



# Minervai 4.0

Policy Submission

RESPONSE TO THE DEPARTMENT OF INDUSTRY,  
SCIENCE AND RESOURCES SUPPORTING  
RESPONSIBLE AI: DISCUSSION PAPER

**August 2023**

## **Introduction & Overview**

Minervai 4.0 is a youth-led, volunteer-based group advocating for the ethical and responsible development and use of technology moving into the “Fourth Industrial Revolution” or “Industry 4.0”. We are made up of volunteers aged 18 to 28 and are passionate about ensuring that technologies inherited by today’s young people embody principles of inclusion, care and intersectionality.

Our submission provides a non technical perspective to the regulation of Artificial Intelligence (AI) in Australia. We note that our submission focuses on social and ethical implications of AI, rather than its technical creation and application. We believe AI will be a fundamental technology to our generation, and we want to ensure its development and deployment is responsible, and ultimately benefits society.

## **Definitions: Question 1**

Definition	Agree/Disagree	Explanation
Governance	Disagree	From a consumer perspective, we believe governance should be sub categorised into ‘governance’ and ‘voluntary governance’ to indicate whether regulatory or voluntary mechanisms are in place. We believe this will provide greater transparency to consumers.
Regulation	Agree	Whilst the definition for regulation is generally representative of both voluntary, market, prescriptive and environmental influences, we believe the word may generally be associated with legislative or government regulation. This should be considered when developing consumer or user advice.
Artificial Intelligence	Disagree	<p>The proposed definition is possibly too narrow in its representation of AI. We note that use of key words such as ‘decisions’, ‘varying levels of automation’, ‘human-defined objectives’ does not necessarily represent the full ambit of AI. For example, AI does not only necessarily generate predictive outputs, but may be used to interpret inputs, process information, learn, take actions, perceive its environment etc.</p> <p>There are also important philosophical and ethical</p>

		questions as to whether the ‘artificial intelligence’ being referred to is true intelligence. While this may not be an imminent issue for the purposes of regulation, this may require consideration in future.
Machine Learning	Agree	
Large language model	Disagree	A more comprehensive definition may be required, for example, incorporating aspects of natural language processing. Further, it may be too narrow to define LLMs as ‘specialising’ in generating human-like text. Their functions include analysis, interpretation and processing of text as well as natural language “requests”.
Multimodal Foundation Model	Agree	Whilst this definition is reasonable, we recommend considering a more nuanced definition referring to MFM structures, rather than only its capabilities/outputs. <sup>1</sup>
Automated Decision Making	Agree	Whilst this is a comprehensive definition for the purposes of regulating AI, we note the significant need to delineate in what circumstances ADM is equivalent to human or even legal decision making.

### Recommendation

1. We recommend making it clear whether regulation or governance refers to voluntary or prescriptive mechanisms.
2. We recommend considering either a broader ‘catch-all’ definition,<sup>2</sup> or a more comprehensive definition of AI. We note the key words and taxonomy developed by the Joint Research Centre of the European Commission which may be used to develop a definition.<sup>3</sup>
3. Consider incorporating the functions and underlying structures in definitions for LLMs and MFMs.
4. Further research and consideration be done on whether Automated Decision Making can be considered a “decision” for social and legal purposes.

<sup>1</sup> See [https://www.adalovelaceinstitute.org/resource/foundation-models-explainer/#\\_ftnref4](https://www.adalovelaceinstitute.org/resource/foundation-models-explainer/#_ftnref4)

<sup>2</sup> See [www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework-glossary-and-methodology](http://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework-glossary-and-methodology); See <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52018DC0795>

<sup>3</sup> [https://eprints.uq.edu.au/28047/1/3.%20irc118163\\_ai\\_watch\\_defining\\_artificial\\_intelligence\\_1.pdf](https://eprints.uq.edu.au/28047/1/3.%20irc118163_ai_watch_defining_artificial_intelligence_1.pdf)

## **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?**

As digital and online platforms have become a central part of life for Australia's young people,<sup>4</sup> we would like to draw attention to AI applications in online settings and regulation in such contexts.

