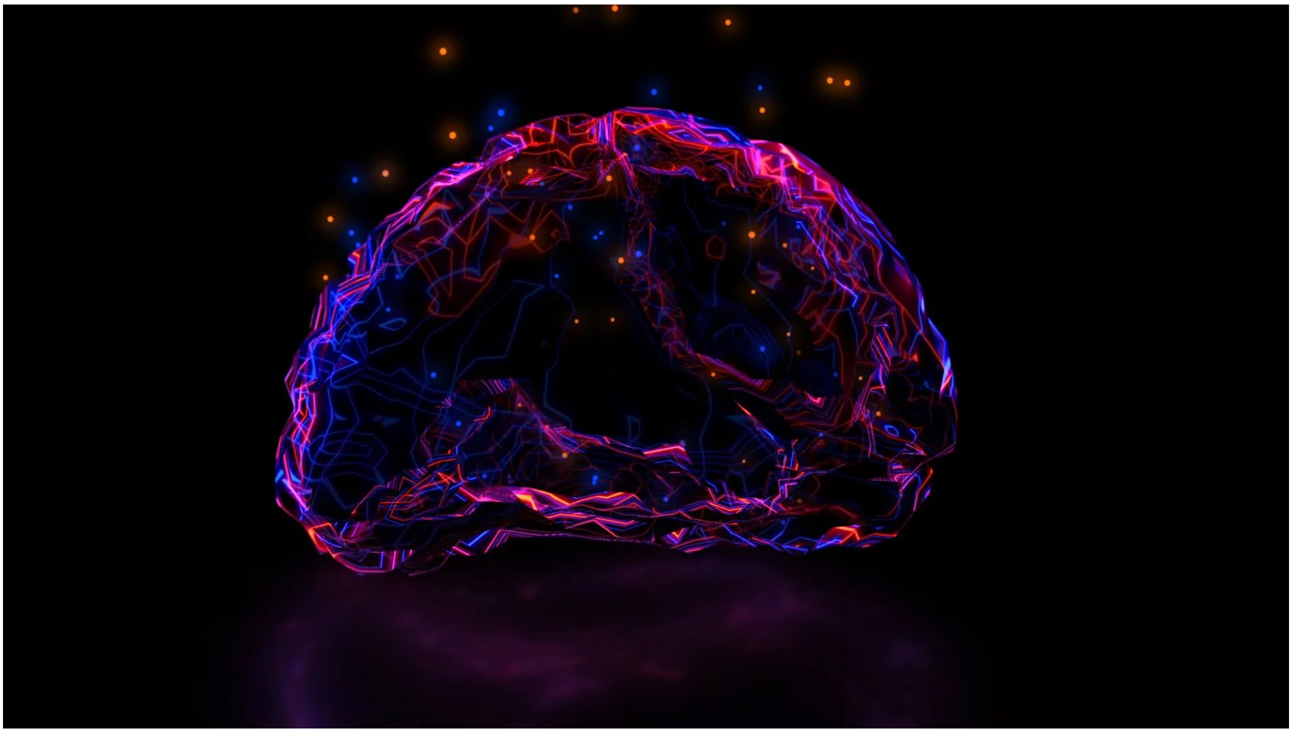

Herbert Smith Freehills submission

Safe and responsible AI in Australia – Discussion Paper

4 AUGUST 2023

HERBERT
SMITH
FREEHILLS

4 August 2023

Australian Government
Technology Strategy Branch

Department of Industry, Science and Resources

DigitalEconomy@industry.gov.au

Our submission to the Discussion Paper

Herbert Smith Freehills appreciates this opportunity to respond to the 'Safe and Responsible AI in Australia' (the **Discussion Paper**) published by the Technology Strategy Branch of the Department of Industry, Science and Resources (the **Department**) in June 2023.

Herbert Smith Freehills is a market leading global law firm, with a large presence in Australia, that advises clients across a wide range of jurisdictions and sectors. We are seeing an increase in the adoption of artificial intelligence (**AI**) technology and automated decision making (**ADM**) (beyond the technology industry), accompanied by uncertainty among our private and public sector clients regarding their legal and compliance obligations but also as to best practices.

