RMIT UNIVERSITY

Enterprise AI and Data Analytics Hub

Safe and responsible AI in Australia

Our response

26 July 2023

Table of Contents

1. Introduction

The RMIT Enterprise AI and Data Analytics hub (the Hub) received an invitation on 1 June 2023 from the Digital Economy team within the Department of Industry, Science and Resources to comment and provide feedback to the discussion paper titled "Safe and responsible AI in Australia" (the Paper) released by the Minister for Industry and Science, the Hon Ed Husic MP.

In developing our response, the Hub hosted a roundtable event on the 21st June 2023 that was attended by 43 academics and industry practitioners (RMIT Enterprise AI and Data Analytics Hub, 2023). The event has a very good mix of academics from economics, finance, social justice, law and information systems, and industry representations from AI and Analytics platforms, HR services, government, consulting, banking, retail, health services, and insurance.

Through this multi-disciplinary consideration and a set of views from across different industries, the Hub aims to achieves a considered response that reflects its unique positioning as a research centre within the College of Business & Law, where AI research is seen through a translational lens focusing on the whole lifecycle of an AI solution rather than just the technology piece on its own.

Inputs from the roundtable, including the data capture, recording and other material, received explicit consent from participants via RMIT's PCIF consent form. The analysis of the data, including the production of this report and subsequent materials from the Hub, has also received ethics approval (Ref:26685) at RMIT University. Any request for clearance information can be obtained by sending the request to one of the Hub's contact details.

2. About authors and contributors

This document is a collective effort among researchers here at RMIT University, led by the Enterprise AI and Data Analytics and members of RMIT Digital3, a coalition of research centres that take an interdisciplinary view to enabling the digital economy of Australia through research, education and thought leadership. Overall this submission does not represent the views of each of the academic and industry contributors, but rather intends to synthesise and explore a wide range of views expressed by the 43 participants in the workshop. Key centres that contributed to this document includes

Enterprise AI and Data Analytics Hub[1]

The RMIT Enterprise AI and Data Analytics Hub (RMIT Enterprise AI and Data Analytics Hub, 2023) is a cross and interdisciplinary research group situated within RMIT's College of Business & Law. Our mission is to undertake research and translation of AI and Analytics solutions to help businesses navigate the challenging environment of technology acceleration and the volatile business environment. The hub brings an enterprise-first focus to help organisations address their challenges through AI and analytics strategy design, applied research, talent development and more. We are also strongly connected with

[1] https://www.linkedin.com/company/rmit-ai-analytics-hub

technology partners to ensure that our work is rooted in seamless research translation into effective AI and Analytics solutions for Australian organisations.

Blockchain Innovation Hub[2]
The RMIT Blockchain Innovation Hub (RMIT Blockchain Innovation Hub, 2022) is the world's first research centre on the social science of blockchain incorporating economics, sociology, public policy and political economy to provide a new way to understand the global blockchain revolution.

Business and Human Rights Centre[3]
Business and Human Rights Centre (BHRIGHT) (Business and Human Rights Centre, 2023) addresses the human rights impact of business through research, education, innovation and collaboration. Our work on technology tackles some of the biggest human rights issues emerging from the increasing pervasiveness of tech in our daily lives.

In addition to these research centres, the document also includes contributors from across the University, industry participants at our roundtable event, and Industry Advisory Board members from RMIT Digital3.

3. Summary of Recommendations

We recommend the following course of actions for the Government of Australia to ensure safe and responsible use of AI.

1. Evaluate current regulations and legislations to incorporate AI into the existing regulatory frameworks.
2. Establish AI-specific regulatory guidelines that define transparency and accountability of human and artificial agents team involved in the design process.
3. Initiate the design of explanations in AI systems that are deployed in public services.
4. Co-create national AI policy by engaging stakeholders such as public, academia, industry partners, regulatory bodies, and other relevant stakeholders.
5. Introduce verification and authorizations systems, where independent bodies evaluate AI systems' risks and performance.
6. Inculcate the approach of "humans in the loop", where AI-generated output is not directly used (or have suitable interventions) in the decision-making process of high risks applications.
7. Encourage inter-disciplinary AI R&D through diverse perspectives from computer science, social science, psychology, and philosophy.
8. Develop public trust through initiatives that explains AI outputs to improve perceptions of fairness and transparency.

[2] https://rmitblockchain.io
[3] https://www.rmit.edu.au/research/centres-collaborations/business-and-human-rights-centre

4. Response to questions outlined in discussion paper

Our response to the discussion paper is guided by the range of industry views and the expertise from our research leaders. We have divided our response into two parts. Section 3, which is this section of the report, will address the specific questions that the government is seeking answers to within the discussion paper. The next section, Section 4, will address aspects of considerations that we have captured from the roundtable as well as inputs from our Industry Board Members at RMIT Digital3.

4.1 Definitions

As the Paper rightly noted, there is no single agreed definition of AI. We see a range of understanding reflected by our participants at the roundtable. We observed that this is very much influenced by one's experience, or the angle in which the technology is experienced or perceived. On the definition of AI that the Paper puts forward, the participants do agree that it does sufficiently reflect their understanding.

From the Hub's perspective, the terminology is meant to be an umbrella term, so its specific definition will evolve as new methodologies or new AI categories are developed. This is the case with Generative AI as a new category, or deep learning as a new methodology. As investment in AI continues to accelerate both in the R&D space and through new start-ups, we expect that the definition of 'AI' will continue to evolve. As such, a definition that sufficient help the community understand what 'AI' does should serve its intended purpose.

The concern with a broad definition, or one that is loosely defined, is that we risk challenges within the legal system where any AI regulatory implementations may have limited reach or protection for the community overtime. Therefore, we believe any definition of AI should be sufficiently clear (or well-understood) for the intended regulatory purpose, and that AI should also be defined by its operational characteristics instead of the specific algorithms and techniques that will quickly become obsolete. This will allow any regulatory design to withstand the rapid pace of development in this space.

We also note that there is broad interpretation of terminologies like "responsible AI", "safe AI" and "explainable AI". Broadly, these terminologies are not well-understood, nor does it reflect the technical definitions that most computer scientists are familiar with. As such, we suggest that considerations for definitions should not be restricted to just "AI" but rather, be extended to AI concepts, including "responsible", "explainable" and "safe".

In our analysis from the data captured at the roundtable, we see the concept of "responsibility" being discussed as an intent or accountability that participants do not expect AI systems to have on their own. In other words, there is a consensus among participants at the roundtable that the designers, users, and decision makers are expected to inculcate responsibility in system design and use.

> *"I would much rather call it the responsible use of AI rather than responsible AI. I think we should be very careful between saying the responsibility of the model itself, and the responsibility of the people who've decided to take the output from the model and automatically executed the decision or not saying that they use that or whatever it happens to be. Whenever an AI makes a decision, someone has implemented some rules that say that's our decision. So now we can sort of drill into how models do this, and who makes rules about that".*

We see this as a positive development in the 'responsible AI' space as the discussion reflects an understanding of shared responsibilities and agreement that the onus is on the humans rather than the technology. On this topic, a common theme that emerged is how "responsible AI" can be as simple as disclosing the model's "error rate" or helping end-users understand the limitations of the AI through other "threshold" details like nutrition labels for food that helped consumers understand what is in the food they consumed.