We are concerned that current laws are not robust enough to protect individual rights, particularly when considering how some AI systems are trained and deployed. Many AI systems require vast amounts of data to be trained on, data which may have been collected with "consent" from various third party platforms. However, consent collected in such circumstances is often limited, and largely ineffective.<sup>5</sup> We note that this is exacerbated through complex and overbearing privacy policies that provide little opt-out mechanisms, some of which are now being updated to include that data collection may be used to train AI systems.<sup>6</sup> While these privacy policies are explicit, there is very limited control or restrictions that users can employ to protect their personal data. We acknowledge that this may be addressed under the current review of the *Privacy Act 1998* (Cth).<sup>7</sup>

Further, we are concerned about applications of AI in social media, a platform widely used amongst young people,<sup>8</sup> and even considered an indispensable part of everyday life. AI in social media is deployed in various ways, including through data analytics, behaviour inferencing, generated content and importantly through the curation and display of content itself. There are currently no known or effective transparency requirements for embedded AI in such contexts; despite social media's extensive use, only 59% of users are aware that social media apps use AI.<sup>9</sup> We are concerned that users, particularly young people, are being subject to possibly exploitative AI systems in such contexts without their knowledge or control. While some aspects of privacy and consumer law may be able to address aspects of this concern, we believe a stronger response to AI is required, through targeted regulation. For example, a regulatory requirement to allow users to turn "off" AI curated social media feeds may provide individuals with greater control over their online lives.

In the context of data collection and analysis, we are also concerned about whether anti-discrimination law can apply to instances of algorithmic bias, and subsequently bias in

---

<sup>4</sup> See <https://www.acma.gov.au/sites/default/files/2021-05/The%20digital%20lives%20of%20younger%20Australians.pdf>

<sup>5</sup> See <https://fpf.org/wp-content/uploads/2022/06/20220628-ABLI-FPF-Consent-Project-Australia-Jurisdiction-Report.pdf>

<sup>6</sup> See e.g., <https://www.searchenginejournal.com/google-updates-privacy-policy-to-collect-public-data-for-ai-training/490715/>

<sup>7</sup> See <https://www.ag.gov.au/rights-and-protections/publications/privacy-act-review-report>

<sup>8</sup> <https://www.acma.gov.au/sites/default/files/2021-05/The%20digital%20lives%20of%20younger%20Australians.pdf>

<sup>9</sup> <https://assets.kpmg.com/content/dam/kpmg/au/pdf/2020/public-trust-in-ai.pdf> Page 43

AI systems. We note the significant implications of algorithmic bias in decision making in the criminal justice system, recruitment processes, law enforcement and financial products. We note, as above, that many may not be aware of AI being used in certain applications, and therefore may not be aware of its possible bias. Therefore, while anti-discrimination laws may apply to algorithmic or AI bias, we are concerned that a lack of awareness and transparency may make these laws less effective.<sup>10</sup> Beyond transparency and awareness measures, we believe a mechanism to audit systems for instances of algorithmic bias may assist in ensuring these systems are compliant with applicable anti-discrimination laws.

Further, we are concerned about the law's ability to prevent AI facilitated sexual exploitation and digital violence. For example, despite the introduction of criminal offences and fines for non consensual image-based sexual abuse, these regulations have either been ineffective or only as effective as their enforcement.<sup>11</sup> We also note the concerning use of technology to facilitate and exacerbate coercive control methods by perpetrators of domestic violence.<sup>12</sup> We are concerned about whether the law can protect victims as AI makes such techniques increasingly accessible.<sup>13</sup> Ultimately, there is a general perception of a lack of protection over predatory uses of AI, particularly in the context of digital and domestic violence.

Overall, it is evident that individuals have too little power in the face of AI systems, and that more distinct individual autonomy and rights should be created through regulation. We also believe that strong government enforcement is critical to mitigating the risk of AI systems. Considering individuals' limited power in the context of this disruptive technology, we are of the view that consistent and strong government responses are what will ensure trust and safety in AI systems.

### **Recommendations**

5. Review regulatory mechanisms' effectiveness against unlawful AI systems particularly focusing on use cases in social media.
6. Consider the introduction of a specific individual right to not be subject to AI systems in particular environments such as social media.
7. Conduct research into the effectiveness of laws to protect against AI facilitated abuse including image-base sexual abuse, coercive control and domestic violence.