Summary

We support a multi-tiered governance framework for AI in Australia that balances clarity required to promote responsible innovation so Australia can benefit economically from AI enabled technology, with the requisite consumer and societal protections to promote public trust in AI systems. To that end, we submit:

- **Regulation as an enabler of innovation:** The AI governance framework must focus on the role of regulation and governance as an enabler of innovation and economic growth recognising potential for Australia to be a key international player in development of AI as well as users/beneficiaries of AI systems.
- **Harmonisation with existing domestic laws and coherence with overseas frameworks:** To minimise regulatory overlap and duplication, Australia should carry out a detailed assessment of existing domestic regulatory frameworks to identify which potential risks of AI use cases are already adequately protected under, or could be addressed by the reform of, existing laws. To the extent existing laws are determined to be inadequate, we support the adoption of a risk-based, proportionate, approach to AI regulation. Such approach should be designed to:
  - give industry the clarity and flexibility required to drive innovation, investment, and employment in this space;
  - consider overlapping implications of emerging technologies (including AI) holistically across reforms of existing legal frameworks as well as the impacts of the convergence of increasing automation, digital assets, and reliance on automated systems; and
  - align with international approaches (which is especially important for multi-national corporations (who are the key drivers in AI innovation and investment).
- **Accountability promotes consumer trust:** A key foundation for a thriving AI economy is strong consumer trust in AI technology, especially AI used to make decisions that will affect individuals. Such trust will largely depend on how companies and governments are held

accountable for the responsible design and use of AI. In the absence of AI-specific legislation, a range of governance mechanisms such as standards and voluntary commitments might assist with setting market expectations and consumer trust around AI safety while an official regulatory framework is being developed.

Our submission focuses on Questions 1, 2, 4, 5, 14, 15 and 17 presented in the Discussion Paper.

Question 1: Do you agree with the definitions in this Discussion Paper? If not, what definitions do you prefer and why?

As a general comment, we note that definitions will need to be precisely formulated in a way that strikes a balance between being sufficiently broad (to be future-proof) and specificity (so that it does not overregulate all forms of machines, automation, or software), bearing in mind not all AI systems that satisfy the threshold definition should necessarily attract application of risk-based AI regulation.

This is especially important if Australia is considering a broad sector-agnostic AI regulation where the scope and application of the regulation hinges upon a few blanket definitions that must be capable of being applied to a wide range of sectors and applications.

We note that the proposed definitions in the Discussion Paper do not explicitly link to each other. In our view, the better approach would be to re-structure the definitions in a hierarchical manner so that each definition cross-references or builds upon the previous definition. We submit that this approach would reduce risk of definitional overlap or conflation. To that end, we propose to add or modify the following definitions.

“Artificial intelligence”

Suggested drafting

“***artificial intelligence (AI)*** includes:

(a) machine learning; or

(b) any other engineered method or functionality that is designed to generate predictive outputs such as content, forecasts, recommendations or decisions for a given set of objectives or parameters”.

Rationale

While we appreciate that the proposed definition in the Discussion Paper is adapted from the corresponding definitions in ISO/IEC 22989:2022[1], we suggest keeping the definition of “AI” as a high-level term which offers the following advantages:

- By cross-referencing to machine learning and non machine learning methods, our proposed definition covers the field of AI without delving into the theoretical complexities of what exactly is AI (which has competing definitions across academic, business and standards). See proposed definition of “machine learning” below.

[1] “*engineered system that generates outputs such as content, forecasts, recommendations or decisions for a given set of human-defined objectives*”. This definition is supported by two notes: (1) “*The engineered system can use various techniques and approaches related to artificial intelligence to develop a model to represent data, knowledge, processes, etc. which can be used to conduct tasks*”, and (2) “*AI systems are designed to operate with varying levels of automation*”.