> *"The elephant in the room with AI, is this error rate. Every model has error rates, nobody wants to go there. They're like, you have to do you have to define your error rate should be at this level, you need to disclose it, you need to disclose"*

We agree that the industry's adoption of AI lacked transparencies around details. It is difficult for anyone outside of the organisation to have an idea of the limitations of an AI model, including the design of the entire software systems that uses the AI. In defining "AI", we see a great opportunity here to also define the "threshold details" that should be disclosed for any implementation of AI products. While we see value in the idea around "nutrition labels" for AI models, we believe a more pragmatic approach is to draw inspiration from financial industries through "product disclosures", where different AI technologies have different types of thresholds that should be disclosed.

While these details are unlikely to be easily understood for the layperson, defining what needs to be disclosed is the foundation for regulators to assess AI models that is consistent and fair to creators of AI. This is an angle that is a different take to the notion of "fair" as defined by computer science literature around "responsible AI". Yet, it is important in terms of implementing any meaningful AI governance, whether it is self- assessed frameworks or in designing compliance measures within the risk levels that the Paper is considering.

4.2 Potential gaps in approaches

The Paper has identified the gaps in existing domestic AI governance landscape. Through the roundtable discussion, we are able to present a coherent and coordinated response that recognises the issues pertaining to regulatory approaches.

We found an incremental approach towards regulatory approaches. First, we found agreement on using existing frameworks and policies to institutionalize AI technologies. The existing frameworks regarding data protection, privacy, copyright, online safety and criminal law were regarded as the first set of solutions. The logic behind using existing frameworks is

to recognise the underlying meaning of regulating AI. As discussed in previous section, “responsible” and “safe” use of AI was suggested. AI systems do not function in isolation and require human agents (mostly decision makers) to enable use of AI systems. Therefore, the same set of regulations regarding data protection, privacy, copyright, online safety and criminal law applicable to human agents can work for human agents in an AI context.

> *“In the very first instance, we should apply our existing framework to this new technology, and see how far that takes us. And I actually think, just in terms of responsible AI and just responsible behaviour, generally, we already have that fairly good system in place, and we should apply that system to see how far I think it'll take us very, very far. Especially if you take them and take AI as another tool and we already regulate human behaviour through these frameworks”*

As a second course of action, AI specific governance regulations were also proposed. One type of AI regulations can be revising the existing legislation and frameworks for AI by understanding the impact of AI in the legislation and make necessary changes.

> *“Government should be revisiting their legislation, and then think about how AI to have experts come in and say, what are the AI implications for this to this legislation to every piece of legislation? What role does AI have here? And then, if you need separate legislation, because of AI, then it's broader than AI”*

Surprisingly, the idea to launch a central coordination body was found instead of having a central body to regulate all sectors

> *“It should be a central body to advise the government, I think there should be a central party that helps coordinate responses, but I don't think there should be a central body, which has the authority to make the decisions”*

The subsequent step in regulating AI is to in enforce principles and legislation for safe and responsible use of AI.

> *“I think the government has to focus on putting some kind of control or procedure for AI output. There has to be a legislation because then if one person decides we are not going to do it, there is nothing stopping the other person from doing it”*

The policies around AI regulation also empower the users and citizens to protect their interest against third parties. Only a regulatory check can place third parties accountable for “irresponsible” use of AI systems.

> *“If the government doesn't mandate a way, for an end user to go and challenge them [private companies] there has to be a way of challenging, no company is going to basically allow you to challenge the system says no. I think that's where the government can play a big role in basically”*

Despite regulatory controls and legislative measures in place, the need to place humans in the loop is essential. The narrative built earlier also strengthens the role of humans as final decision-makers of using AI-based output. To promote the responsible use of AI, placing humans in the loop was a frequently discussed concept.

> *"So the use case is how are we going to turn it completely over to the AI or we're going to keep the human in the loop where their recommendations coming from the AI doctor. So that's where I think the government has to stop"*

> *"And we already regulate human behaviour. And it's actually ultimately the human in the loop that is going to be regulated here"*

It is also important to note that distinguishing between human roles and machine is important. The "humans in the loop" strategy might not work if there are no clear boundaries defined between AI and human jurisdiction. There has to be a mechanism to isolate system objective and human involvement. One participant mentioned that we need to start isolating between AI system output and its deployment by a human agent. It also includes human judgement on using the system output.

> *"If the human was in the loop in all things [system performance] there is a clear responsibility for when things go wrong because we know from the beginning and no one can deny everything for us"*

> *"To start off with as a tractable regulatory approach, if you are using AI to make decisions, a person is making a decision, not an AI. What, what am I using? How is it inputted into your decision? That makes more sense from a regulatory perspective to me"*

To ensure responsible use of AI, we suggest engaging with AI testing and and verification bodies that issue verification of AI system accuracy. The concept of AI audit was discussed that means testing AI performance against objectives.

> *"Government can lay down the rules of what a valid AI audit is"*

We also suggest that government can play an intermediary role in coordinating between AI system designers and testing bodies.

> *"I think if we look at what the government can do, in that space, they can actually facilitate that will actually provide the hubs which you can integrate with us and test your algorithms"*

There are multiple avenues of cost-saving for AI testing. The incentives by the government to encourage algorithmic and AI testing can play a significant role in improving the overall performance of AI system in a responsible and safe manner.

> *"If you have a way to test your algorithms with some government entities, I think that will actually create a lot of value. Because other than that, you have to invest in order to go and test it. But if there is already a testing sense of build, because the more you do the testing, the more you will learn and grow, and you will improve your standards"*

4.3 Responses suitable for Australia

The Hub is aware of the various developments in AI governance around the world, including voluntary best practices as well as regulatory approaches among countries with high degree of AI activities. We are also aware of the sector specific (bottom-up) developments of 'AI ethics' and 'responsible AI', which also reflected the community's interest.

A key observation that we draw is that these proposals share a similar 'soul', i.e., there are similarities in their approaches, and all try to achieve a similar outcome. It would make sense for Australia, being a part of the world economy, to adopt similar responses to ensure trade and economic alignment with its trading partners, especially in the exchange of products and services with AI embedded in it.

That said, we recognise trying to harmonise standards for AI between countries to be challenging. This can be seen in the case of the risk-based approaches, that while similar in principle, differ in the determination of risk levels and risk response. Such difference(s) *cannot* be avoided as is the case with other regulations, e.g., medical devices. As such, our view is that Australia should adopt a response where the 'soul' of the approach is similar to the key economies that Australia aligns itself with. In doing so, (i) it will be easier to articulate Australia's response to the public, and (ii) the level of trust in the regulatory approach will be higher among the community.

We emphasise the 'soul' of an international response because we believe Australia's response, while aligned with international developments, should be and needs to be different because of our own unique characteristics. Hence, the Hub would like to submit the following considerations to aid the development of a response. We believe these considerations (including other considerations in Section 5) will give Australia's AI regulation framework the ability to interoperate internationally in its trade while protecting the interests of the Country and its people.

