KING&WOOD MALLESONS
金杜律师事务所

Level 27
Collins Arch
447 Collins Street
Melbourne VIC 3000
Australia

T +61 3 9643 4000
F +61 3 9643 5999

www.kwm.com

TO Technology Strategy Branch
The Department of Industry, Science and Resources
Commonwealth of Australia

28 JULY 2023

Submission to Discussion Paper: Safe and Responsible AI in Australia

We are pleased to provide our submission to the Department's Discussion Paper on Safe and Responsible AI in Australia.

AI presents significant opportunities for Australian organisations and the broader Australian economy. However, while there may be some risks associated with AI, any AI related regulation must be clearly targeted at, and proportionate with, identifiable and serious risks to individuals, society or the environment.

This will involve taking into account the varying contexts in which AI systems can be deployed throughout the Australian economy and the need to support and promote innovation by Australian companies.

Yours sincerely

Cheng Lim | Partner
Sector Leader Technology, Media, Entertainment + Telecoms
King & Wood Mallesons

T +61 3 9643 4193
M +61 419 357 172
F +61 3 9643 5999
E cheng.lim@au.kwm.com

Bryony Evans | Partner
King & Wood Mallesons

T +61 2 9296 2565
M +61 428 610 023
F +61 2 9296 3999
E bryony.evans@au.kwm.com

Kim O'Connell | Partner
King & Wood Mallesons

T +61 2 9296 2188
M +61 413 618 299
F +61 2 9296 3999
E kim.oconnell@au.kwm.com

SAFE AND RESPONSIBLE AI IN AUSTRALIA DISCUSSION PAPER: KWM SUBMISSION

King & Wood Mallesons welcomes the opportunity to provide feedback on the "*Safe and responsible AI in Australia: Discussion Paper*" (**Discussion Paper**). As a leading Australian law firm, we have been working closely with Australian and international companies over the past few years to address legal and ethical risks presented by the development and deployment of AI systems. Our work has included the development of bespoke AI governance frameworks, the design and execution of internal AI impact assessments and the design and execution of contractual arrangements to operationalise AI systems.

1 Summary of KWM's submission

We think it is crucial for the Commonwealth to exercise caution and refrain from hastily implementing regulation of AI in response to the hype and panic that we have seen during the last few months as companies (and the media) rush to capitalise on the uptake of generative AI (both in Australia and overseas).

While there may be some risks associated with AI, any AI related regulation must be clearly targeted at, and proportionate with, identifiable and serious risks to individuals, society or the environment. This will involve taking into account the varying contexts in which AI systems can be deployed throughout the Australian economy and the need to support and promote innovation by Australian companies.

It will also be important for the Commonwealth to minimise regulation that would impose regulatory burdens unique to Australia. Minimising such regulation will allow Australian businesses to rapidly adopt AI technologies that are available overseas (making them more efficient domestically and more competitive globally) and will ensure that Australian organisations and consumers have access to global advances in AI systems that will be of benefit to them.

In summary, we consider that the Commonwealth should:

1. Focus its initial energies on reviewing how existing Commonwealth and State legislation already addresses specific risks and harms presented by AI and, if there are gaps, consider how that legislation can be amended to address those risks and harms.
2. Only consider introducing new horizontal (sector-wide) legislation if identified harms presented by AI cannot be adequately addressed in existing legislative or regulatory regimes (or amendments to them).
3. If new horizontal AI legislation is to be introduced, ensure that it is principles-based legislation that implements a set of AI Governance Principles. We suggest an approach similar to the way in which the *Privacy Act 1988 (Cth)* (**Privacy Act**) implements the Australian Privacy Principles. Adopting a principle-based approach will be sector neutral and enable the legislation to be flexible to changes in underlying technology. Such legislation should be combined with specific guidance from an appropriate regulator where that specific use cases for AI would benefit from more details in relation to how the AI Governance Principles apply. The AI Governance Principles should:

(a) promote the establishment of internal mechanisms for organisations to identify whether they are deploying an AI system that will (or is likely to) involve a serious risk of harm to individuals, society or the environment;

(b) require that organisations undertake an AI Impact Assessment against the Australian AI Ethics Principles if they plan to deploy an AI system that will (or is likely to) present a serious risk of harm to individuals, society or the environment; and

(c) require organisations take reasonable steps to mitigate the risks of serious harm to individuals, society or the environment arising from their deployment of any AI system.

4 Co-ordinate approaches to the regulation of AI through the establishment of a new agency or regulator (or designating an existing agency or regulator) as a centralised point for regulating, coordinating and sharing expertise with other regulators responsible for sector-specific and domain-specific legislation relevant to AI technology.

5 Not introduce regulation that could inhibit the development and adoption of low risk, open-source AI models in Australia.

We have explored this proposal in greater detail in our responses to the specific questions below. Ultimately, any regulation would need to promote responsible AI development and usage, and foster collaboration and innovation in the field while avoiding undue regulatory burdens.

2 Question 2: Potential gaps in approaches

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

2.1 Although Australia does not have AI-specific regulation, AI is not unregulated in Australia. Rather:

(a) existing legislation already addresses many of the risks associated with the use of AI. For example:

(i) the Privacy Act 1988 (Cth) imposes guardrails on how APP Entities may utilise personal information throughout the lifecycle of an AI system. Upcoming amendments to the Privacy Act (as foreshadowed in the Attorney-General's recent review of the Privacy Act)1 may also impact the use of automated decision-making and other uses of personal information in AI systems;

(ii) the *Copyright Act 1968* (Cth) gives rights holders some protection against use of their copyrighted materials to train AI systems and determines when, if ever, copyright attaches to the outputs generated by AI systems (recognising there is no express provision for computer-generated works to be protected by copyright in their own right currently in Australia);

[1] Attorney-General's Department, *Privacy Act Review, Report 2022* (Final Report, 16 February 2023) https://www.ag.gov.au/sites/default/files/2023-02/privacy-act-review-report_0.pdf>.