- Paragraph (b) of our suggested definition is adapted from the Discussion Paper's original definition of AI. We think that the words "*generate predictive outputs*" (which we consider is a core characteristic of AI) adequately mitigate the risk of the term "*artificial intelligence*" extending to non-AI software and machines.

For noting, we note in the footnotes other alternative legislative definitions of AI which the Department could consider for other variations of paragraph (b). [2]

"**Artificial intelligence system**"

Suggested drafting

"***AI system*** means a system including one or more components or processes implementing AI."

Rationale

We suggest distinguishing between "*artificial intelligence*" (i.e. which captures the science, field or methodology) and "*AI system*" (which captures the implementation of AI into networks, processes and systems) so that regulation only applies to the development, deployment and use of AI rather than the science or field of AI itself.

The above drafting is adapted from ISO/IEC/IEEE 29119-1:2022, though we consider it worthwhile adding the words "*or processes*" to capture not only technical processes, but also business or human operations and actions that surround the implementation of the technology.

"Machine learning"

Suggested drafting

"***machine learning***" *is a method whereby a machine derives* ~~*are the*~~ *patterns or inferences* ~~*derived*~~ from ~~*training*~~ data ~~*using machine learning algorithms*~~, *which can be applied to new data for generating predictive outputs* ~~*prediction or decision-making purposes*~~ *without being explicitly programmed to do so.*"

[2] We note below some other definitions of AI sourced from regulation or guidelines:

- *"information-processing technologies that integrate models and algorithms that produce a capacity to learn and to perform cognitive tasks leading to outcomes such as prediction and decision-making in material and virtual environments. AI systems are designed to operate with varying degrees of autonomy by means of knowledge modelling and representation and by exploiting data and calculating correlations. AI systems may include several methods, such as but not limited to:(i) machine learning, including deep learning and reinforcement learning; (ii) machine reasoning, including planning, scheduling, knowledge representation and reasoning, search, and optimization. AI systems can be used in cyber-physical systems, including the Internet of things, robotic systems, social robotics, and human-computer interfaces, which involve control, perception, the processing of data collected by sensors, and the operation of actuators in the environment in which AI systems work"* (UNESCO Recommendation on ethics of AI).
- *"means a technological system that, autonomously or partly autonomously, processes data related to human activities through the use of a genetic algorithm, a neural network, machine learning or another technique in order to generate content or make decisions, recommendations or predictions*" (Bill C-27 (Canada)).
- "*a machine-based system designed to operate with varying levels of autonomy and that can, for explicit or implicit objectives, generate outputs such as predictions, recommendations, or decisions, that influence physical or virtual environments*" (Draft AI Act (European Union)).
- "*discipline concerned with the building of computer systems that perform tasks requiring intelligence when performed by humans*" (ISO/IEC 39794-16:2021).
- "*capability to acquire, process, create and apply knowledge, held in the form of a model, to conduct one or more given tasks*" (ISO/TR 5255-2:2023).
- "*capability of a functional unit to perform functions that are generally associated with human intelligence such as reasoning and learning*" (ISO/IEC 2382:2015).

Rationale We have proposed minor tweaks to the definition, particularly removing circularity (i.e. removing the reference to 'machine learning' in its own definition) and focusing the definition on the process of finding patterns rather than the patterns themselves (consistent with other established definitions).[3]

Other definitions

We note that the Discussion Paper proposes definitions for "*generative AI model*", "*large language model*" and "*multimodal foundation model*". While we broadly accept these definitions for the purposes of the Discussion Paper, we query whether these terms are necessary and/or appropriate for a risk-based approach regulation given that these terms are inherently technology-specific. We recommend that Government avoids new definitions for specific applications (unless there is strong policy rationale to include them) but rather express an application through the purpose of an "AI system" (defined above) – for example, generative AI could be referenced as 'AI systems used for the purpose of generating novel content', whereas large language model could be referenced as 'AI systems used for the generation of human-like text'.

We note under the current drafting in the Discussion Paper, the definition of Automated Decision Making would capture "AI Systems" as defined above as a possible (but not requisite) subset or component of ADM.