An AI regulatory framework that understands the diversity of the Australian people and its links to the world. Australia's demographic is changing, and this change is accelerating as seen in the census. Having an eye to how a diverse group will interpret any regulatory framework will be crucial. This is especially so given how this diverse demographic will bring their views of AI, including use of AI in 'high risks' applications, into Australian society. Our response on this challenge is not one that is negative. Rather, we believe there is an opportunity here to engage the collective wisdom of the population on how any AI regulatory framework can be shaped through the lived experience of its people.

> *"I think, that if you actually look at the government of Australia, and the population in Australia, so we have a big migrant crowd here. So what is trust for them can be different, you know, so therefore, it is the responsibility of the government to increase the trust of different migrant communities with respect to AI, maybe perhaps starting from the schools, because when I migrated to Australia I learned things from my kids"*

A more diverse and inclusive representation to regulatory framework design. At the roundtable, our participants made observations that the current representation in government may not be reflective of the Australian population. If AI regulation is to work for all, the view is that the body ultimately responsible for the design of an AI regulatory framework should be one that has a diverse and inclusive representation.

> *"There is no diversity, this is the root of the problem. So you need to, like dive a little bit into the root of the problem if you want to solve the problem. And I don't think the government is really an example of diversity and inclusion"*

> *"Do our discrimination laws have enough information about our diverse society about the Aboriginal community being so small, are they so there are discrimination laws at the moment which force it, but are they equipped to deal with the bias that AI will bring? Is there equal representation of the numbers and, and statistics that currently represent the makeup of Australia and make sure that those biases do not occur?"*

A government body that shows that they understand the technology as well as the public. While the Hub is of the view that the government has managed the developments of AI well, including the establishment of the National AI Centre, RAIN and this discussion paper, we do find that the general consensus of the participants at the roundtable to be different. Some held the view that there is a need for the government members to be more in touch with the developments of fast-paced AI developments. Due to time constraints, we were not able to investigate this further but it could be a symptom of a broader 'trust' issue that was discussed. Perhaps, not only should we be looking at building trust to engagement of the public, there is also a need to show the community that the government is well-informed about a disruptive technology like AI.

> *"The first step is the government should educate themselves about the technology, and the second step is, of course, focus on educating the general public about it"*

We believe the government can play a leadership role to demonstrate their understanding of the technology. One initiative that the government can pursue is to look at providing guidance to the public around how the redesign of common job roles in the industry can be achieved to harness the benefits of AI while retaining a human-centred approach. Not only will this demonstrate the government's grasp of the technology to the public, but it will also create confidence in the industry (particularly small and medium enterprises) on how productivity gains can be achieved with AI. Furthermore, this exercise will help clarify areas of investments that government will need to support and subsequently, build clarity on

grant schemes to support SMEs in their digitalisation uplift. While this is not necessarily part of regulation per se, it is part of a response that we think is suitable for Australia.

An AI regulatory framework that demonstrates the Government's will to ensure a 'fair-go' for all Australian. As AI plays an increasingly important role in the digital economy, there are already signs that the development of AI will only exacerbates the inequality in the Australia society. While this won't be a problem that is isolated to Australia, there are real risks in the Australia society particularly at a time where the economy is undergoing a significant structural change due to the pandemic.

Our concern here is not only restricted to how AI may replace labour and change the employment structure but also how AI may not be accessible to the disadvantaged if access to AI is determined by one's social economic status. This divide can be a real possibility if our AI regulatory framework does not put in place systems that ensure access to key AI capabilities are not a function of one's social economic status.

The concern of a 'fair go' is not only restricted to individual Australians. With the current AI developments, there are signs of reliance on large institutions to provide the underlying AI capabilities. Will our AI regulatory framework ensure that our small and medium enterprise are given alternatives so that they are not locked into a specific AI service that limit a firm's economic competitiveness? Besides protecting our community from the downside risks of AI, should the AI regulatory framework to put in legislation that will give our local entrepreneurs a chance to compete?

4.4 Target areas

We form the view that we should differentiate between how AI is used in the public and private sector. This was also the views shared by participants at the roundtable event. The consensus is that while the public and private sectors share common operations, their operating goal is different, i.e., the private sector is accountable to the shareholders while in the public sector, the accountability is to the community it serves. In turn, this fundamental difference dictates how AI deployment should be managed.

One of the views put forward by participants at the roundtable is the fact that any use of AI within the public service is likely to come without the opportunity for the community to opt-out. For example, an AI-based model used at the airport to manage biosecurity and immigration risks, or an AI model to identify irregularity in tax lodgements at the ATO would all come without the option for individuals to opt-out. In that sense, many felt a stronger need to ensure that these systems emphasise more on establishing 'trust' and managing 'risks' given the lack of choice.

> *"When you're dealing with government agencies, they don't do user experience because they don't need you to approve of what they are doing. And so I think the managing AI risks are going to be very different to the public in the private sector"*

This is in-contrast to public sectors, where in delivering a product or service, there is an opportunity for an individual to opt-out – either by declining to accept the T&C, or through operation system environments, e.g., not allowing an app feature that contains the AI technology. In that sense, the degree of 'trust' that one needs to have for the system is not as stringent as that of systems used by public services. Furthermore, users can decide not to participate in a service or product by declining the T&Cs in the private sector. There is also the possibility of choice or options. This is very different to citizens interacting with the public service like the ATO, where options or opting out is not an option.

In the absence of choice, or the opportunity to opt-out, the need to address 'trust' becomes an important target area. There are many ways to earn 'trust' and one of them is the expected accountability of agencies deploying AI models for decision making. This can be in how AI decisions is transparent, or that the decision can be explained or understood by the receiver of the AI's decision.

> *"The accountability that you expect in public and private is different, and levels of explainability that you expect in public and private, again, are going to be quite different. And, and I certainly think public sector does need to be able to explain how and why a decision was made"*

The view from participants confirmed the difference in expectations between the private and public sector. This was again reflected among participants when discussing confidentiality requirements, another angle to 'trust'.

> *"When it comes to public and private use of AI, so there's going to be a different commercial confidentiality, I reckon in that domain, it's probably regular in the public sector, compared to the private sector. The private sector, they can always use excuses, is confidential. Whatever the calls, whatever data goes on behind us is confidential. So we cannot make it public. However, there is more reason for the private sector to disclose"*

Quite interestingly, we note that our participants felt that the public sector has used 'confidentiality' to avoid disclosure of the details of their AI systems. This perception that "it's an excuse" suggests that on this issue, the Freedom of Information Act that governs our public sector, has given the community a higher level of 'trust' on matters around confidentiality. We are not suggesting extending the FOI Act to the public sector nor we are of the view that we should have a specific FOI Act for them. We do not have an answer to this question as we recognised the complexities around industry verticals within the private sector. We therefore agree with the discussion paper that we need to not just consider AI regulation as a new piece of legislation but also, take upon reviews of sector specific laws with the view of bringing them up to date with the developments of AI.

Take the financial services sector for example. Currently, existing regulatory approaches in financial services may not comprehensively cover certain aspects of artificial risks. One area is the regulation of algorithmic decision-making and the potential for unintended

consequences, or systemic risks arising from the use of AI/ML systems. While ASIC can draw inspiration from the International Organization of Securities Commissions (IOSCO) to improve its algorithmic decision-making, we believe an AI regulatory framework supported by sector specific regulation will deliver more benefits to the community. For a start, this will eliminate definition differences of AI and sets the framework in which sector specific laws can be put in place.

