To: Department of Industry, Science and Resources

I have written this brief submission both as a technologist as well as a father – in my view there is opportunity to leverage AI for the benefit of the Australian society, but it must be done with a balanced approach. There are tremendous advantages that AI can deliver but also raises a few areas of concern that, based on specific use-cases, warrant further legal controls to be in place to protect individuals as well as different sections of a diverse society.

By way of experience, I have a technology background and have worked with clients in many industry sectors and geographies, so my intention is to provide context and examples from a broader perspective on the topic where possible. Additionally, I have general understanding of the Australian laws and potential gaps that may exist in context of governing the use of AI systems.

Most of my content is primarily aimed at Generative AI – the branch of AI that automatically generates new content levering underlying technology and training datasets but (in many cases) without the ability to provide better transparency (explainability) and accountability. These factors have high risk of eroding trust in AI. Generative AI should therefore be used with robust governance and responsible legal controls in place to protect end stakeholders and societies at large who could be impacted the most (if appropriate moderation is not in place). This applies even more so to government organisations in particular who choose to use such technologies for efficiency gains in their service provision to end citizens, students, patients, etc., either directly or indirectly.

My views are also largely aimed at procured generative AI services and solutions from 3rd party vendors (large multinational technology vendors). Based on the current rate of evolution of this technology and the significant influx of venture capital in the past 12 months, driven primarily from the US market, I have assumed most if not all Australians will interact with generative AI via procured solutions – that is not to say generative AI solutions cannot be built here. As it stands, most of the mainstream offerings as candidates for early adoption seem to be currently offered by vendors originating from outside Australia. Applying an Australian legal framework to procured technologies hosted overseas and offered by non-Australian providers itself creates some potential governance challenges.

I would like the Australian government to recognise lessons learnt from the introduction of social media solutions into our society (early versions of societal-scale AI solutions) and how they are now entangled into everyday life, the impacts (both positive and negative) of these solutions across different generations and sectors of our population, and then utilise these learnings to introduce better legal protection from emerging solutions like generative AI - and mitigate against adverse outcomes appropriately where avoidable. The pace at which technology evolves cannot be matched by the rate of change in legal reform, but extensive effort must be invested to review this iteratively with a balanced perspective and not merely as a reactive or one-off exercise.

This is not an exhaustive submission but can contribute to the wider debate on the topic of safe and responsible use of AI. This response does not include any generative AI output and is formed from my own research and opinions.

Answer to Q1:
No answer provided.

Answer to Q2:
To my understanding the following potential risks appear to not be covered by existing regulatory approaches.

In general, it appears that there is no specific legislation in place to expressively deal with decisions made by AI systems and any flow on liabilities that could arise from such decisions. The "at fault" or "bad actor" components in the decision-making processes are hard to identify/pinpoint definitively. The attribution of liability is unclear whether it is the end user of the solutions, the organisations deploying the solutions, the provider of the AI solutions or the AI technology itself that bears the primary responsibility when adverse outcomes do occur from decisions made by AI systems. Traceability in this regard is vague. Similarly, the notion of ownership when it comes to generative AI e.g., with respect to copyright, is also ambiguous as per the current interpretation/definition of Australian laws that cover this.

Furthermore, on generative AI: Whilst lighter-weight generative AI solutions are available for deployment at a local geographic level, they often lack robustness. The more demonstrable and capable solutions like OpenAI's ChatGPT underpinned by Large Language Models (LLMs) are primarily hosted in USA. By the very nature of their magnitude, it seems unlikely that they will be hosted in other jurisdictions any time soon. Therefore, when it comes to Australian end users of such solutions (organisations as well as individuals) access is provided via a chatbot or API interface, directly or embedded in other 3rd party solutions. The notion of how any Australian data privacy and any future legislative oversight could be applied to such foreign-domiciled offerings therefore remains ambiguous.

- Australian end users could easily share private data unintentionally (or even maliciously) and there appears little to no recourse to enforce these LLMs to unlearn once queries with private data are submitted. What happens to the data in the queries, who sees it, how is it stored or disposed off, etc. - all remain unclear to or poorly understood by most lay people. And yet the technology is openly being used/experimented with in Australia and in many cases, decisions taken based on outputs from such solutions.
- Furthermore, the risk of correcting misinformation (hallucination) and how any legal oversight can protect end users also remains poorly understood (cases of copyright infringements and defamation have already surfaced in US and other countries). These solutions appear to provide answers with confidence but often lack the support of supplying supporting facts or accurate reference data.
- Based on what is published about training LLMs, who trains them, how they are trained, based on what training datasets etc., some form of bias seems to be built into the solutions (intentional or otherwise) and therefore the credibility of information being returned to end users could impact different parts of society in inequitable ways.