Question 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?

Prior to introducing new AI-specific regulation, the Australian government should first consider the existing regulatory frameworks and identify which potential risks of AI use cases are already adequately protected under, or could be addressed by the reform of, existing laws.

- ***Assessment of existing domestic laws***

  Many existing technology-neutral laws already apply to the development, use and consequences of AI in Australia including privacy, data security, product safety, consumer protection and human rights-based anti-discrimination statutes.

  We recommend the Government coordinate a multi-agency review of relevant federal and state laws to identify which potential risks of AI use cases are already adequately protected under, or could be addressed by the reform of, existing laws.

  We recommend this analysis consider legal issues or risks arising from AI to:

  - better understand whether an AI-specific regulation is truly necessary for the Australian context; and
  - develop more targeted regulatory solutions based on existing laws on an issue-by-issue basis.

  It is likely that this analysis will identify issues or risks where the relevant law is known but the application of the law is not clear in an AI context. In other words, these issues or risks

[3] For example, the ISO defines "machine learning" as "*process by which a functional unit improves its performance by acquiring new knowledge or skills, or by reorganizing existing knowledge or skills*" (ISO/IEC 2382:2015).

are neither novel nor unusual (and have been broadly considered under existing law) but need to be reconsidered through the new lens of AI. Some examples include:

- How does copyright apply to protect AI-generated works?
- Does a company 'collect' personal information (for the purposes of the Privacy Act) if such information was generated using data analytics and machine learning? If yes, how does the company procure consent for that collection if the generation of information occurs after the point where the consumer reviews the privacy notice?
- Is there a substantial lessening of competition (under the Competition and Consumer Act) where an organisation uses an AI system to set or fix prices (especially in response to a competitor's price moves)?
- Should the concept of personhood extend to AI systems?

These existing harms / issues do not necessarily require new AI-specific regulation but may be addressed by reforms to the relevant existing law at the first instance. Depending on the complexity of the issue, reform could range from guidance memorandums/notes released by the relevant regulator (to the extent the issue is a matter of unclear enforcement or legislative interpretation) to legislative amendment (to the extent the issue requires the actual text of the existing statute to be reconsidered).

By contrast, certain novel AI specific issues may be identified which raise unprecedented issues or risks posed by AI for which there is no obvious or direct applicable treatment under existing law. Some examples include:

- prohibition of high-risk systems
- transparency requirements for AI systems
- registration of AI systems

We submit that such a review will assist the Government to develop more targeted regulatory solutions based on existing laws on an issue-by-issue basis.

If the Government considers it necessary to have a new AI-specific legislation, we consider that such legislation would primarily address these novel issues (and defer to other existing laws for identified existing harms protected under existing technology-neutral or sector specific laws).

- ***Avoid fragmented approach to law reform***

  As emerging technologies drive towards an increasingly digital economy, different existing legal frameworks will be challenge by common themes. Government law reform processes should consider overlapping implications of emerging technologies (including AI) holistically across reforms of existing legal frameworks. For example, the impacts on our existing laws relating to privacy, anti-discrimination, competition, consumer protection, IP, cyber security, digital platforms, misinformation and disinformation, as well as industry or sector specific laws. In addition to impacts of increasing adoption of AI, law reform should also consider the impacts of the convergence of increasing automation, digital assets, and reliance on automated systems. See response to question 4.

- ***Harmonisation with existing domestic laws and standards***

  Further, in support of a multi-tiered approach to AI governance, government should continue to promote adoption of certifications and standards for AI, including audit standards, testing protocols and mitigating strategies for algorithmic bias, with incentives for industry participants who implement solutions for responsible AI.

If and when special purpose AI legislation is deemed necessary in Australia, it will be important to adopt a harmonised approach to promote consistency, avoid fragmentation and duplication of regulation and existing processes that organisations deploy to manage risk including standards, risk frameworks, and accepted industry methodologies. See our response to Question 14 for our specific recommendations.

Question 4: Do you have suggestions on coordination of AI governance across government?