This whole of regulation approach also means that the government is clear about its interest in 'responsible AI', which is already set in motion through the CSIRO National AI Centre's Responsible AI Network (RAIN). Therefore, on the question if the government should be taking a lead in implementing 'responsible AI', we believe we should. Not only because it's an important leadership to demonstrate to the private sector, but it also goes back to our point on developing 'trust' in the community for agencies in the public sector.

While the early signs are positive, we believe there is more to be done. So far, most initiatives out of RAIN have focused on developing the technical aspects of 'responsible AI'. This computer-science centric approach is necessarily to deliver the tools that enterprises in both sectors will need. However, we are of the view that the government can go further in leading the implementation of 'responsible AI' by

- putting in schemes or initiatives to encourage private sector enterprises in adopting these 'responsible AI' toolkits;
- building trust in the public sector through government lead implementations of 'responsible AI' patterns developed out of RAIN;
- engaging the community on these initiatives to help build understanding of AI as a piece of technology within the community; and
- developing a culture of shared responsibility on embracing a complex technology like AI or Machine Learning.

On the point above on building AI understanding within the community, we came to this conclusion from our experience as a research centre interacting with our industry partners and having drawn the same observation from our participants at the roundtable. That is, (i) there isn't a shared understanding of 'responsible AI' at best, and (ii) there were some interesting viewpoints as noted below.

> *"There is no such thing as responsible AI, it is not made by itself, we, as humans, design and use it"*

> *"At the end of the day, the fundamental responsibility of the government is to help drive trust in the system. And you think in anything that we use today, regulations exists so that we can trust. We trust that we get on an airplane and it's not going to crash and trust starts with principles of how it's built. How cooperated. And more importantly, if things go bad, what are the consequences? So trust in the government, all of the government uses to help establish trust"*

On the issue of 'risks', we believe the 'traffic light' system proposed in the discussion paper is a good place to start. Its greatest strength is in its ease of understanding and a sensible approach to AI governance, comprising of self-assessment to enforced obligation depending on the perceived risk level. The downside to this is how an AI artefact is perceived at a specific level of risk. While examples are good, they are not exhaustive. A given risk level is subjected to dispute, and how a risk level is determined can be subjected to challenge. Thus, any AI regulator will need to develop a well-thought implementation plan to the risk management system.

If we get the risk management system right, we believe there shouldn't be a need to manage risks differently. We take this view because purpose-built AI systems are likely to use off-the-shelves AI packages or Cloud services. These off-the-shelves packages/service is likely to be developed across different geographical regions and therefore, are subjected to different levels of AI regulation (if there is one in place). Therefore, trying to have a different set of risks management is, in our view, not a fruitful exercise.

Take ChatGPT as an example, which would be considered off-the-shelves. However if an organisation decides to use ChatGPT's API with an in-house AI model to create a purpose-built system, trying to assess them different is not likely to make sense. Instead, our view is that there should be a way to cascade the AI risks alongside a set of disclosure requirements (i.e., disclosure of what off-the-shelves components were used) to go with the traffic light system. This ability to cascade risks means systems built with off-the-shelves packages/services will have a baseline risk level that further assessment of the risks of the purpose-built system can be undertaken.

Here, the major concern is obviously the 'high risks' applications. On the point of whether such applications should be banned or regulated, we hold the view that the risk level associated with an AI system should evolve overtime in the consumer space. Therefore, regulation would be a better approach as a 'ban' is likely to impede progress or limit Australia's ability to compete globally. This is also a view that is seen in the roundtable as reflected in their various comments.

> *"I don't think autonomous vehicle driving should be banned. I think you if we allowed it to be on the road, the authority responsible for these things to apply the risk based principles to safety principles and say is it safe to be on the road doesn't need to be an AI around a road safety issue"*

> *"I think blanket bans are probably not there. I think machine doing ethics is a great idea. You know, I think we think about an AI providing counsel as being really, really bad. But I can see good use cases of it. Because we all know humans are not flawless. Even good practitioners can go through a bad period, right? So the risk of something bad happening by human practitioners is equally there. And so, therefore, we can't just keep blaming machines. Machines are probably more objective. They can crunch a lot of data, and can be more standardised. You can control them a lot better than you can control humans. So I think therefore, I'm not very intent on banning"*

> *"I think if AI is used for any kind of decision-making, then it's medium. And if AI is used for anything to do with the health or safety of individuals, then that becomes high risk. And the thing that they actually proposed that anything high-risk has to have a third-party audit. So I think it's going in the right direction when you take risks into account"*

Our emphasis here is that risks level change for any systems as technology evolves rapidly. Therefore, risk management or assessment should not be a one-off but instead, have a timeframe in which an assessment applies. Risk levels may change due to technological advances or a change in the operating environment in which the AI system exists. Therefore, our view is that risks management should have a 'valid until' date, and that the validity of such a date should vary according to the risk level.

4.5 Implications and infrastructure

Earlier, we discussed 'high risks' applications in the context of the consumer space. We made the qualification as we believe there are 'high risks' applications where a traffic light risk management system may be inadequate. We have seen how disruptive events like 9-11 changed aviation, or how the recent pandemic transformed the use of technology. In that sense, there should be consideration for government and in AI regulation, where specific assessment should be made on grounds of the 'public good', whether that is in community safety, mobility, or economic prosperity.

To some extent, the examples raised in the Paper is already embedded within the community with very specific use-cases, as compared to its use for a different or undeclared purpose. As the Paper rightly though implicitly pointed out, any considerations for a ban should be on the risks level of the specific 'activity' rather than the AI technology itself. By extension, because the consideration for ban is on the activity itself, we believe the evaluation criteria established within the risk management system for AI will be different though, we would agree that there is capacity for any AI risk assessment to inform decisions around the banning of a specific activity.

We also see similar views reflected at our roundtable, where participants discussed the balance that is calibrated overtime between the advantages and risks that AI (or any technology) mature. This also includes the need for 'time' for the public to embrace and understand the perks and perils of AI as we go forward.

> *"First of all, you know, there have been stigmatism issues that we are not talking about the technologies learn over time, they become more mature, you know, maybe AI will become less the responsibility. So don't worry, the risks will become less and less as the technology matures, essentially:*

The question of 'time' will in turn depend on the community's confidence in regulatory guidelines, initiatives such as what has been carried out in RAIN in cultivating a culture of 'responsible AI' in the public and private sectors, etc.

"The trade-off between risk and explainability. You're allowed to use them, as long as you can perfectly explain exactly why this is made"

"But yes, policing, transparency, user testing, governing the laws governing the data, and algorithms, proper testing ethics. I think these are these are the key things that we can do to make sure that we reduce the high-risk"

4.6 Risk-based approaches

In addition to our commentary on risks already made, through our roundtable discussion, we unveiled potential risk-based approaches that can inform the development of 'responsible AI'. They would complement the computer-science approach that RAIN is currently developing. We believe this mix of approaches will be needed to address community concerns as AI permeates into different parts of our society.

Principles-Based Approach: We found a strong agreement, whereby, participants emphasised on the respect for autonomy, nonmaleficence, beneficence, justice, and explicability as fundamental AI design and use principles. They argued that these principles aren't isolated; rather, they form a synergistic network that contributes to responsible AI usage. As AI continues to evolve, adherence to these principles is crucial to ensure that AI technologies are developed and used responsibly, for the benefit of all.