(iii) various discrimination Acts[2] regulate decisions made using AI systems that impact individuals with protected attributes;

(iv) the *Competition and Consumer Act 2010* (Cth) imposes restrictions (such as in relation to unconscionable conduct, product liability, misleading and deceptive conduct and consumer guarantees) on the outputs of AI systems marketed and sold to individuals and small businesses; and

(v) section 180 of the *Corporations Act 2001* (Cth) imposes a general duty on directors to ensure there are effective risk management and compliance systems in place to address material risks to their organisations – which would include material risks arising out of their use of AI; and

(b) (although not mandatory) Australia's AI Ethics Principles[3] provide voluntary principles designed to ensure AI is safe, secure and reliable.

2.2 Nonetheless, it is possible that there are some gaps in this existing regulation. There are at least two gaps that we consider warrant the Commonwealth's attention:

(a) the possibility that certain risks will fall between the various regulatory remits of existing legislation; and

(b) a lack of guardrails and guidance on how companies can identify and mitigate the risks presented by their proposed uses of AI systems. This gap has been recently highlighted by the Human Technology Institute at the University of Technology Sydney, whose research found that few Australian organisations have implemented systemic and structured forms of governance around AI systems.[4]

2.3 We suggest these gaps should be addressed in the following ways:

(a) the Commonwealth should undertake a review of existing Commonwealth and State legislation to comprehensively map the existing restrictions and obligations that affect AI systems. This will assist in establishing a common body of knowledge for companies implementing AI systems and for the Commonwealth to identify gaps that currently exist or will likely exist in the future. Such work should also take into account current legislative reform processes (including the current Australian Cybersecurity Strategy 2023-2030 review and the Privacy Act Review);

(b) if the Commonwealth identifies harms arising from the deployment and use of AI systems that are not adequately addressed in existing legislation, the Commonwealth should first consider amending existing legislation rather than introducing new horizontal (sector-wide) legislation. Only if these harms cannot be adequately addressed in the existing regulatory regime should the Commonwealth introduce horizontal AI legislation to specifically address those harms;

(c) if horizontal AI legislation is to be introduced, it should be:

(i) principles-based legislation that implements a set of AI Governance Principles similar to the way in which the Privacy Act implements the Australian Privacy Principles (supported by detailed guidance on specific areas where necessary);

---

[2] This includes the *Age Discrimination Act 2004* (Cth); *Disability Discrimination Act 1992* (Cth); *Racial Discrimination Act 1975* (Cth); and the *Sex Discrimination Act 1984* (Cth).

[3] Australian Government, Department of Industry, Science and Resources, 'Australia' AI Ethics Principles', *Australia's Artificial Intelligence Ethics Framework* (Web Page) <https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles>.

[4] Lauren Solomon and Nicholas Davis, 'The State of AI Governance in Australia' (Human Technology Institute, University of Technology Sydney, 31 May 2023) 28 <https://www.uts.edu.au/sites/default/files/2023-05/HTI%20The%20State%20of%20AI%20Governance%20in%20Australia%20-%2031%20May%202023.pdf> (*'HTI UTS'*).

KING&WOOD MALLESONS

(ii) supported by either a new regulator or (subject to an expansion in regulatory ambit) an existing regulator; and

(iii) complemented and supported by guidance and template documents developed by the regulator to support how companies implement the AI Governance Principles. This should include, but not be limited to, template AI Impact Assessments and guidance on how companies should approach the AI Governance Principles for specific use cases. Such guidance will be crucial to promoting confidence in how companies should approach the adoption of high risk AI systems while not stifling innovation. We explore this further in the risk section of our submission below; and

(iv) to the extent that there are concerns of a regulatory patchwork emerging as a result of having a sector-specific regulatory approach with governance focussed horizontal legislation, establish a formal mechanism (such as a cross-regulatory forum) that promotes and facilitates co-ordination across the economy.

3 Question 9: Transparency

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

3.1 The principle of transparency is a lynchpin of many AI ethics frameworks, including Australia's AI Ethics Principles.5 Nonetheless, while the principle of transparency is uncontroversial at a high level, we consider that the Commonwealth should avoid mandating transparency requirements without regard to:

(a) how these requirements will operate in practice across the different use cases in relation to Al; and

(b) the existing technical limitations of AI technology.

3.2 In particular, the Commonwealth should take into account the following:

(a) 'transparency' as it relates to AI systems can refer to a number of closely related, but different, concepts. These include transparency about when an AI system is being used, how the system works, what data the system was trained on and how it was trained, why an AI system produced a particular output (often known as 'interpretability')[6] and so on. Each of these different concepts warrant their own consideration;

(b) what constitutes an appropriate level of transparency is highly context specific. For example, there may be some cases in which it is appropriate and desirable to require disclosure to a natural person that they are interacting with an AI technology (e.g. a direct interaction with a public-facing voice system or chatbot), but other cases where this requirement may be unnecessary (e.g. image generation tools). Similarly, in the

[5] Ibid 3.

[6] See, for example, the National Institute of Standards and Technology, 'Artificial Intelligence Risk Management Framework' (2023) *AI RMF 1.0* <https://doi.org/10.6028/NIST.AI.100-1> ('***NIST AI Risk Management Framework***').

context of interpretability, there will be some use cases where interpretability of an output may be of greater importance in light of the uses and impacts of that output (e.g. if the output is used as the basis for a significant decision that impacts an individual or in a medical context). By contrast, interpretability will be less important in some other contexts (e.g. AI chatbots designed for entertainment purposes).[7] Accordingly, careful consideration should be given to how any government mandated transparency requirements will operate in practice, particularly as AI technology becomes commonplace across all software and products containing software. This may include consideration of who should bear the primary burden of complying with disclosure requirements (e.g. the developer or the deployer / user of the technology); and