---

<sup>10</sup> See <http://www5.austlii.edu.au/au/journals/ANZCompuLaw/Jl/2021/4.pdf>

<sup>11</sup> <https://www.theguardian.com/society/2020/may/09/revenge-porn-in-australia-the-law-is-only-as-effective-as-the-law-enforcement>;  
<https://www.sbs.com.au/news/the-feed/article/a-streamer-was-caught-looking-at-ai-generated-porn-of-female-streamers-the-story-just-scratches-the-surface/vfb2936ml>

<sup>12</sup> See <https://lens.monash.edu/@politics-society/2022/07/28/1384928/half-of-australians-will-experience-technology-facilitated-abuse-in-their-lifetimes-new-research>

<sup>13</sup> See <https://www.ucl.ac.uk/computer-science/research/research-groups/gender-and-tech/tackling-technology-facilitated-abuse-protect-victims-and>

**3. Are there any further non-regulatory initiatives the Australian Government could implement to support responsible AI practices in Australia? Please describe these and their benefits or impacts.**

Education

We believe there are two ways the education system can be bolstered to build public trust in AI. Firstly, we believe school students should be taught how to safely and effectively use AI systems. Just like how students were once taught how to use new text applications, students should be taught how to use AI systems, a technology likely to be an essential tool. We also note that there are several examples of how AI applications in both teaching and learning can be positive for students.<sup>14</sup>

Secondly, at a tertiary level, ethical and social learning components should be embedded in computer science, coding and technology courses. We believe emphasising the development of human-centred technology at a tertiary level will have a positive flow on effect when students enter their respective industries.

Public Awareness

Public awareness on the types and safe use of AI technology is critical to gaining public trust. As mentioned above, public awareness of AI use in environments such as social media is limited. We believe raising public awareness of AI systems is particularly important in cases where AI is “embedded” (i.e. used as an underlying tool such as in data analytics or curated social media feeds). Raising public awareness about what AI is and how it is used will bolster transparency and accountability of AI systems.

**Recommendations**

8. Review primary and secondary education to include teaching syllabuses on the safe and responsible use of AI.
9. Encourage tertiary education institutions to mandate ethical and social teachings as part of computer science, technology and engineering courses.
10. Create a public education campaign to raise awareness of types of AI and increase awareness of where AI systems are in use.

**4. Do you have suggestions on coordination of AI governance across government? Please outline the goals that any coordination mechanisms could achieve and how they could influence the development and uptake of AI in Australia.**

---

<sup>14</sup> <https://educational-innovation.sydney.edu.au/teaching@sydney/how-ai-can-be-used-meaningfully-by-teachers-and-students-in-2023/>

Considering the 2021 Human Rights and Technology Project Final Report,<sup>15</sup> establishing an AI Safety Commissioner may assist in the coordination of AI governance across government. As per the report, the creation of an AI Safety Commissioner could support regulators, policy makers, government and business in applying laws and standards for AI-informed decision making. A centralised AI Safety Commissioner will assist in streamlining AI regulation while also encouraging its lawful and ethical use in government and non-government settings. With government bodies made accountable to an AI Safety Commissioner, it may also bolster public trust in AI and therefore encourage its uptake.

#### **Recommendation**

11. Consider incorporating the recommendations of the 2021 Human Rights and Technology Project Final Report in any regulatory model for safe and responsible AI, including establishing an AI Safety Commissioner.

### **Responses suitable for Australia**

#### **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?**

Minervai 4.0 is supportive of the measures being employed by the European Union, through the *AI Act*. While the Act may be considered far reaching, its risk based approach prevents stifled innovation whilst protecting individuals from risky AI systems.

#### **Recommendation**

12. Consider, in addition to recommendation 11, the incorporation of a risk-based legislative mechanism as seen by the European Union's *AI Act* in any regulatory model for safe and responsible AI.

### **Target areas**

#### **6. Should different approaches apply to public and private sector use of AI technologies? If so, how should the approaches differ?**

Public sector use of AI should take into consideration the public expectation for transparency, efficiency and ethical practice. We believe public sector applications of AI should be used in low-risk settings, such as in communication triaging. Arguably, such applications can have significant efficiency gains within the public service, freeing up

---

<sup>15</sup> [https://tech.humanrights.gov.au/sites/default/files/2021-05/AHRC\\_RightsTech\\_2021\\_Final\\_Report.pdf](https://tech.humanrights.gov.au/sites/default/files/2021-05/AHRC_RightsTech_2021_Final_Report.pdf)

resources for more complex matters. We note that in a regulatory setting, public sector applications of AI will be subject to a higher public expectation and scrutiny. As many things in the public sector, we believe the sector should be a “model user” of AI.