Attachment A of the Discussion Paper highlights 16 current government initiatives underway in respect of AI. These initiatives are likely being run in parallel with limited cross agency coordination or consultation.

A dedicated taskforce could be established with a mandate to provide advice to government on AI governance reform and to assist with coordination and harmonisation of laws, processes, and methodologies.

Such an expert body would provide proactive policy and legal guidance on rapidly evolving technologies including AI. Such taskforce would be extremely beneficial to the various aspects of governance and law reform required for AI and the common issues arising with adoption of other key technological advances. Transformational emerging technologies have broad application and invariable touch many technology-neutral laws and sector specific laws.

Question 5: Are there any governance measures being taken or considered by other countries (including any not discussed in this paper) that are relevant, adaptable and desirable for Australia?

- *Coherence with international approaches*

  There is considerable activity taking place in relation to AI law reform globally. For Australia to become a leading AI economy, AI governance will need to be coherent with international approaches. This is especially important for multi-national corporations (who are the key drivers in AI innovation and investment). Clear and coherent international frameworks will promote compliance and avoid duplication and fragmentation.

We submit that the Government should:

- promote adoption of overseas certifications and standards for AI, including audit standards and testing protocols, with incentives for industry participants who implement solutions for responsible AI;

- consider how existing or proposed AI and data-related regulations, frameworks or strategies in other jurisdictions may inform, interact with or impact any measures taken in Australia, particularly where those measures will apply to actors or AI systems that are cross-border in nature;

- continue our leading role in the development of international standards recognising the key part standards will play in achieving international coherence in governance of AI systems.

Australia is unlikely to derive benefits from moving quickly to introduce specific AI regulation that may result in inconsistencies with other international frameworks. Given our position in the global economy, there is greater merit in monitoring the approaches adopted by other leading jurisdictions and at the appropriate time, adopting elements that are suitable and beneficial in the context of Australia's political, economic and legal system.

On that note, we outline below some overseas developments which have not been noted in the Discussion Paper but may still be of interest for the Department to consider.

United States of America

On 21 July 2023, seven big technology companies (Alphabet, Meta, Amazon, Anthropic, Inflection AI, OpenAI and Microsoft) made voluntary commitments to the White House to promote safe, secure and transparent development and use of AI.[4] Notably, some of the commitments included: (1) committing to internal and external red-teaming of AI models or systems, (2) developing and deploying mechanisms that enable users to understand if audio or visual content is AI-generated, including robust provenance, watermarking, or both, and (3) developing and deploying frontier AI systems to help address society's greatest challenges. [5]

While it remains to be seen whether these commitments will be effective, the fact that market-leading companies have been able to put aside their competitive interests and agree to set of principles is a noteworthy development which could potentially establish a market standard around the development, deployment and use of AI by other market players. The highly publicised nature of these commitments, which were announced by President Biden in a televised speech, would also likely have the effect of putting participating companies are under public scrutiny to fulfil their commitments. This could add a level of market accountability and help build consumer trust.

Key takeaway: Procuring voluntary commitments from market-leading industry players might assist with setting market expectations and consumer trust around AI safety while an official regulatory framework is being developed.

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

To the extent that the Government considers it necessary to implement new AI-specific regulation, we generally support a risk-based approach to AI regulation subject to our comments below:

- We note that the Discussion Paper (pg 32-33) identifies the risk categories according to impact of individual use cases or applications. While we appreciate this is an intuitive approach for formulating risk-based regulation, we note that to genuinely protect against certain high-risk applications of AI, risk-based regulation may also need to have regard to how individual applications could work in concert with other applications to create

[4] US White House, 'Fact Sheet: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI' (21 July 2023) at https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/.

[5] US White House, 'Ensuring Safe, Secure and Trustworthy AI' (21 July 2023) at https://www.whitehouse.gov/wp-content/uploads/2023/07/Ensuring-Safe-Secure-and-Trustworthy-AI.pdf.

systemic risks or harms. This is a more difficult challenge to solve for as it may involve multiple entities.