(c) any mandated transparency requirements will need to be cognisant of the technical limitations of the technology itself, and flexible enough to adapt as that technology evolves. One key challenge in implementing transparency requirements is that neural networks (which underpin many recent advances in AI technology) are generally not considered to be 'inherently interpretable'. Although there are numerous post-hoc methods that can be used in an attempt to interpret these models, this is an emerging field and these post-hoc methods currently have many limitations. For example, different post-hoc methods can disagree with each other on interpretations,[8] suffer from bias and can be vulnerable adversarial attacks.[9] Furthermore, improving interpretability can also involve trade-offs (including with privacy and accuracy), which means that maximising interpretability is not always appropriate in a particular context.[10]

3.3 As a result, one key risk in mandating transparency requirements is that it could have the effect of inhibiting the adoption, and in some cases prohibiting or restricting certain uses, of AI technology. While this might be an acceptable outcome in specific use cases (after appropriate consideration has been given to the costs and benefits of doing so), we recommend that the Commonwealth should avoid mandating general transparency requirements which, while well-intentioned on their face, may unintentionally inhibit the use of certain AI technology in existing, or yet to emerge, use cases.

3.4 In light of the above, we recommend that the Commonwealth:

(a) avoid mandating transparency requirements without first carefully considering how they can be implemented in practice;

(b) consider dealing with any detailed transparency requirements through sector-specific regulation or guidance (and only after appropriate consideration is given to the costs and benefits); and

(c) ensure any regulation on transparency is supplemented by sufficient guidance to enable developers and deployers to understand what is required in specific scenarios.

[7] Alistair Reid, Simon O'Callaghan and Yaya Lu, 'Implementing Australia's AI Ethics Principles: A selection of Responsible AI practices and resources' (Gradient Institute and CSIRO National Artificial Intelligence Centre, June 2023) 19 <https://www.csiro.au/-/media/D61/NAIC/Gradient-Report/23-00122_DATA61_REPORT_NAIC-ResponsibleAITools_WEB_230620.pdf.

[8] Satyapriya Krishna et al, 'The Disagreement Problem in Explainable Machine Learning: A Practitioner's Perspective' (2022) <https://arxiv.org/pdf/2202.01602.pdf>.

[9] Dylan Slack et al, 'Fooling LIME and SHAP: Adversarial Attacks on Post hoc Explanation Methods' (2020) <https://arxiv.org/pdf/1911.02508.pdf>.

[10] *NIST AI Risk Management Framework* (n 6) 12.

4 Question 10: Prohibitions

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

4.1 We recommend that the Commonwealth exercise caution before introducing any legislation that prohibits certain uses of AI technology by the private sector.

4.2 There are legitimate concerns about some potential uses of AI technology. In some cases, carefully considered and narrowly defined prohibitions may end up being the appropriate course of action to protect both individuals and society from particular harms, while increasing public confidence in AI.

4.3 However, outright prohibition of certain AI technology or use cases through legislation is an imprecise regulatory approach which raises several challenges:

(a) as discussed in our response to questions 14 and 15 below, the appropriateness of many use cases can be context-dependent, with different considerations applicable in each case. Banning the use of a technology outright may preclude deployment of safe and sensible applications of AI, or applications of AI that can be made acceptable with appropriate safeguards;

(b) in some cases, the risks of these use cases are yet to emerge in Australia, and it will be difficult to draw the line between appropriate and inappropriate use cases in circumstances where the technology cannot be tested, applied and monitored;

(c) as legislation takes time to amend, there are significant risks associated with creating new prohibitions instead of relying on existing legal mechanisms to prevent harm.

4.4 The risk of overregulation is particularly acute with emerging technologies like AI in which the dystopian possibilities may be front of mind (and are constant tropes of popular fiction), but the potentially beneficial use cases may be less well-known or yet to emerge. A rush to regulate in fear of hypothetical use cases (e.g., widespread social scoring in Australia) may have the unintentional consequence of both inhibiting the adoption of actual beneficial use cases and increasing the uncertainty and public distrust of AI.

4.5 We acknowledge that the European Parliament's proposed AI Act seeks to prohibit a range of harmful AI practices.[11] However, in addition to the issues highlighted above, it is important to note that these prohibitions are not easily adapted to an Australian context. For instance:

(a) the prohibited list has been compiled on the basis that the prohibited AI practises "go against the EU values of democracy, freedom and human dignity, and violate

[11] At a broad level, the current list of prohibited practises (as modified by the European Parliament) covers AI systems: that deploy subliminal techniques beyond a person's consciousness or purposefully manipulative or deceptive techniques causing them to take a decision they otherwise would not have; that exploit any of the vulnerabilities of a person or a specific group of persons (e.g. age/personality traits) with the aim to materially distort behaviour; that are biometric categorisation systems that categorise based on sensitive or protected attributes/characteristics (or the inference of those attributes/characteristics); for social scoring that leads to detrimental or unfavourable treatment (a) of persons/groups in social contexts that are unrelated to the contexts in which the data was originally generated or collected, and/or (b) unjustified or disproportionate to their social behaviour or its gravity; AI systems that can be used to assess offending or reoffending or for predicting the occurrence or reoccurrence of an actual or potential offences based on profiling or assessing personality traits/characteristics; that create or expand facial recognition databases through the untargeted scraping of facial images from the internet or CCTV footage; or that infer emotions of a natural person in the areas of law enforcement, border management, in workplace and education institutions.

fundamental rights, including privacy and consumer protection”.[12] Such an approach cannot be easily adapted to Australia given we do not have a comparable clearly articulated legislative approach to human rights; and

(b) the European Council has acknowledged that some of the prohibitions fit within the context of more focused regulation (e.g., the prohibition of manipulative practice could be integrated into the Unfair Commercial Practice Directive and the prohibitions of the general-purpose social scoring could be included in General Data Protection Regulation).[13]

4.6 Before legislating any prohibitions, we recommend that the Commonwealth:

(a) consider whether the harm that the prohibition is seeking to address is or can be mitigated, or reduced, via existing laws (e.g., concerns about use of AI to manipulate individuals might already be adequately captured under Australian Consumer Laws and torts law);