Transparency and Explainability: Another approach that gained much traction was the demand for increased transparency and explainability in AI systems to reduce misuse and improve accountability. Participants indicated the need for robust monitoring systems, regulatory measures, and third-party audits to ensure AI accountability, a risk-based strategy to prevent potential misuse.

"I think there needs to be more transparency around the input data that is used, because we're talking about transparency and mitigating risks. So, transparency of input data and the weighted measures, because a lot of times, so for example if you collect data only from New South Wales and Victoria, and then you inform AI policies for overall Australia, then you're missing out on a large. But also, then, I think, going into the technical stack, there are different weighting parameters. And I think there needs to be more transparency around that. Because you know, whatever your input data, if your weighted scores are not transparent, then you can create bias. I think they are very obvious."

Implementing principles of privacy-by-design (PbD) to reduce data privacy risks: A key theme that emerged was the importance of integrating the principles of Privacy-by-Design (PbD) to mitigate data privacy risks effectively. The discussion underscored PbD not just as a concept but as a proactive methodology that firmly places privacy at the heart of system design. This forward-thinking approach is aimed at eliminating privacy risks before they even emerge and providing optimal privacy assurances to all stakeholders.

The insights from the roundtable workshop suggested that addressing privacy as a primary design objective, rather than as an afterthought, presents significant advantages. First, it foregrounds the importance of privacy and ensures it is not compromised or neglected due to other design considerations. Second, participants noted that when privacy is integrated from the onset, it can minimize costly re-engineering or modifications that may arise due to privacy issues identified later in the development process. These sentiments align with regulatory requirements more effectively. Many jurisdictions around the world, such as the EU with its General Data Protection Regulation (GDPR), have legal requirements to consider privacy in the early stages of system design. Thus, adopting PbD can help ensure compliance and avoid potential legal complications.

> *"We've heard about, we need to consider the potential for harm, we need to focus on the outcomes of where the technology is applied, and we need to consider not just the technology, but the institutions and decisions around technology."*

Additionally, one of the attendees cited an example to illustrate the same,

> *"Our unique thing, we'll say, if you're sending 10 people's data, you have 10 unique IDs for that, because that way it reduces the risk"*

Collaborative Governance: Further, collaborative governance was recognised as a necessary risk management strategy. Participants emphasised the importance of cross-border and cross-sector cooperation among various stakeholders, including governments, industries, academia, and civil society. For instance, the European Commission's High-Level Expert Group on Artificial Intelligence, a diverse group of stakeholders from academia, industry, and civil society, was tasked with drafting AI ethics guidelines that have influenced AI policy worldwide. On this, the following experts quoted,

> *"It's global. It's dynamic. It's constantly changing."*

> *"This is where the governance role and all the stakeholders come into play, to say like, this is a trust entity in an organization, and every set of outputs that we will develop or data products will develop, need to go to that entity."*

Ethics Training: Last but not the least, one major approach that was consistently and strongly voiced at the roundtable was training AI practitioners and users on the ethical design and application of AI is critical for mitigating AI risks.

> "*The question is not whether or not machines can think; rather, it is whether or not humans can think."*

There is emphasis that users and developers are primarily responsible for the ethical deployment of AI, not the technology itself. This has been widely recognised by educational institutions around the globe. For instance, the Institute for Human-Centered Artificial Intelligence at Stanford University, offers an interdisciplinary AI ethics course, demonstrating

that academia is also recognising the significance of this training. Similarly, RMIT has incorporated Information Systems Ethics training (including AI Ethics) into a number of its undergraduate and graduate foundation courses.

This demonstrates that comprehending the ethical implications of AI is not a luxury, but rather a necessity. Training AI practitioners and users on these ethical considerations can help mitigate risks by ensuring that AI systems are designed and used responsibly, with the well-being of all affected stakeholders as the top priority.

5. Other considerations

In addition to the questions posted by the Paper, we felt that the following considerations should also be taken into account for any AI regulatory framework given the potential disruptive nature of the technology to humanity.

5.1 Environmental, Social, and Governance (ESG), or Corporate Social Responsibility (CSR)

Corporate actions in the arena of social welfare are often referred to as Environmental, Social and Governance, or Corporate Social Responsibility (Gillan et al., 2021). ESG/CSR is commonly used as a framework to assess an organisation's business practices and performance on various sustainability and ethical issues. The ESG factors span from greenhouse gas emissions, water usage, waste management, employee relations, diversity and inclusion, and human rights to governance metrics such as board composition, executive compensation, risk management, and ethical considerations (Safari et al., 2021; Safari & Areeb, 2020; Safari & Parker, 2023). Investors are increasingly seeking out companies that are committed to sustainability and ethical practices based on ESG/CSR criteria to avoid holding companies engaged in risky or unethical practices.

Artificial Intelligence (AI) and ESG/CSR

In the environmental and social domain, AI can be deployed to optimize energy consumption, reduce waste generation, enable predictive maintenance, enhance environmental monitoring, and also can contribute to social inclusivity. The positive deployment of Artificial Intelligence (AI) could help achieve ESG/CSR intended goals, by ensuring equal access, reducing bias in decision-making, and enabling personalized experiences. However, the misuse of AI could also pose various threats in areas such as privacy and security concerns, data and unintended biases, job displacement, and ethical decision-making. Proper governance and ethical considerations appear to be crucial in AI algorithm development and AI deployment to ensure transparency, accountability, and fairness of AI and responsible use of AI.

One of the key challenges with AI seems to be privacy and security considerations. AI raises concerns related to data privacy, security, and potential misuse. It is deemed that establishing robust data protection frameworks and ethical guidelines is essential to address these challenges. AI systems also heavily rely on training data, which may reflect existing biases. Efforts should be made to ensure quality, diverse and representative datasets to avoid perpetuating societal inequalities. In addition, while AI has the potential to create new

job opportunities, certain roles may become automated, leading to job displacement. Reskilling and upskilling employees should be implemented to mitigate the impact on the workforce. Furthermore, ensuring ethical decision-making by AI systems is another complex challenge. Our roundtable data indicates that it would involve developing ethical frameworks, implementation of accountability mechanisms, and human oversight associated with AI Systems, especially for high-risk applications involving vulnerable groups.

5.2 Sustainable Development Goals (SDGs)

Sustainable Development Goals comprise 17 interlinked objectives designed to serve as a "shared blueprint for peace and prosperity for people and the planet, now and into the future". The goals were adopted by all United Nations in 2015 and are intended to be achieved by 2030. While SDGs propose an ambitious agenda, they provide a framework for action that can help us to address the most pressing challenges facing our world. The goals include no poverty, zero hunger, good health and well-being, quality education, gender equality, clean water and sanitation, affordable and clean energy, decent work and economic growth; industry, innovation and infrastructure; reduced inequalities; sustainable cities and communities; responsible consumption and production; climate action; life below water; life on land; peace, justice and strong institutions; and partnerships for the goals. Our roundtable participants underscored the importance and opportunities associated with the use of AI for social good.

> "*I guess I don't want to just be focused on the no-harm aspects, but also in terms of how we can incentivize* [AI for social good]."