#### **Recommendation**

13. Ensure a regulatory model encourages only low-risk AI to be deployed in public sector environments.
14. Engage in strategies to ensure the public sector and public service are “model” developers and users of AI systems.

### **8. In what circumstances are generic solutions to the risks of AI most valuable? And in what circumstances are technology-specific solutions better? Please provide some examples**

While a risk or harm based solution to AI is valuable, we are concerned about the nuances and application of this approach on specific AI use cases. We believe that regulation should not necessarily stifle innovation or positive use cases of AI. We point to examples of AI in construction and health as key areas where AI may be considered a regulatory risk under “generic” solutions, but have great benefits to individuals and broader society.

#### Construction

AI could assist in expediting parts of the construction process, much like what we witnessed with the implication and standardisation of Building Information Modelling (BIM) technologies (used to design and manage construction and infrastructure projects).<sup>16</sup> Application of AI in the construction industry may invoke many benefits such as analysing material use to reduce wastage. AI also presents an opportunity to further the effectiveness of BIM programs, creating systems to quickly identify faults within designs and documentations. There is also the possibility of using AI through robotics to work alongside builders on site,<sup>17</sup> to bolster efficiency and effectiveness of the construction process. While AI systems present significant efficiency and sustainability possibilities, safety risks also emerge. For example, AI system mistakes or failures in the design and construction process of a building may not only endanger those involved in the project, but occupiers of buildings itself.

#### Health

It is well established that AI applications in health care can provide great benefits to healthcare systems and patients. This includes diagnosis and treatment and patient

---

<sup>16</sup> <https://www.standards.org.au/news/australia-adopts-international-standard-for-bim-data-sharing>

<sup>17</sup> <https://www.sciencedirect.com/science/article/pii/S219985312201054X>



engagement.<sup>18</sup> However, there are critical ethical considerations to be made when employing AI systems in healthcare, considering not only the sensitivity of the data used in these systems, but the significance of the system outcomes. For example, a missed diagnosis could have devastating impacts on individuals.

#### **Recommendation**

15. For systems that may be considered “risky” but may have significant benefit to individuals or broader society, specific regulatory mechanisms should apply. This may include specific accountability, transparency, privacy and stringent safety requirements.

**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.**

We would like to briefly suggest how transparency requirements could be implemented in relation to generative AI. Minerva 4.0 supports some method of labelling of generative AI material, such as text, images or other content. Similarly to the Classifications Scheme or advertising disclosure requirements on social media, labelling generative AI content could assist users in making an informed choice about information and products. We believe this will enforce a greater sense of accountability on those using generative AI, and instil greater trust from the public.

#### **Recommendations**

16. Mandate labelling of AI generated or manipulated material in consumer settings, such as images, videos, text, advertisements, publications etc.
17. Mandate content labels for where AI systems and embedded AI are being used in consumer settings to invoke greater transparency and accountability.

**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?**

Minerva 4.0 is greatly concerned about the use of AI to exploit and harm. For example, AI can be used to create non consensual deepfake pornography and exacerbate image-based

---

<sup>18</sup> See <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6616181/>

sexual abuse. This application of AI is becoming increasingly accessible, with downloadable mobile apps bringing the technology to smartphones. Essentially any image online can be used to create non consensual sexual images which can be used to exploit and victimise anyone. We note that these technologies are almost exclusively used to harm women: a report found that 96 percent of deep fakes were non consensual deep fake pornography and 99 percent of those were made of women.<sup>19</sup> We also note the possible uses of this technology to increase AI-generated child sexual abuse material.<sup>20</sup> These use cases of AI incite significant harm on women and children, and we strongly believe that such applications of AI should be banned.

#### **Recommendation**

18. Mandate and enforce the total ban of AI generated deep fake pornography, to prevent and discourage the creation of non-consensual, exploitative sexual abuse material.