(b) to the extent the harm is not mitigated by existing laws or other measures, consider whether it is more appropriate to implement a prohibition on a sector-specific basis; and

(c) undertake consultation with impacted parties and a detailed cost-benefit analysis to ensure that any specific prohibitions are proportionate and appropriately adapted to address particular targeted harms, without undermining legitimate use cases. As an example of this, we draw the Commonwealth’s attention to the UTS Facial Recognition Technology Model Law that demonstrates the different considerations that apply in a single use case.[14]

4.7 Finally, if a particular use case is to be prohibited:

(a) the prohibition should be narrowly defined and carefully drafted to ensure it captures only the most objectionable uses cases (for which the risks cannot be mitigated by other means) and that uncertainty in the application of the prohibition does not discourage innovation in use cases outside the intended scope of the prohibition.

(b) Aside from prohibitions implemented through legislation, prohibitions and bans can also be imposed through means such as policies adopted by public sector entities or by voluntarily codes adopted in particular industries or sectors. Although these measures also face some of the issues outlined above, we consider that these measures should be considered as an alternative to legislated prohibitions given that they can be quickly imposed, modified or removed and can be more targeted - and so are less likely to have a negative impact on innovation.

5 Questions 14 and 15: Risk-based approaches

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

[12] European Commission, Commission Staff, *Impact Assessment accompanying the Proposal for a Regulation of the European Parliament and of the Council, Laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain union legislative acts* (Doc No SWD(2021) 84 final, 21 April 2021) 46 <https://parliament.bg/pub/ECD/4327901_EN_impact_assessment_part1_v7.docx> (‘***EU Commission Impact Assessment***’).

[13] Ibid 46 n 219.

[14] *HTI UTS* (n 4).

KING&WOOD MALLESONS

[15] What do you see as the main benefits or limitations of a risk-based approach? How can any limitations be overcome?

5.1 We broadly support the adoption of "risk-based" approaches in AI regulation. It is, however, important to clarify what "risk-based" means in this context because, as Professor Cary Coglianese of University of Pennsylvania Law School has noted:

> *"The widespread enthusiasm for risk-based regulation may be partly a function of the ambiguity of the "risk-based" concept: it can mean different things to different people."*[15]

5.2 For example, the OECD has described risk-based regulation as "about focusing on *outcomes* rather than *specific rules and process* as the goal of regulation"[16] while, in the context of GDPR, the EU Article 29 Data Protection Working Party view risk-based regulation as "a scalable and proportionate approach to compliance".[17] In other contexts, it has been variously used to refer to the process of determining whether an activity should be regulated, to the use of internal risk management tools as a means of compliance, and to approaches for regulators to supervise regulated entities and enforce law based on risk.[18]

5.3 In the context of AI regulation, there is currently a strong international trend of governments preferring approaches that they describe as "risk-based" (see, for example, Europe's *Proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)* (**Draft EU AI Act**),[19] Canada's draft *Artificial Intelligence and Data Act*,[20] Brazil's draft *Bill No. 2338 of 2023*[21] and (most recently) China's *Interim Measures for the Administration of Generative Artificial Intelligence Services*).[22]

5.4 However, while these approaches are all "risk-based" in some sense, it is important not to conflate the following separate concepts:

(a) **proportionate legislation**: legislation that has been drafted to be proportionate to the risks it is attempting to address;

(b) **risk categorisation:** legislation that imposes obligations based on some categorisation of risk (e.g., the categorisations proposed in the Draft EU AI Act and Canada's draft *Artificial Intelligence and Data Act*); and

---

[15] Cary Coglianese, 'What does Risk-Based Regulation Mean?' *The Regulatory Review* (online, 8 July 2019) <https://www.theregreview.org/2019/07/08/coglianese-what-does-risk-based-regulation-mean/>.

[16] OECD, *OECD Regulatory Policy Outlook 2021* (online, 2021) ch 6 <https://www.oecd-ilibrary.org/sites/9d082a11-en/index.html?itemId=/content/component/9d082a11-en>.

[17] Statement of the WP29 on the role of a risk-based approach in data protection legal frameworks: Rather, the scalability of legal obligations based on risk addresses compliance mechanisms. This means that a data controller whose processing is relatively low risk may not have to do as much to comply with its legal obligations as a data controller whose processing is high-risk.

[18] Julia Black, 'Risk-based Regulation: Choices, Practices and Lessons Being Learnt', *Risk and Regulatory Policy: Improving the Governance of Risk* (OECD, online, 2010) p 187 <https://read.oecd-ilibrary.org/governance/risk-and-regulatory-policy/risk-based-regulation_9789264082939-11-en#page3>.

[19] *Artificial Intelligence Act* (EU) (retrieved 25 July 2023) <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206>; *Amendments adopted by the European Parliament on 14 June 2023* (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)) <https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html>(collectively, '**Draft *EU AI Act***').

[20] *Bill C-27* (Canada) pt 3 *"Artificial Intelligence and Data Act"* (retrieved 25 July 2023) <https://www.parl.ca/DocumentViewer/en/44-1/bill/C-27/first-reading>.

[21] *Draft Bill 2338 of 2023, Provides for the use of Artificial Intelligence* (Brazil) (retrieved 25 July 2023, only available in Portuguese) <https://legis.senado.leg.br/sdleg-getter/documento?dm=9347622&ts=1683827933990&disposition=inline&_gl=1*gdi95o*_ga*ODMxNzgxMDUzLjE2ODAxMDM2NTc.*_ga_CW3ZH25XMK*MTY4NDI1MDY2OS4xLjAuMTY4NDI1MDY2OS4wLjAuMA >.

[22] *Interim Measures for the Management of Generative Artificial Intelligence Services* (China) (retrieved 25 July 2023, only available in Simplified Chinese) <http://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm>.