Artificial Intelligence (AI) and SDGs

Similar to ESG/CSR domain, while the positive deployment of Artificial Intelligence (AI) could perform as an accelerator for the support and achievement of UN Sustainable Development Goals, the misuse of AI could equally pose some threats and bring along some negative impacts on Sustainable Development Goals. Sustainable AI development is an emerging term promoting the responsible use of AI for long-term benefits to the public globally. Sustainable, green, and responsible AI is expected to enable empowerment and achievement of SDGs. For example, AI could be used to facilitate climate actions, biodiversity conservation, and enhance cultural interactions.

> *"It becomes a policy question. Right. So how do governments regulate technology? And how do they integrate them in a way that is fair and democratic?"*

The Roadmap for Digital Cooperation report (United Nations, 2020) provides a vision for a digitally interdependent world. It lays out goals and actions for the global community to connect, respect, and protect people in the digital age, with some goals expected to be achieved by 2030 with a collective global effort. The report stressed that while digital technologies could have enormous potential for positive change but can also worsen economic and other inequalities and increases worldwide data breaches, spreading disinformation and hate speech, online harassment and violence, and other complex issues significantly. The roadmap is deemed to be a living document that will be updated as the

digital landscape evolves. It is an important tool for ensuring that digital technologies are used for good, not for harm. While recommending governments as the main actors at the centre, digital cooperation emphasizes the importance of a multi-stakeholder effort for making realistic and effective decisions and policies to seize on the opportunities that are presented by technology while mitigating the risk.

The participants at the roundtable appear to be in agreement with taking a multi-stakeholder approach to the use of digital technologies for good. One participant stated,

> "*I think the technology is actually allowing more individuals to coordinate and solve broader social issues.*"

The roadmap identifies eight key areas for action: Achieving universal connectivity by 2030, digital public goods, digital inclusion, digital capacity-building, Digital human rights, artificial intelligence, Digital trust and security, and global digital cooperation.

Achieving the universal connectivity goal requires every person should have safe and affordable access to the Internet by 2030, including meaningful use of digitally enabled services in line with the Sustainable Development Goals.

The digital public goods goal promotes creating a more equitable world. In addition to suggestions for undertaking a collective global effort to encourage and invest in the creation of digital public goods such as open-source software, open data, open AI models, open standards, and open content, it is recommended to ensure that these digital public goods do no harm and help achieve the SDGs.

The digital inclusion goal ensures mitigating divisions and including all and the most vulnerable. It is recommended to close the gaps through better metrics, data collection, and coordination of initiatives. The role of governments in the protection of vulnerable users also emphasized by our roundtable participants.

The digital capacity-building goal aims to strengthen capacities and skills crucial to the digital era and to attain the SDGs. Digital capacity building is recommended to be more needs-driven and tailored to individual and national circumstances and coordinated globally.

The digital human rights goal aims to ensure the protection of human rights in the digital era. It is deemed that while digital technologies could provide new means to exercise human rights, they are also repeatedly used to violate human rights. It is suggested that regulatory frameworks and legislation on the development and use of digital technologies should have human rights at their centre. Some considerations in the digital human rights area comprise concerns regarding data protection, digital ID, the use of surveillance technologies, online harassment, and content governance. The protection of human rights and ensuring no harm are also mentioned by our roundtable participants.

> *"So I think [the best approach would be] policing, transparency, user testing, governing the laws governing the data, and algorithms, proper testing ethics. I think these are the key things that we can do in making sure that we reduce the high risk."*

The artificial intelligence goal requires multi-stakeholder efforts on global AI cooperation to help build global capacity for the development and use of AI in a manner that is trustworthy, human rights-based, safe and sustainable, and promotes peace.

The digital trust and security goal promotes trust and security in the digital environment. It is suggested that the digital technologies that underpin core societal functions and infrastructure, including supporting access to food, water, housing, energy, health care, and transportation, need to be safeguarded. The roundtable participants believed that cyber risk and cybersecurity need to be taken seriously at the government level.

It is also pointed out by roundtable participants that there is a need for more transparency around how AI models are built and used to mitigate risks and build trust. This includes transparency around input data, weighting parameters, and the technical stack.

The global digital cooperation goal requires building a more effective architecture for digital cooperation. As a starting point, the Internet Governance Forum is recommended to be strengthened, in order to make it more responsive and relevant to current digital issues.

In all, the roundtable data revealed that while AI may transform some jobs, and if used responsibly it has the potential to expand human creativity and productivity. But policies are needed to manage this transition. In addition, banning specific AI applications may not be effective on its own. The incorporation of ethical considerations in AI governance especially when AI is interacting with vulnerable groups could help mitigate some risks. These may involve proper testing and requirement for certifications, and human oversight of high-risk uses as likely more effective approaches.

5.3 Human-centred approach

A further area in which there are concerns about the impact of automated decision making on humans is in the area of algorithmic management. The use of software algorithms to automate organisational functions traditionally carried out by human managers has been termed "algorithmic management". Research into algorithmic management focuses on software algorithms, defined as 'computer-programmed procedures for transforming input data into a desired output'. It entails human jobs being assigned, optimized, and evaluated through algorithms. The capacity of algorithmic management has expanded with the increased skill levels of machine learning.

Algorithmic management has been researched in greatest detail in the settings of platform work and warehousing but also noted to a lesser extent in retail, manufacturing, marketing, consultancy, banking, hotels, call centres, and among journalists, lawyers and the police.

Level of automation in decision making (Wood, 2021)

Level of automation	**Narrative definition**	**Direction, Evaluation, Discipline**	**Review** (in case of system failure)	**Mode specific** (human manager can ignore/ overrule system)
No automation	Full-time performance by *human manager* of all aspects of direction, evaluation and discipline	Human manager	Human manager	n/a
Management Assistance	Assistance in either direction, evaluation **or** discipline **with the expectation** that *human managers* perform other management tasks **and** use own judgement to review, ignore and overrule system.	Human manager **and** algorithmic system	Human manager	Yes
Partial Automation	Mode specific execution of either direction, evaluation **or** discipline **with the expectation** that *human managers* perform remaining functions.	Algorithmic system **or** human manager	Human manager	Yes
Algorithmic management				
Conditional Automation	Mode specific execution of direction, evaluation **and** discipline **with the expectation** that *human managers* will respond appropriately to a request to intervene.	Algorithmic system	Human manager	Yes
High Automation	Full-time performance by an algorithmic system of direction, evaluation **and** discipline **without the need** for *human managers* to intervene.	Algorithmic system	Algorithmic system	Yes
Full Automation	Full-time performance by an algorithmic system of direction, evaluation **and** discipline **without the possibility** for *human managers* to intervene.	Algorithmic system	Algorithmic system	No

The existing evidence suggests that algorithmic management may

- Curtail the scope for decision making of low-level managers and supervisors and confining their organisational function to offering workers encouragement and support in navigating the algorithmic direction and evaluation.
- Accelerate and expand precarious fissured employment relations (via outsourcing, franchising, temporary work agencies, labour brokers and digital labour platforms).
- Worsen working conditions by increasing standardisation and reducing opportunities for discretion and intrinsic skill use. Evidence from platform work and logistics highlights the danger of algorithmic management intensifying work effort, creating new sources of algorithmic insecurity and fuelling workplace resistance.
- Make management decisions less transparent or accountable. Various studies have found performance systems in algorithmic management to be 'black boxes' to workers, making platform workers feel compelled to accept jobs. For example, one worker interviewed said:

> *"I've never cancelled any of them. I try to accept all of them because if you drop below 85% [the acceptance rate], then you'll – I don't know – something happens like you have to be [stood] down for a while. Sometimes the app will crash [while] accepting a job, so some of them can … [reject] but not from my doing." (Interview 31) (Veen et al., 2020)*

- Reduce worker discretion in choosing how to undertake their job as well as limited discretion over the ordering of their day-to-day tasks.
- Increase workforce stress and anxiety and be harmful to wellbeing and health. In particular, in location-based platform work (e.g. rideshare, food delivery), AI systems of surveillance and performance management are often utilised to achieve what the UK Supreme Court in *Uber BV and others (Appellants) v Aslam and others (Respondents)* [2021] UKSC 5 described as 'a classic form of subordination' - typical elements of which include: the use of passenger ratings to rank drivers (a factor which the algorithm then considers when allocating trips), with consistently low customer ratings leading to deactivation from the app; and the monitoring of a driver's level of acceptance or cancellation of trips, creating pressure to accept a high proportion of trips or be automatically logged off from the app (as outlined above).

Possible policy responses to this problem include

1. A requirement for firms to report on the impacts of digital technologies on jobs, wages and the quality of work.
2. An individual right for the data subject to have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her, as is the case in Article 22 of European Union Regulation (EU) 2016/679 (General Data Protection Regulation) (Burgess, 2019)
3. A requirement for human reviewers to be involved in checking the system's recommendation and not "routinely" apply the automated recommendation to an individual; and that the reviewers' involvement must be active and not just a token gesture. The reviewer should have "meaningful" influence on the decision, including the "authority and competence" to go against the recommendation; and reviewers must 'weigh-up' and 'interpret' the recommendation, consider all available input data, and also take into account other additional factors' (Information Commissioner Office, 2019).
4. The development of collective rights that protect workers from algorithmic management in the Fair Work Act, Awards and Enterprise Agreements. Specifically, in relation to platform work, a first step would be to override through legislation the High Court of Australia rulings in *ZG Operations v Jamsek* [2022] HCA 1 and *Construction, Forestry, Maritime, Mining and Energy Union v Personnel Contracting Pty Ltd* [2022] HCA 2. These decisions require courts and tribunals to defer to the formal written terms entered into between platforms and workers, which in most cases are imposed by the platform without negotiation, as a basis for legally categorising the worker as an independent contractor. These rulings have been applied in *Deliveroo Australia Pty Ltd v Franco* [2022] FWCFB 156 and *Nawaz v Rasier Pacific Pty Ltd T/A Uber B.V.* [2022] FWC 1189, to deny gig workers employment rights including the ability to bring unfair dismissal claims (contesting the use of AI systems in decisions leading to dismissal). Redefining the definition of 'employee' in the *Fair*

Work Act 2009 (Cth) to encompass these workers would give them not only unfair dismissal protections but also access to awards and collective bargaining, enabling unions to negotiate a stronger framework of AI rights in the gig economy. The federal government's current proposals to give the Fair Work Commission powers to set minimum standards for 'employee-like' workers may also assist in this area.

5.4 Improving public trust in AI deployment – a cybersecurity perspective

Integrating AI into various aspects of our lives can bring numerous benefits, however, the concerns of public trust in AI deployment have also hindered the adoption of AI. We thus need to improve public trust in AI to boost the widespread adoption and acceptance of AI in our society. From a cybersecurity perspective, ensuring the confidentiality, integrity, and availability of AI systems is crucial in building and maintaining public trust. This section examines the key factors and approaches for improving public trust in AI deployment, with a specific focus on cybersecurity considerations.

Implementing Robust Authentication and Authorization Mechanisms: Once an AI system has been deployed, one must ensure that only authorized individuals/entities can access and interact with the system. Implementing strong authentication and authorization mechanisms can prevent unauthorized access, tampering, or manipulation of AI algorithms and confidential data. Multi-factor authentication, access controls, and secure user management practices are essential components of an effective cybersecurity strategy when deploying AI systems.

Ensuring Data Privacy and Protection: One of the primary concerns associated with AI deployment is the security and protection of user data. Cybersecurity measures are instrumental in guaranteeing the secure collection, storage, and processing of sensitive information. Robust encryption protocols, proper access control mechanisms, differential privacy and data anonymization techniques must be implemented to protect the data privacy. In addition, transparent data governance practices and compliance shall also be applied to further enhance the data protection in AI systems.

Enhancing AI System Security: Securing AI systems from cyber threats is crucial in maintaining public trust. AI systems are known to be vulnerable to adversaries seeking to manipulate their outputs by crafting adversarial examples or exploit vulnerabilities to steel information either about the model itself or the training data. To mitigate such risks, it is essential to employ cybersecurity measures, such as penetration testing, vulnerability assessments, and secure coding practices to identify and remediate such weaknesses. Continuous monitoring, incident response plans, and security audits increase the resilience of AI systems, which can further improve trust in their reliability and integrity.

Mitigating Algorithmic Bias and Fairness: Algorithmic bias, leading to discrimination by AI systems based on factors such as race, gender, or other characteristics, can pose a significant threat to public trust in AI. Cybersecurity mechanisms can help mitigate bias by ensuring the fairness and transparency of AI algorithms. Conducting regular audits, utilizing bias detection

tools, and incorporating diverse training data sources are effective approaches to identify and rectify bias within AI models. Through actively addressing bias, promoting fairness and inclusivity of AI, we can ensure AI's equitable treatment of all individuals and generate the public confidence in AI systems.

Ensuring Ethical and Transparent AI Practices: Transparency and ethical considerations form the foundation of public trust in AI. Cybersecurity professionals and researchers play a vital role in ensuring that AI systems comply with ethical guidelines and standards. They can help Implement transparency measures, such as explainable AI techniques to enable users comprehend the decision-making process of AI systems, which can build trust and promote accountability. Effective collaboration among cybersecurity experts, AI developers, and regulatory bodies is also essential in establishing ethical AI practices that align with public expectations.

Establishing Regulatory Frameworks: Regulatory frameworks that govern AI deployment are instrumental in ensuring public trust. Cybersecurity experts can make significant contributions to the development of these frameworks by offering insights into potential security risks and vulnerabilities associated with AI systems. Embedding the cybersecurity experts' insights can enable the creation of effective regulations that balance the AI innovation and security. Through these regulatory and governance frameworks, public concerns can be effectively addressed, leading to enhanced trust in AI systems.

6. Summary

In response to provide feedback to the Government's Paper, the RMIT Enterprise AI and Data Analytics Hub ran a roundtable event to hear from its stakeholders on the various questions put forward in the Paper. With over 40 attendees and the contributions of academics in their own research domain, we present our views to the questions raised, with key findings as follows.