### **11. What initiatives or government action can increase public trust in AI deployment to encourage more people to use AI?**

Similarly to other products and industries, if AI systems were to be subject to tests in order to meet specific standards, we believe this will increase trust in AI and therefore encourage its use. For example, there is a general consensus of trust in products that are subject to specific Australian Standards, ranging from consumer electrical products to industrial materials. If a similar mechanism applied to AI systems, this may instil greater trust in their deployment. We draw attention to the White Paper ‘The Metaverse and Standards’ written with the Responsible Metaverse Alliance, to demonstrate how emerging technologies can interact with Australian Standards.<sup>21</sup>

#### **Recommendation**

19. Consider the creation of nuanced Australian Standards for AI systems to ensure the creation of safe and responsible AI.

### **Risk-based approaches: Question 14 - 20,**

Arguably, a risk-based approach to regulating AI strikes an appropriate balance between protecting individual rights, and not stifling innovation. If implementing an approach similar to the European Union’s *AI Act*, a risk-based approach ensures that harmful applications and

<sup>19</sup> [https://regmedia.co.uk/2019/10/08/deepfake\\_report.pdf](https://regmedia.co.uk/2019/10/08/deepfake_report.pdf) ;

<https://www.cigionline.org/articles/women-not-politicians-are-targeted-most-often-deepfake-videos/>

<sup>20</sup> <https://www.washingtonpost.com/technology/2023/06/19/artificial-intelligence-child-sex-abuse-images/>

<sup>21</sup> <https://www.standards.org.au/documents/h2-3061-metaverse-report>

uses of AI are subject to restrictions, whereas less risky AI will be subject to less restrictions. In theory, this will deter the development and use of AI that poses greater risks or harms to its end users, but still welcome innovation.

As alluded to above, some sectors using AI, such as health or law enforcement may require specific approaches. This is because of the nature of the outcomes, being ones that have significant implications on individual lives. This is where stringent standards processes prior to the release of AI systems may be required in combination with a risk-based approach to regulation. Notably, this is one of the benefits of a risk-based regulatory model, as the “level” of restriction or requirements can be nuanced according to the specific AI application.

In relation to specific elements that should be considered in a risk-based approach, we note that an element specifically addressing harm should be considered. We are concerned that narrowing regulatory considerations to AI “risk” may overshadow the very human effects of AI systems. For example, while embedded AI use in social media may be categorised as low risk, the possible long term effects of AI curated or generated material may be overlooked. Contrarily, if harm is used as an element for categorising risk, a more nuanced approach to regulating AI systems can be employed, ensuring greater protection for users. We note that harm as an element, can also ensure that specific harms are considered, such as those against specific groups including First Nations people, women and other marginalised populations.

Further, although AI can be used for positive environmental applications, we are concerned about the negative environmental impacts of the computational power required to train some AI systems.<sup>22</sup> For example, it is estimated that the amount of power required to train and search a particular neural network involved roughly 626,000 pounds of carbon dioxide emissions.<sup>23</sup> It is therefore critical that AI systems are subject to sustainability requirements so that their benefits can be fully realised without excessive environmental damage.

### **Recommendations**

20. Employ a risk-based approach to regulating AI systems, considering a range of restrictions and requirements according to the “riskiness” and context of the system.
21. Include a specific element of harm (being harm generally, and to specific populations) to be assessed as part of a risk-based approach to regulating AI.

---

<sup>22</sup> See <https://insights.groglobalgroup.com/the-environmental-impact-of-ai>; See <https://ec.europa.eu/research-and-innovation/en/horizon-magazine/ai-can-help-us-fight-climate-change-it-has-energy-problem-too>.

<sup>23</sup> See <https://arxiv.org/pdf/1906.02243.pdf>

22. Ensure AI systems are subject to stringent sustainability requirements, whether through a specific element, or through a separate regulatory mechanism.

### **Conclusion & Other Matters**

While this discussion document did not consider employability themes, we are concerned about the effect of AI on the future of the workforce. We urge the government to consider the possibly significant effects of automation in the workplace. As a youth-led organisation, we want to ensure that AI can be used as a positive tool in the workforce, rather than one to replace skilled workers.

Finally, we believe it is critical to ensure young voices are considered in the development of regulating such a fundamental technology. Should any further opportunities arise to contribute to the regulation of safe and responsible AI, we welcome and look forward to participating. We again thank the Department for the opportunity to provide our views on the significant policy issue of responsible AI.