(c) **risk management obligations:** legislation that requires entities to adopt the tools of risk management (e.g., obligations to undertake ex ante impact assessments or to maintain risk management systems).

5.5 In fact, because of this ambiguity, whether particular legislation is “risk-based” is often disputed. For instance, while the European Parliament refers to the Draft EU AI Act as risk-based,[23] some commentators have objected to this characterisation on several different grounds.[24] Similarly, others have pointed out that while the GDPR has adopted some risk management tools, it remains at its core a form of rights-based regulation.[25]

5.6 As a result, while we broadly support “risk-based” approaches (in all three senses mentioned above), this does not necessarily mean following any specific jurisdiction in their approach to Al regulation. In particular, the issues of risk categorisation and the use of risk management obligations are quite separate.

5.7 There will also be many details to consider if the Commonwealth decides to adopt risk-based legislation. These include the following two key issues:

(a) What risks are being addressed and how?

(i) Undertaking a risk assessment approach requires determining the risk criteria, including which harms are be taken into account, how the level of risk will be determined, the timeframes of interest and which levels of risks are acceptable.

(ii) If the Commonwealth uses risk assessment as the basis for determining whether particular obligations apply (i.e., risk categorisation), we consider that this will raise substantial challenges, including:

- whether the Commonwealth can adequately define the harms that it is seeking to address; and
- whether the Commonwealth can adequately define the levels of risk with sufficient certainty such that industry participants are able to determine whether their AI systems will be subject to increased regulations.

(iii) In doing this, the Commonwealth will need to strike a balance between over-regulating (which will impose unnecessary regulatory burdens) or under-regulating (which may mean that the regulation is ineffectual).

(iv) These issues have already emerged in the debates over how to define ‘high-risk AI system’ for the purposes of Europe’s proposed AI Act. The EU Council and European Parliament have offered differing definitions:

- the European Council has proposed that an AI system listed in Annex 3 of the proposed AI Act is considered high risk “*unless the output of the system is purely accessory in respect of the relevant action or decision to be taken and is not therefore likely to lead to a significant risk to health, safety or fundamental righ*ts”;[26] and

[23] Natascha Gerlach, ‘Why we need to return to a risk-based approach’, *iapp* (online, 23 March 2023) <https://iapp.org/news/a/the-case-of-the-eu-ai-act-why-we-need-to-return-to-a-risk-based-approach/> (‘***Gerlach***’).

[24] Lilian Edwards, ‘Regulating AI in Europe; four problems and four solutions’ (Ada Lovelace Institute, March 2022) <https://www.adalovelaceinstitute.org/wp-content/uploads/2022/03/Expert-opinion-Lilian-Edwards-Regulating-AI-in-Europe.pdf>; *Gerlach* (n 23).

[25] Fanny, Daniel Leufer and Estell Masse, ‘The EU should regulate AI on the basis of rights, not risks’ *accessnow* (online, 13 January 2023) <https://www.accessnow.org/eu-regulation-ai-risk-based-approach/>.

[26] Council of the European Union, *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative act - General Approach* (doc 14954/22, 25 November 2022) <https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf>.

- the European Parliament has proposed that an AI system listed in Annex 3 of the proposed AI Act is considered high risk "*if they pose a significant risk of harm of the health, safety or fundamental rights of natural persons*". [27]

(v) Defining risk categories will likely have significant consequences in practice. For example, the original European Council's Impact Assessment estimated that "no more than" 5-15% of AI systems in Europe would be classified as high-risk AI systems.[28] However, a recent survey of 106 AI systems in Europe by *applied AI* found that although 18% were clearly high-risk AI systems, 40% could not be clearly classified as high-risk AI systems. As companies are more likely to err on the side of caution and treat these AI systems as 'high-risk', this practically suggests the total of high-risk AI systems in the sample is up to 58%, potentially leading to unintended regulatory consequences for those creating and deploying some of these systems.[29]

(b) Will a risk-based approach remain sufficiently flexible?

(i) A key risk in setting out risk categories in regulation is that the regulation may quickly become outdated as existing risks evolve and new risks emerge (especially in a field as complex and fast moving as AI).

(ii) This issue has also already arisen in relation to the Europe's proposed AI Act, which has focused its risk-based categorisation on the 'intended purpose' of the AI system. In doing so, the EU presumed that an AI system would always be developed, or deployed, with an intended purpose. In early 2023, this approach was challenged by the sudden upswing in popularity of generative AI systems that use 'foundation models'.[30] As a result, the European Parliament has introduced the concept of 'foundation models' into their proposed legislation and set forth particular obligations applying to them.[31]

(iii) The need to amend these key definitions so soon after the original draft legislation was proposed suggests that the evolving nature of AI systems, their use cases, and the possible resulting harms are likely to be a continuing challenge in formulating appropriate risk categories.

(iv) It also demonstrates the particular difficulty in applying a risk categorisation approach to foundation models or general purpose AI systems that are capable of being used in a variety of ways and for a wide range of purposes. As we note below, it reinforces the importance of ensuring that any risk-based approach to the regulation of AI systems focuses on the risks inherent in the particular use or application of the AI systems, rather than in the technology itself.

(v) Accordingly, although Australia must be mindful of the approaches being developed overseas, and the Commonwealth must ensure that regulation does not prevent companies from complying with the Draft EU AI Act (once passed), we encourage the Commonwealth not to rush to copy the "risk based" approaches adopted in other jurisdictions and instead carefully consider how any proposed "risk-based" approach would apply in practice.

[27] *Draft EU AI Act* (n 19) art 3.

[28] *EU Commission Impact Assessment* (n 12) 68.

[29] Dr Andreas Liebl and Dr Till Klein, 'AI Act: Risk Classification of AI Systems from a Practical Perspective' (appliedAI, March 2023) <https://aai.frb.io/assets/files/AI-Act-Risk-Classification-Study-appliedAI-March-2023.pdf>.

