployers and end-users may be low.[34]

88. We suggest that further clarity is needed about the protections that would be in force for end users within Australia under the current proposals. Whilst enforcement will be critical, so too will be ensuring that the legislative framework includes sufficient protections and avenues to pursue overseas developers or deployers.

Specific comments

89. We have received the following specific comments from our membership about each of the proposed mandatory guardrails.

Proposed mandatory guardrail	Comments
1. Establish, implement and publish an accountability process including governance, internal capability and a strategy for regulatory compliance.	The process for accountability needs to be a best practice process.
2. Establish and implement a risk management process to identify and mitigate risks.	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 that 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 that examples of risk management processes be provided.
3. Protect AI systems, and implement data governance measures to manage data quality and provenance.	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.	The testing of AI models and AI systems on an ongoing basis needs to be performed by persons with accreditation in the subject matter.

[34] Lisa M. Given, 18 January 2024, *Australia plans to regulate 'high-risk' AI. Here's how to do that successfully*, The Conversation (Online, 18 January 2024).

Proposed mandatory guardrail	Comments
5. Enable human control or intervention in an AI system to achieve meaningful human oversight.	This guardrail is supported.
6. Inform end-users regarding AI-enabled decisions, interactions with AI and AI-generated content.	This guardrail is supported. Given the recent introduction of the Privacy and Other Legislation Amendment Bill 2024 (Cth), 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 where 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.	Dispute processes must be accessible and responsive. However, members of our Business Law Section's Intellectual Property Committee, with experience engaging in the complaint processes for some global entities, have reported that a complaint about a breach of legal rights, or a breach of terms of use, does not necessarily get answered. As such, processes should include clear guidelines for any challenge of the use or outcomes of AI systems to be dealt with—and resolved—within a reasonable timeframe, 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. Other issues that could be considered to support transparency include: • how the regulatory framework can respond to circumstances where the use of AI is not known. This will include continued exploration of 'marking' techniques to identify AI-generated outputs. This would assist in clarifying for organisations and users the extent to which a product has an AI component; • where its use is difficult to contract out, the interaction between AI use and the unfair contract terms regime; and • requiring developers and distributors to be transparent about data providence and copyright works used, and how to implement options for opting out of machine learning.
9. Keep and maintain records to allow third parties to assess compliance with guardrails.	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).

Proposed mandatory guardrail	Comments
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 Guardrails 1 and 9. Conformity assessments should be carried out by an accredited third party, or by government entities or regulators, rather than the organisations themselves. The accountability and quality assurance metrics should be standardised.

Other comments

Environmental harm

90. The Proposals Paper observes that the precautionary principle is usually applied where a technology has significant uncertainty about its risks, including environmental harms. Nonetheless, we are concerned that environmental risks may not be sufficiently addressed in the Proposals Paper.

91. We submit that this should be a core issue for consideration. For example, environmental protection is a core value put forward in the EU AI Act with the energy consumption of AI models a central consideration.[35] We support the inclusion of enhanced environmental impact considerations in the Australian Government's regulatory response (whatever form that takes).

Oversight

92. Compounding the abovementioned concerns regarding uncertainties with a principles-based approach are issues of ensuring compliance and enforcement.

93. It has been suggested to us that a mechanism of licensing is most appropriate. We note that there is some support within our membership for an independent oversight body that can consider testing and scientific results in a transparent way (at development and deployment levels). For example, we note the Proposals Paper references Canada's requirements 'for developers to prepare and share model cards for machine learning models'.[36] Model cards aim to provide information about the trained machine learning model, and can include evaluation data, training data, performance metrics, ethical considerations, and any relevant model-specific details.[37]

94. An oversight body with visibility of such information would support regulatory certainty and ensure that, where an AI tool has a particular impact, then appropriate enforcement and penalties regimes are applied.

[35] Sasha Luccioni et al, *The Environmental Impacts of AI*—Policy Primer (Report, 2024) 8.
[36] Proposals Paper, 41.
[37] Ibid.

Use of AI to replicate personhood

95. As a related matter, our Business Law Section's Intellectual Property Committee recommends that consideration be given to whether Australia needs specific laws to address the use of an individual's voice, or image, without consent, in the context of an AI system.[38] These may be used, for example, in a political context,[39] or in an arts and entertainment context (e.g., the use of a deepfake of a performer's voice).

96. As Australia does not have a broad right of privacy that is attached to personal image, or rights of publicity, an individual who is affected by AI system outputs of this kind would currently be required to rely on laws relating to:

    (a) copyright, although this may be of limited utility, depending on whether an original copyright work can be identified as having been incorporated, and then only if the affected individual is the owner of that copyright work;

    (b) misleading and deceptive conduct, but principally if the individual's voice or image is used in certain commercial contexts; or

    (c) defamation, noting that this may be difficult to prove, depending on the use.

97. If there is no amendment to Australian law, an individual who is affected by an AI system that generates this type of output may not be able to point to an 'adverse legal effect' (as required under proposed principle (c)).

[38] The recent *Criminal Code Amendment (Deepfake Sexual Material) Act 2024* (Cth) addresses similar issues, with respect to 'deepfakes' and pornographic content.
[39] See, e.g., Rachel Looker, *Trump falsely implies Taylor Swift endorses him*, BBC (Online, 20 August 2024).