- Given the diverse representations, we did find broad agreement to the various AI definitions presented in the Paper.
- There is a consensus that the Government's interest to develop "responsible AI" is a step in the right direction, with emphasis that the focus should be on its use rather than the technological aspects of AI.
- Besides government owned/supported AI audit bodies, substantial importance should be given to the existing regulatory and legislative regimes to ensure responsible AI, i.e., it must be a collective effort.
- Those involved in the development of this response felt that the best way to ensure safe and responsible AI is a multi-prone approach, e.g., having humans in the loop, establishing independent AI testing body, encouraging diversity at each step of the AI lifecycle, and consideration of factors that is unique to Australia, e.g., diversity of its population.

- Consensus is also found in different considerations to private and public sectors, given the nature of their operations and the intent of using AI.
- There is a broad agreement that risk management is important, and that the dynamic nature of AI applications will require further thought on how this can be best managed.

Despite with varying degree, we are delighted to observe the broad positive development in the mindset of stakeholders, particularly the consensus in the separation of AI as a technology versus its use. The technology itself is not responsible but how it will be used is the question that AI regulation will need to answer.

Contributors and acknowledgement
This submission is the collective effort of academics, industry board members within RMIT University, and industry participants. We acknowledge that the arguments, views or recommendations presented in this Paper may not fully reflect any individual's views. This is expected given the number of participants and the nature of such a submission, which we have prioritised a coherent submission.

Contributing authors
- Kok-Leong Ong, Professor and Director, Enterprise AI and Data Analytics Hub
- Samar Fatima, Research Fellow, Enterprise AI and Data Analytics Hub
- Abhinav Shrivastava, PhD Candidate, Enterprise AI and Data Analytics Hub
- Anthony Forsyth, Distinguished Professor, Graduate School of Business and Law
- Darcy Allen, Senior Research Fellow, Blockchain Innovation Hub
- Chris Berg, Director, Blockchain Innovation Hub
- Shelley Marshall, Director, Business and Human Rights Centre
- Marta Poblet Balcell, Professor, Graduate School of Business and Law
- Maryam Safari, Senior Lecturer, Department of Accounting
- Humza Naseer, Information Systems, Enterprise AI and Data Analytics
- Sophia Duan, Information Systems, Enterprise AI and Data Analytics
- Angel Zhong, Associate Professor, Department of Finance
- Chao Chen, Deputy Director, Enterprise AI and Data Analytics

Academics and Industry participants
- Robert Wickham, Vice-President and COO, Salesforce APJ
- Brad Kasell, Principal Technology Strategist, Domo Inc
- Jason Ball, Assistant Director, Foreign Investment Division, Commonwealth Treasury
- Ramah Sakul, Head of Government Affairs, SAP ANZ
- Yogesh Nerurkar, CEO, Infonyx Pty Ltd
- Buddhi Jayatilleke, Chief Data Scientist, Sapia.AI
- Madhura Jayaratne, Lead Data Scientist, Sapia.AI
- Geoff Brim, President, Brimwood Consulting

- Matthew Cooke, Senior Director of Technology, NTT Data
- Melroy Vanlangenberg, Head of Data Governance, Australian Unity
- Ryan Cushen, Analytics and Data Science Lead, Prezzee
- Fei Tony Liu, CEO, Bijak.AI
- Stuart Thomas, Director of Learning, RMIT Digital3
- Jason Murphy, Business Analyst, RMIT University
- Steven Li, Professor of Economics, RMIT University
- Robyn Moroney, Professor of Accounting, RMIT University
- Emmanuelle Walkowiak, Vice Chancellor's Senior Research Fellow, RMIT University
- Say Yen Teoh, Senior Lecturer in Information Systems, RMIT University
- Sonia Magdziarz, Lecturer in Accounting, RMIT University
- Ariel Kam Ha Lui, Lecturer in Information Systems, RMIT University
- Yiliao (Lia) Song, Research Fellow, Enterprise AI and Data Analytics Hub
- Nataliya Ilyushina, Research Fellow, Blockchain Innovation Hub
- Vy Nguyen, Research Fellow, Blockchain Innovation Hub

References

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B., Anderson, H., Roff, H., Allen, G. C., Steinhardt, J., Flynn, C., hÉigeartaigh, S. Ó., Beard, S., Belfield, H., Farquhar, S., … Amodei, D. (2018). *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation* (arXiv:1802.07228). arXiv. https://doi.org/10.48550/arXiv.1802.07228

Burgess, M. (2019, January 21). What is GDPR? The summary guide to GDPR compliance in the UK. *Wired UK*. https://www.wired.co.uk/article/what-is-gdpr-uk-eu-legislation-compliance-summary-fines-2018

Business and Human Rights Centre. (2023). *Business and Human Rights Centre*. https://www.rmit.edu.au/research/centres-collaborations/business-and-human-rights-centre

Caplan, R., Donovan, J., Hanson, L., & Matthews, J. (2018, April 18). *Algorithmic Accountability: A Primer*. Data & Society; Data & Society Research Institute. https://datasociety.net/library/algorithmic-accountability-a-primer/

Costanza-Chock, S. (2018). *Design Justice: Towards an Intersectional Feminist Framework for Design Theory and Practice* (SSRN Scholarly Paper No. 3189696). https://papers.ssrn.com/abstract=3189696

Gillan, S. L., Koch, A., & Starks, L. T. (2021). Firms and social responsibility: A review of ESG and CSR research in corporate finance. *Journal of Corporate Finance*, *66*, 101889. https://doi.org/10.1016/j.jcorpfin.2021.101889

Information Commissioner Office. (2019). *Automated Decision Making: The role of meaningful human reviews | Information Commissioner's Office*. https://www.wired-gov.net/wg/news.nsf/articles/Automated+Decision+Making+the+role+of+meaningful+human+reviews+15042019122500?open

Müller, V. C. (2021). Ethics of Artificial Intelligence and Robotics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2021). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2021/entries/ethics-ai/

Safari, M., & Areeb, A. (2020). A qualitative analysis of GRI principles for defining sustainability report quality: An Australian case from the preparers' perspective. *Accounting Forum*, *44*(4), 344–375. https://doi.org/10.1080/01559982.2020.1736759

Safari, M., & Parker, L. D. (2023). Understanding Multiple Accountability Logics Within Corporate Governance Policy Discourse: Resistance, Compromise, or Selective Coupling? *European Accounting Review*, *0*(0), 1–30. https://doi.org/10.1080/09638180.2023.2194028

Safari, M., Tsahuridu, E., & Lowe, A. (2021). Big4 responses to the COVID-19 crisis: An examination of Bauman's moral impulse. *Accounting, Auditing & Accountability Journal*, *35*(1), 131–145. https://doi.org/10.1108/AAAJ-08-2020-4818

United Nations. (2020). *Road map for digital cooperation: Implementation of the recommendations of the High-level Panel on Digital Cooperation: Report of the Secretary-General | Policy Commons*. https://policycommons.net/artifacts/72465/road-map-for-digital-cooperation/141436/

Veen, A., Barratt, T., & Goods, C. (2020). Platform-Capital's 'App-etite' for Control: A Labour Process Analysis of Food-Delivery Work in Australia. *Work, Employment and Society*, *34*(3), 388–406. https://doi.org/10.1177/0950017019836911

Wood, A. J. (2021). *Algorithmic management consequences for work organisation and working conditions* (Working Paper No. 2021/07). JRC Working Papers Series on Labour, Education and Technology. https://www.econstor.eu/handle/10419/233886