[30] See, eg, *Draft EU AI Act* (n 19) recital (60e) <https://www.europarl.europa.eu/resources/library/media/20230516RES90302/20230516RES90302.pdf>.

[31] *Draft EU AI Act* (n 19) art 28(b).

6 Questions 17 and 18: Elements of a possible risk-based approach

> [17] What elements should be in a risk-based approach for addressing potential AI risks? Do you support the elements presented in Attachment C?
>
> [18] How can an AI risk-based approach be incorporated into existing assessment frameworks (like privacy) or risk management processes to streamline and reduce potential duplication?

6.1 In Attachment C of the Discussion Paper, the Commonwealth has presented measures as possible elements of a “risk-based” approach. While we broadly support risk-based approaches, we consider that the Commonwealth should avoid regulation that sets out a risk categorisation in legislation and then imposes rules-based regulation in relation to particular “high risk” systems. Such a proposal might look “risk-based” on the surface, but in fact requires compliance with fixed rules regardless of actual risks (and would be based on a risk assessment conducted once, without consideration of emerging risks.

6.2 Instead, we recommend that:

(a) any regulatory regime that applies to “high risk” AI systems (and addresses the issues in Attachment C of the Discussion Paper) should be approached as principles-based regulation rather than rules-based regulation; and

(b) the test for whether an AI system is “high risk” should be flexible enough to take into account the specific use cases of that AI system.

Rules-based versus principle-based regulation

6.3 Rules-based regulation is generally considered to be most well-suited to regulating something (i.e., conduct or technology) that can reasonably be expected to lead to the harm the regulation is seeking to address. However, rules-based regulation may lead to undesirable outcomes or become outdated where the thing being regulated may be harmful, neutral or beneficial depending on the circumstances. In contrast, principles-based regulation requires companies to adhere to broad principles rather than prescriptive rules. It provides companies (and regulators) with greater flexibility as to how to reach the required outcome while reducing the risks of unnecessarily stifling innovation.[32]

6.4 In the context of regulating the risks of AI, principles-based regulation is particularly appropriate given (as explored above) the harms presented by AI are highly contextually dependent and are likely to evolve over time. Accordingly, we consider there can be merit in

[32] Julia Black, Martyn Hooper and Christa Band, ‘Making a success of Principles-based regulation’ (2007) Law and Financial Markets Review 191 <https://www.lse.ac.uk/law/people/academic-staff/julia-black/Documents/black5.pdf>; Tania Van den Brande, ‘Rules-based versus principles-based regulation – is there a clear front-runner?’ *Ofcom* (online, 3 August 2021) <https://www.ofcom.org.uk/news-centre/2021/rules-versus-principles-based-regulation>; Australian Law Reform Commission, *For Your Information: Australian Privacy Law and Practice* (Report No 108, August 2020) <https://www.alrc.gov.au/publication/for-your-information-australian-privacy-law-and-practice-alrc-report-108/4-regulating-privacy/regulatory-theory/>.

introducing principles-based regulation that addresses gaps in sector-specific regulation (such as the current lack of specific guidance on appropriate internal AI governance mechanisms).

Combining risk-based and principle-based approaches

6.5 If some sort of AI regulation is to be introduced, we think that practically, and to ensure that such regulation is introduced in a manner that companies in Australia are familiar with and to minimise regulatory burden, consideration should be given to the introduction of high level "AI Governance Principles," similar to the Australian Privacy Principles, that draw upon the AI Ethics Principles and the concepts of responsible AI being developed by the CSIRO.[33]

6.6 The AI Governance Principles should:

(a) promote the establishment of internal mechanisms for organisations and agencies to identify whether they are deploying an AI system that will (or is likely to) involve a serious risk of harm to individuals, society or the environment. As demonstrated by the Draft EU AI Act (see above), narrowly defining this trigger is problematic. Accordingly, we suggest a regulator prepare and update guidance regarding when an AI system (in the context of its use or deployment) will, or is likely to, be considered as presenting a serious risk of harm to individuals, society or the environment;

(b) require that organisations and agencies undertake an AI Impact Assessment against the Australian AI Ethics Principles if they plan to deploy an AI system that will (or is likely to) present a serious risk of harm to individuals, society or the environment. In regard to AI Impact Assessments, we note the following:

(i) an AI Impact Assessment is broadly similar to a Privacy Impact Assessment (**PIA)** in that it will be a systematic framework for entities to understand the impacts (both positive and negative) of an AI system's actions so that they can be addressed, managed and mitigated. The appropriateness of PIAs as a risk mitigation tool within the data protection is demonstrated by the fact they are highly embedded within privacy regulation (including the GDPR and the Privacy Act). The suitability for the Australian context has also been recently confirmed in the Attorney-General's recent proposal that APP entities must conduct a PIA prior to undertaking a 'high-risk activity';[29]

(ii) in proposing mandatory PIA for companies as part of Privacy Act reforms, the Attorney-General has not proposed exhaustively defining when a PIA is required. Rather, "*The Privacy Act should provide that a high privacy risk activity is one that is 'likely to have a significant impact on the privacy of individuals.' OAIC guidance should be developed which articulates factors that that may indicate a high privacy risk, and provides examples of activities that will generally require a privacy impact assessment to be completed. Specific high-risk practices could also be set out in the Act.*"[34]

Accordingly, empowering companies to determine whether they are developing or deploying an AI System that will involve a reasonably expected risk of harm to individuals, society or the environment is in alignment with the proposed amendments to the Privacy Act; and

(c) require that organisations and agencies who undertake an AI Impact Assessment then take reasonable steps to mitigate risk of serious harm to individuals, society or the environment identified from the AI Impact Assessment as arising from the deployment of any AI system.

---

[33] CSIRO, *Responsible artificial intelligence research* <*https://www.csiro.au/en/research/technology-space/ai/responsible-ai*>

[34] Attorney-General's Department, *Privacy Act Review, Report 2022* (Final Report, 16 February 2023) <https://www.ag.gov.au/sites/default/files/2023-02/privacy-act-review-report_0.pdf>.