Answer to Q3:
The following non-regulatory initiatives and examples (non-exhaustive) should be explored further as part of deploying responsible and safe AI practices in Australia:

- Continuous community engagement and education. The means by which generative AI has been introduced to our society (largely due to the meteoric rise and instant popularity of ChatGPT since Nov 2022), has resulted in governments and society at large being caught up in a reactive mode. The technology continues to evolve at a phenomenal pace with heightened investment interests but our society's understanding of the technology, its limitations and issues is not growing at a similar rate. The tech providers in the meantime have pivoted the responsible use of the technology largely back to the end users and governments with a thin veil of accepting responsibility of how their solutions are performing and what corrective measures they have introduced or could introduce in future versions. As the Australian government develops supporting AI legislation in the coming years, in parallel it should also strive to drive educational campaigns at all levels of government and across all pockets of society. Absence of this could lead to some end user decisions being made by different stakeholders, underpinned by misinformation from generative AI, all because end users do not understand the full spectrum of possibilities and limitations of the technologies more broadly speaking.
- Mainstream education should also teach the future generations and workforce about safe and responsible use of AI, including what legal frameworks already exist and what other developments are under consideration.
- Taxes and penalties. The evidence of growing data security breaches and increased adverse impacts of social media solutions on our society (the first wave of mass-scale AI solutions) should serve as a learning opportunity and be used as material input into AI-related legislation reform. Governments have been reactive and/or unable to properly enforce better governance controls to protect end customers. This second wave of AI solutions and vendors should therefore be subjected to more stringent penalties, taxes and controls for specific use cases (identified via a risk-based approach). These are use-cases where decisions made by humans are potentially (mis)guided by incorrect information.

  By way of an example, at a surface level, it appears that solutions like generative AI can act as a tremendous enabler in education for instance. But how can the vendors be held responsible if their offerings start educating children and end users with misinformation, with bias, etc.? How much effort is required on the part of the teaching fraternity to provide additional oversight checks on top of already heavy workloads to monitor this? The quality of education cannot be compromised by utilisation of poor-quality information/data and certainly not without accountability controls in place to ensure problems do not perpetuate continuously.

  This needs to be examined through several perspectives and use-cases as poorly informed decisions could have adverse effects on many fronts e.g., social welfare outcomes for underrepresented classes of our society, discrimination, defamation, equal opportunity, etc.

  Furthermore, (beyond the cover of legislation) if becoming more analytical and resilient are attributes and skills of the future (as published by many sources), how

does the government ensure its population does not blindly depend on generative AI as a default solution and defeat the very goal of cultivating these future skills in the next generation?

Whilst innovation in generative AI should not be stifled or penalised, equally the counter argument also holds true where irresponsible innovation should not have free reign on how it released to society and where controls are not set in place to protect our society at large.

Answer to Q4:
Australia should continue observing and learning from Europe, US and other countries as a worthwhile exercise. At a local level, the government should ensure there is adequate and coordinated coverage across appropriate stakeholders and organisations to review the progress of legislation reform and representation of different perspectives reflected in these changes.

Answer to Q5:
No response provided.

Answer to Q6:
This is the opportunity for government organisations to lead by example and set benchmarks for others to follow – where generative AI is deployed, government agencies should have clear governance defined prior to the deployment, demonstrate best practice principles of safe and responsible use of (generative) AI, have proper disclosure in place to inform citizens and the wider community of how the technology is being used, for what kinds of use-cases, what risk mitigations and controls are in place etc.
Private organisations by their nature are more likely to be driven by efficiency drivers/cost saving factors when it comes to the use of generative AI – the rhetoric of how many job types could be replaced by generative AI as an efficiency enabler has heightened interest in this technology at board levels and beyond. Other than broad legislative oversight to protect rights of individuals, it is unclear if the same levels of rigour could be adopted by the private sector.

Answer to Q7:
Government agencies must prioritise educating their workforce (without compromise) on the advantages and disadvantages of generative AI. This could include annual recertification of agency personnel (especially primary users & custodians of the deployed technology) to reflect how they keep abreast of technology and legal evolutions and updates. This could also include audits of how the technology is being used, how risk registers are in place to capture feedback and any adverse impacts of decisions made from the continued use of generative AI. Such proactive measures and actions would demonstrate the safe and responsible use of AI as best practice in government agencies.