- Aside from the question of whether a risk-based approach is appropriate, we think the more difficult question is what rules, rights and obligations will apply for each risk category. While we appreciate the EU's draft AI Act as an instructive example of a risk-based approach legislation, it is important to consider the interoperability of such a risk-based approach with the features and nuances of Australia's common law and existing legal frameworks.

- A key concern of regulating certain activities in relation to AI under a specialised piece of legislation, will be defining the rights and remedies of consumers (and society more generally). Government should consider questions of enforcement and remedies carefully to ensure access to justice (including process, time and cost) for consumers. For example, merely giving the user a right to sue the counterparty for the counterparty's breach of their obligations the AI-specific regulation might not be practical where the user is an individual minor and/or is under-resourced to commence proceedings. We also recognise that many remedies will lie in technology-neutral laws that will need to work in concert with any specific AI act.

- The regime should provide sufficient transition periods and processes to facilitate a smooth adjustment to new laws.

- The regime be crafted through a collaborative, transparent, multi-disciplinary and multi-stakeholder process (including consultation with diverse interest groups), guided by rights-based approach to law reform, ethical principles of responsible innovation and responsible use.

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

Given the wide application of use cases for AI, we note a risk-based approach offers several advantages by:

- providing the flexibility to address a broad range of issues and risks that may arise now and in the future. Risk-based assessment provides the opportunity of proportional and tiered compliance requirements;

- limiting the negative impact of regulation on innovation, if only the AI systems with genuine risks are subject additional regulation; and

- conveniently housing AI rules in one place. This provides consistent standards across the board and has the potential to be consumer-centric.

However, we note some challenges that would need to be considered with this approach which include:

- It is possible that certain outcomes cannot be predicted when assessments are made, and unforeseen risks might arise through the combination of multiple AI systems controlled different AI actors. Any risk assessment process should therefore be an ongoing test rather than one made at a point in time.

- A further limitation to the effectiveness of a risk-based regulatory approach, lies in the potential lack of skilled personnel able to conduct risk assessments of potentially complex AI system. In addition to having government policies to encourage increases skilled personnel available in Australia (including through the education system), this limitation can be minimised by clear and comprehensive guidance on factors that indicate high or low levels of AI risk.

- The introduction of holistic AI statute could potentially overlap with other existing regulations which, if not addressed, could add to the increasing regulatory compliance burden on businesses in Australia. A single AI regulation should focus on those elements of AI that make it distinctive and avoid impinging in other areas unnecessarily.

- The effectiveness of a holistic, sector-agnostic framework approach to AI regulation will depend heavily on clear and flexible taxonomy. In particular:

  - risk categorises should not be defined by prescriptive blacklists but rather technology-agnostic terms accommodating of changes in AI technology.

  - risk definitions or standards should reflect contextual factors (e.g. intended purpose of the technology, the circumstances of its use, purported and actual impact, nature of the user/beneficiary of the technology) so that risk is assessed on a case-by-case basis.

Question 17: What elements should be in a risk-based approach for addressing potential AI risks?

As a general comment, we submit that the risk-based approach must support the role of regulation as an enabler of innovation and economic growth. We specifically note:

- **The rationale for, and the purpose of, a risk-based law would need to be clear.** A primary objective must be to inform and control the development and deployment of relevant AI systems. However, we assume, to the extent possible, such legislation would defer to other relevant laws to provide specific remedies and enforcement for consequential harms resulting from use of the AI system. Using privacy as an example, a face filter app may use an AI system that collects biometric facial data. The key risk relates to the sensitive nature of the data collected and potential impact of harm from misuse of that data. Would demonstrating compliance with privacy laws negate or diminish the risk-level classification under AI legislation and thereby reduce the compliance burden on developers, or would the risk-based classification remain regardless of demonstrated compliance with other law based on the potential for impact / harm? The Government will need to be clear about the obligations/remedies specific AI law could impose that would mitigate privacy risk in ways not already provided for under the Privacy Act. This assessment is necessary before subjecting AI systems to obligations/rules that would not ultimately mitigate the identified risk.