6.7 The Commonwealth should also co-ordinate approaches to the regulation of AI through the establishment of a new agency or regulator (or designating an existing agency or regulator) as a centralised point for regulating, coordinating and sharing expertise with other regulators responsible for sector-specific legislation and domain-specific legislation. The new agency or regulator should also be responsible for developing guidance and templates that will assist organisations to assess whether an AI system or the particular use case will involve a reasonably expected risk of presenting a serious harm to individuals, society or the environment and to mitigate those risks after having carried out an AI Impact Assessment. This guidance and any templates will need to be updated over time.

6.8 For completeness, in considering the role of AI Impact Assessments, we note that consideration must also be given to:

(a) whether transparency requirements need to be imposed on developers (or other participants in the AI value chain) to enable downstream deployers to effectively undertake AI Impact Assessments;

(b) when in the lifecycle of the AI system they need to be conducted (including whether AI Impact Assessments must be updated or replaced from time to time as the AI system evolves, and whether AI Impact Assessments are required if the AI system is only for internal use);

(c) whether AI Impact Assessments are distinct obligations on organisations (and if, so, the penalty for non-compliance and whether AI Impact Assessments can be challenged by third parties) or require specific evidence of compliance of other substantive obligations in relation to how they are performed; and

(d) whether (and when) a regulator should be able to audit AI Impact Assessments (for example, in response to complaints from individuals or demonstrated harm) and the consequences of audit results showing that the relevant requirements have not been complied with.

Elements in Attachment C

6.9 We do not otherwise support the elements in Attachment C being imposed as requirements separate to any requirement to conduct AI Impact Assessments. In our view, requirements in relation to notice, human in the loop/oversight requirements, explanations and training should not be considered separate elements. Rather, they are considerations that should be identified and implemented as risk mitigation measures following an effective AI Impact Assessment.

7 Question 19: General purpose AI systems and foundation models

[19] How might a risk-based approach apply to general purpose AI systems, such as large language models (**LLMs**) or multimodal foundation models (**MFMs**)?

General purpose AI and foundation models

7.1 Any regulation of AI should take into account the varied and evolving value chains through which AI systems are developed and used.

7.2 The terms ‘foundation models’ and ‘general purpose AI’ (**GPAI**) do not have a settled definition and are used slightly differently by different people in different contexts.[35] However, for the purpose of our submission, we use ‘foundation models’ to refer to AI

[35] Eliot Jones, *Explainer: What is a foundation model?* (Resource, 17 July 2023) <https://www.adalovelaceinstitute.org/resource/foundation-models-explainer/>.

models that are capable of a broad range of possible downstream tasks (and we do not intend to draw any distinction between foundation models and GPAI).

7.3 There have been several calls to regulate foundation models due to their centrality to the AI value chain.[36] The reasons for these calls include that:

(a) foundation models can be "singular points of failure that can radiate harms (e.g., security risks, inequities) to countless downstream AI applications";[37] and

(b) there is an asymmetry between developers of foundation models and downstream participants deploying foundation models in terms of information and negotiating power.[38]

7.4 As noted above, in response to these and other concerns the European Parliament amended the AI Act on 14 June 2023 to include provisions specifically relating to foundation models.[39] Despite these actions by the European Parliament, we recommend that the Commonwealth exercise caution in enacting any new regulation that is specific to foundation models:

(a) First, any regulation should be future proofed by not being tied to a particular business model or distribution channel. AI value chains are already complex and involve numerous participants, structures and models of distribution.[40] As such, simple distinctions between a developer and a deployer, or a deployer of a foundation model and a deployer of a downstream model are unlikely to reflect the full range of possibilities, especially as the technology and markets continue to evolve.[41] Making assumptions about how these value chains will operate could lead to confusion or may result in inflexible regulation that quickly becomes outdated or deters innovation and adoption of beneficial use cases.

(b) Secondly, as a general principle, regulation of AI should focus primarily on high-risk *uses* of AI, rather than the development of the technology itself, as this is more likely to be effective in preventing harms that arise from the use of AI systems. In particular, we recommend the Commonwealth avoid any licensing systems that will prevent the development of AI, noting that such restrictions are likely to be ineffectual and counterproductive if imposed by Australia.[42]

(c) Thirdly, any regulatory burden should be placed on those participants best placed to mitigate the relevant risks.[43] The complex value chain of AI systems makes this a challenge as there may be no single entity that is capable of assessing the risks of employing an AI system in a particular use case and of mitigating those risks.[44]

---

[36] See, eg, Bommasani et al, *Response to Request: AI Accountability Policy Request for Comment* (Letter, 12 June 2023) <https://hai.stanford.edu/sites/default/files/2023-06/Response-to-Request.pdf> ('***Stanford and Princeton Letter***'); AI Now, *Five considerations to guide the regulation of "General Purpose AI" in the EU's AI Act* (Letter, 13 April 2023) <https://ainowinstitute.org/publication/gpai-is-high-risk-should-not-be-excluded-from-eu-ai-act>; Future of Life Institute, *General Purpose AI and the AI Act* (Report, May 2022) <https://artificialintelligenceact.eu/wp-content/uploads/2022/05/General-Purpose-AI-and-the-AI-Act.pdf>; Natali Helberger and Nicholas Diakopoulos, *ChatGPT and the AI Act* (2023) *Internet Policy Review* 12(1) <https://policyreview.info/essay/chatgpt-and-ai-act>.

[37] Catelijne Muller et al, *AIA in depth - Objective, Scope, Definition* (Report, 13 February 2022) 12 <https://allai.nl/wp-content/uploads/2022/03/AIA-in-depth-Objective-Scope-and-Definition.pdf>.

[38] Ibid 14.

[39] *Draft EU AI Act* (n 19) s 28b.