Answer to Q8:
Any legal solutions should include both generic and technology-specific solutions. For generative AI in particular, a technology-specific solution is warranted given the technology is creating new content and the responsibility of the accuracy of information derived or utilised is ambiguous.
A different combination of both generic and technology-specific solutions may apply to other types of AI solutions and other specific use-cases e.g., use of AI in the health sector.

Answer to Q9:
Transparency breeds trust – whether for public or private sector use. It could be applied via a combination of both mandated controls and best practice guidance (voluntary). Given the probabilistic nature of generative AI for information-seeking use-cases, transparency must be applied across the entire lifecycle – from training data sets to development practices and controls to 3rd party independent audits and risk registers. Whilst protecting the commercially sensitive aspects of AI systems is appreciated, it must not be done so at the expense of adverse impacts to society and individuals where generative AI solutions have been deployed. Informed or educated users are better placed to decide on how best to accept or reject any advisory information derived from generative AI. A gold standard in cultivating an ethical culture around use of generative AI in organisations must be aimed for without compromise. Some level of prescriptive intervention/legislation is warranted to demonstrate better transparency and increase the levels of trust in technologies like generative AI.

Answer to Q10:
Given that generative AI cannot statistically provide a 100% accurate answer (at best, only probabilistic answers can be generated), the government needs to define high-risk AI use-cases accordingly. E.g., using generative AI for efficiency gains like drafting emails is very different to relying on generative AI as a bona fide source of accurate information.

And due to my limited understanding of risks related facial recognition and bias, discrimination, privacy, profiling etc., I remain extremely conservative on when and where any supporting AI systems should be deployed.

Answer to Q11:
To be read in conjunction with the response provided in to question 9 above.

Examples of initiatives or government actions to increase public trust in generative AI include community education, capacity building initiatives, innovation awards for responsible and safe use of AI solutions, community engagement and consultation forums, potential use of generative AI in education and other government services, disclosure principals and transparency reporting, etc.

Answer to Q12:
No response provided.

Answer to Q13:
No response provided.

Answer to Q14:
Yes, a risk-based approach is warranted – not all AI solutions are alike, and neither are all identifiable risks comparable to each other per se. There are benefits to using generative AI solutions for some use cases for instance but also risks associated with other types of use-cases. The EU AI Act is a good example of a risk-based approach that Australia can learn from/leverage, but the government should not shy away from taking more stringent actions and decisions should gaps be identified in the legal frameworks from around the world in their current form. Given that legislation reform can be a slow process, Australia should aim to ensure a wide variety of perspectives can be incorporated, and that any new legislations/changes to existing laws can provide maximum protection against the different types of AI solutions that could be deployed.

Answer to Q15:
A risk-based approach encourages responsible and balanced use of AI solutions. Higher risk areas/use-cases warrant better mitigation controls be in place whilst not stifling innovation and other positive outcomes for lower-risk areas/use-cases.

Answer to Q16:
I view this more so from a use-case perspective irrespective of sector, organisation size, AI maturity etc. – regardless of public or private sector deployment of generative AI, which specific use-cases could lead to unintended outcomes? Primarily, generative AI use-cases pertaining to information generation are most susceptible to hallucination as they lack the support of referenceability (in the current versions of technologies released) – presenting information with confidence to users that may not be formulated from factual data. What does “failing to be fair” legally look like for these use-cases? Any augmentation or changes to Australian laws must take this into account.

Answer to Q17:
The more comprehensive the list of elements are, the more protected society is from potential harms resulting from generative AI-based decision-making processes. I support the elements presented in Attachment C.

Answer to Q18:
No response provided.

Answer to Q19:
Given that the providers of general-purpose AI systems like LLMs and MFMs largely originate from outside of Australia (as it stands today), applying a risk-based approach will require the Australian government to introduce supporting legislation in tandem with how legal coverage

is evolving in other geographies. Additionally, given that the AI technology is evolving at such a rapid pace, a national risk register comprised of identified risks/examples of adverse impacts submitted from different sections of our society (individuals, public and private organisations) coupled with a review of ongoing legal reform in parallel might provide a practical approach to dealing with a lot of unknowns that are yet to be identified. The reform process should be continuous given that no one can predict what direction the AI solutions are taking beyond what is known about them in the short-term future.

Answer to Q20:
This will depend on specific use-cases – but suffice to say, where impacts to individuals and groups within our society are the greatest currently, is where generative AI and its vendors cannot be held accountable for bias, misinformation, etc. In these situations, a level of mandated risk-approach should be applied regardless of whether the organisations are public or private and should also apply to both developers and deployers of generative AI solutions. To build trust with citizens, customers, employees, etc. organisations of all types will have to cultivate their own culture of ethical and safe use of AI solutions.