- **Risk categories need to be clearly defined.** Obviously, clearly defining the regulatory perimeter of any risk-based AI-specific law will be important for clarity of application. We note:

  - The EU's risk-based approach specifies different tiers of risk level in way that is linked to existing AI use cases. Although there is a mechanism for the list of high-risk use cases to be updated, it is possible that risky AI systems are missed and fall outside of the scope of the regulations.

  - Consequently, a more holistic approach, based on individual risk assessments applying a set of core criteria, should be adopted. This approach could build on sector-agnostic risk classification frameworks such as the OECD's Framework for the Classification of AI systems[6] and the NIST AI Risk Management Framework[7] (**NIST AI RMF**) which look at a range of interested or affected parties at different stages of the AI system's life cycle.
  - Depending on the ultimate definitions implemented and suite of compliance standards / obligations required, we assume any AI system that does not satisfy the criteria for any of the specified risk categories would remain entirely 'unregulated' by the ambit of the act.

- **For AI systems that fall within the regulated risk categories, regulatory responses and compliance obligations should be proportionate to the level of perceived potential risk.** We agree with the current thinking in the Discussion Paper that regulatory consequences should be tied to governance and transparency in the operation of the AI system. We also agree that the higher the risk, the greater likelihood the AI system will require stronger governance controls (including in design, testing and operations) and greater transparency. However, we note:
  - Low risk AI systems should be subject to light touch governance obligations in terms of ethical principles or voluntary codes of conduct. In particular, the regime should provide clear but voluntary standards that: (i) encompass the essence of universal AI ethical principles; and/or (ii) promote responsible AI use and innovation and best practice. While the regime could offer incentives (whether social, economic, such as grants, or through a statutory mechanism) to encourage businesses to comply with these standards, it should encourage, rather than penalise businesses for non-adoption. This targeted focus will promote responsible innovation.
  - The NIST AI 'RMF Core' reflects these principles through the cyclical process of govern, map, measure and manage. The Government should consider a framework along similar lines.

- **Australia should also consider introducing a clear set of consumer-centric principles** (similar to the Australian Privacy Principles) that can guide policy relating to technology and the digital economy holistically and provide the foundation for responsible development and use of AI.

We are pleased to provide this submission to the Department and would welcome the opportunity to discuss our comments further.

[6] OECD Framework for the Classification of AI Systems at https://www.oecd.org/publications/oecd-framework-for-the-classification-of-ai-systems-cb6d9eca-en.htm

[7] NIST AI Risk Management Framework at https://www.nist.gov/itl/ai-risk-management-framework

About Herbert Smith Freehills

Herbert Smith Freehills is one of the world's leading commercial law firms, bringing the best people together across our 24 offices globally.

In our capacity as a trusted professional advisor to a large number and variety of clients, across a wide range of industries and sectors, we have experience supporting our clients to thrive in the digital age and navigate novel technological changes. We were pleased to be a major project partner of the Human Rights and Technology Project of the Australian Human Rights Commission.

Herbert Smith Freehills has a number of specialist practice areas that consider the legal and regulatory issues arising in connection with technology and data. These specialists work closely with our business and human rights specialists. This submission was prepared by our cross-practice group, multidisciplinary Emerging Technology Group who are providing bespoke advice and practical solutions to the opportunities, risks and ethical and regulatory requirements brought on by digital transformation and the adoption of emerging technologies.

These experiences mean that we have a multi-dimensional perspective on the issues raised by new and emerging digital technologies and their impact.

Key contacts

Julian Lincoln
Partner
+61 3 9288 1694
julian.lincoln@hsf.com

Kwok Tang
Partner
+61 2 9225 5569
kwok.tang@hsf.com

Susannah Wilkinson
Digital Law Lead – Australia
+61 7 3258 6786
susannah.wilkinson@hsf.com

Alex Lundie
Senior Associate
+61 3 9288 1918
alex.lundie@hsf.com

Raymond Sun
Solicitor
+ 61 2 9322 4173
raymond.sun@hsf.com