[40] See Alex C Engler and Andrew Renda, *Reconciling the AI Value Chain with the EU's Artificial Intelligence Act* (Centre for European Policy Report, 3 September 2022) 6-14 ('***CEPS Report***'). See generally, Sabrina Kuspert, Nicolas Moes and Connor Dunlop, *The value chain of general-purpose AI* (Blog Post, 10 February 2023) <https://www.adalovelaceinstitute.org/blog/value-chain-general-purpose-ai/> ('***Ada Lovelace Institute Post***').

[41] See *CEPS Report* (n 40).

[42] Sayash Kapoor and Arvind Narayanan, *Licensing is neither feasible nor effective for addressing AI risks* (Blog Post, 10 June 2023) <https://www.aisnakeoil.com/p/licensing-is-neither-feasible-nor>.

[43] *CEPS Report* (n 40) 24.

[44] Ibid 14.

However, we consider there to be good reasons for risk management processes to focus on the downstream use cases.

7.5 Foundation models are, by their nature, intended to be generic tools. As noted by Alex C Engler and Andrew Renda, "a [foundation model] provider cannot reasonably predict the many diverse applications and contexts for which these systems will be used".[45] Similarly, the NSTC's Rapid Response Information Report on generative AI noted that "risk management approaches are most effective within a specific context of use".[46]

7.6 As a result, it may not be possible for a provider of a foundation model to undertake an effective general risk assessment of all possible uses cases of a foundation model and to then implement mitigations that would be well adapted to all these possible use cases. By contrast, a downstream deployer who is using a foundation model in a concrete use case will be better placed to identify the specific risks of that use case. The downstream deployer will also be able to decide upon the appropriate risk mitigation strategy (e.g., human oversight, notices to end users) that is appropriately adapted to the specific use case or decide not to use the foundation model at all if the risks cannot be appropriately mitigated.

7.7 One challenge for downstream deployers is that they may not have access to all the information necessary to assess the risks (e.g., access to the training data or sufficient knowledge of the operation of the foundation model). To overcome this asymmetry, the Commonwealth could consider targeted transparency obligations on providers of foundation models designed to enable downstream deployers to undertake appropriate risk assessments. We consider that the Commonwealth should, however, avoid the temptation of attempting to address all foreseeable harms of AI systems at the foundation model level, given (as noted above) the appropriate mitigations will often depend on the relevant use case.

7.8 In summary, we recommend that the Commonwealth:

(a) ensure that any regulation is future proofed by not being tied to a particular business model or distribution channel;

(b) focus any regulation on the use of AI systems by deployers in the context of concrete high risk use cases; and

(c) consider addressing asymmetries between market participants through transparency requirements, rather than attempting to impose granular requirements on developers and deployers of foundation models.

The open-source ecosystem

7.9 The Commonwealth should also avoid imposing regulation that would discourage the development or use of open-source AI models.[47]

7.10 In this context, we use 'open-source' to refer specifically to models that have been publicly released (including the weights of the model) free of charge and under a permissive licence.[48] A key feature of open-source releases is that they are often one-off transactions,

[45] Ibid 27.

[46] Genevieve Bell et al, *Rapid Response Information Report: Generative AI - language models (LLMs) and multimodal foundation models* (Report, 24 March 2023) 150 <https://www.chiefscientist.gov.au/sites/default/files/2023-06/Rapid%20Response%20Information%20Report%20-%20Generative%20AI%20v1_1.pdf>.

[47] See *CEPS Report* (n 40) 29; Alex Engler, *The EU's attempt to regulate open-source AI is counterproductive* (Commentary, 24 August 2022) <https://www.brookings.edu/articles/the-eus-attempt-to-regulate-open-source-ai-is-counterproductive/>; Jeremy Howard, *AI Safety and the Age of Dislightenment* (Article, 10 July 2023) <https://www.fast.ai/posts/2023-11-07-dislightenment.html>.

[48] *Ada Lovelace Institute Post* (n 40).

meaning that the developer may not have any ongoing oversight or control over the downstream uses of their models.[49]

7.11 Open-source distribution has numerous benefits, including spurring innovation, encouraging competition, enabling research, and improving cybersecurity.[50] Australian businesses often consider using open-source models as this enables them to host the models onshore. As a result, we consider that the Commonwealth should be careful about imposing regulation that will damage the open-source ecosystem and adoption of AI technology in Australia.

7.12 As noted above, as a general rule, the burden of regulation should fall on the participant in the value chain best able to assess and mitigate the relevant risks. In the case of open-source AI models, the developer of the model will have no (or a very limited) ability to control the downstream uses of the model or to mitigate new or previously unforeseen risks of a given concrete use. By contrast, the deployer or user of the model will have much better insight into the specific risks involved in their particular use case and the ability to implement mitigations tailored to those risks. Where the developer of the open-source model has provided the deployer/user with sufficient information about the model, we consider that the risk assessment can be left entirely to the deployer/user without further assistance from the developer.[51]

7.13 As a result, if any regulation is to be imposed on developers of foundation models (or those releasing them to the market), an exception should be considered for open-source releases. This exception could be conditional on certain transparency requirements, such as the release of information relating to the training data or training techniques necessary for downstream deployers/users to conduct their own risk assessments and meet any regulatory requirements. A similar exception has already been proposed in the context of the Draft EU AI Act[52] (although some concerns have been raised about the scope of this exception).[53]

King & Wood Mallesons
28 July 2023

[49] Ibid.

[50] *Stanford and Princeton Letter* (n 36) 7; Creative Commons et al, *Supporting Open Source and Open Science in the EU AI Act* (Letter, 26 July 2023) 3-4 <https://huggingface.co/blog/assets/eu_ai_act_oss/supporting_OS_in_the_AIAct.pdf> (‘***Supporting Open Source and Open Science Letter***’).

[51] See *CEPS Report* (n 40) 25.

[52] Draft EU AI Act (n 19) recital (12b).

[53] *Supporting Open Source and Open Science Letter* (n 50) 1.