ACS response

Safe and responsible AI in Australia discussion paper

July 2023

Australian Computer Society Inc. (ACT)
ARBN 160 325 931
National Secretariat
Tower One, 100 Barangaroo Avenue, Sydney NSW 2000
PO Box Q534, Queen Victoria Building, Sydney NSW 1230
T +61 2 9299 3666 | F +61 2 9299 3997
E info@acs.org.au | W www.acs.org.au

To the Department of Industry, Science and Resources, the Australian Government

ACS response
Safe and responsible AI in Australia Discussion paper

26th July 2023

Dear Sir or Madam.

Thank you for the opportunity to contribute to this important issue.

The Australian Computer Society (ACS) is the peak professional association for Australia's information and communications technology sector. We represent over 40,000 members working in all sectors of the economy and in all states and territories across the nation.

The ACS works to grow the technology sector while making sure IT professionals act ethically, responsibly, and in keeping with the best interests of not only their employers, but the wider community.

This review clearly comes at a critical time, when many in the business and wider community are finally waking up to the full potential of the risks and opportunities presented by AI. Ensuring that the good outweighs potential harms will put Australia on the right track for the future.

We would urge the government to look towards a model of careful progress. Vitally, we would urge consideration of not just risks and negative effects (which we imagine will be the focus of many responses to this paper), but also to ensuring that Australian organisations large and small are empowered and encouraged to govern their own applications of AI with transparency, responsibility, care and fairness.

We don't know what the future holds for AI, but if Australia's organisations have the right frameworks and governance in place to manage them, then we have the best opportunity to ensure that the Australian people can benefit from this new technology.

In the following pages, we've outlined some of the broader issues at play, and offer a number of suggested government interventions to make sure we're on the right track. This response was developed by the expert advisory panel of the ACS AI Ethics committee and led by Professor Peter Leonard, principal of Data Synergies and a Professor of Practice at UNSW Business School.

We'd be happy to discuss it further. If you would like to discuss any part of this response or simply seek further clarification or input, please feel free to contact myself by email at troy.steer@acs.org.au or by phone on 0417 173 740.

Yours sincerely

Troy Steer
Director of Policy, Advocacy and Communications
Australian Computer Society

Introduction

Few technologies will have a greater impact on Australian society than AI. The response of government to AI today is likely to shape the nation for years to come, for good or bad.

During the course of this review, the Department of Industry, Sciences and Resources is likely to get a huge number of responses outlining the risks and appealing for the government to apply bans, blocks or requirements to 'high-risk' and even regular AI activities.

While guardrails are critical, there is a danger of over-regulation, and of writing laws today that impact technologies that we don't even know about yet. The current EU laws, for example, were drafted before ChatGPT and Stable Diffusion brought generative AI into the limelight. We have no idea what AI technologies will be around in 10, 15 or 20 years, so trying to regulate them now will be at best challenging.

We believe the government needs to take a broader approach, beyond blacklists and risk assessment requirements. There is a need to look at the broader context of governance, of workforce skills, and of education. *Everybody,* from small business owners to billion-dollar corporations, will be using AI, and will need to have the knowledge, skills and motivation to deploy and employ that AI responsibly.

Summary and key recommendations

This response argues for a number of key considerations:

- Guardrails are important, and there should be context-specific rules about high-risk applications of AI.
- There should be a considerable focus on incentivising, enabling and assisting organisations (large and small) to develop competencies and capabilities with respect to AI governance and the ethical, transparent and safe use of AI. A model of enforced and assisted self-regulation can deliver that.
- Prescriptions for safe and responsible AI should focus upon the contexts in which AI is used by people in processes of decision making by organisations operating in Australia, and not exclusively focussed upon AI itself. Regulated prescriptions and requirements have a role to play, but are not the full picture.
- This consultation should adopt a primary objective of improving the reliability, quality, safety and accountability of decisions made by organisations operating in Australia where those decisions have been assisted or influenced by uses of AI.

In terms of specific government interventions, we are recommending three types of intervention:

Top down:

- higher level guidance and guidelines as to safe and responsible deployments of AI
- enforced self-regulation of and by organisations through regulated incentives
- a requirement that each organisation designate a senior officer who is responsible for ensuring safe and responsible deployment of AI
- coordination between regulators to ensure commonality of approach and avoid duplication and conflict in requirements
- support for complementary (but not conflicting or duplicative) initiatives to assure safe and responsible deployments of AI in particular contexts: for example, ASX guidance for listed corporations, APRA and ASIC standards and mandated requirements for regulated entities, international and Australian standards
- prohibitions (blacklists) in relation to use of AI in particular contexts
- prescriptions as to when and how a structured risk of harms assessment should be conducted by organisations where decisions by an organisation are to be assisted or influenced by uses of AI
- requirements to provide appropriate transparency and 'audit trail' for subsequent scrutiny by a regulator as to risk of harms assessments that have been conducted by an organisation.

In the middle:

- support for further development and modification of enterprise risk and operational risk frameworks and methodologies to ensure that AI-affected decisions are fully addressed by these frameworks and methodologies. This initiative is particularly important and time critical. There is currently a gap between enterprise risk and operational risk programs, and technology project management frameworks and methodologies, in relation to evaluation of AI-affected decisions, as distinct from deployment of AI as a technology project
- support for development of best practice in data sheets, system cards, and model cards for AI applications and services, particularly focussed upon non-enterprise AI. The need in non-enterprise AI sector is greater than for enterprise AI applications, as enterprise AI applications are likely to be evaluated by resources and skilled project managers and demand-side competitive pressure upon providers of enterprise AI is likely to lead to continuous improvement in transparency and disclosures
- support for development of international and Australian standards for AI impact assessment, including sector-specific, application-specific and task-specific standards that focus upon practical steps for non-specialist personnel in evaluation and mitigation of risk of harms from AI-affected decisions
- support for development of better understanding within organisations of the respective roles of C-suite executives, generalist managers, HR professionals, technology professionals, data science and AI/ML engineers, and lawyers,

privacy professionals and prudential and regulator specialists, in assuring safe and responsible use of AI by organisations.

Bottom up:

- an information campaign and publication of explanatory materials about safe and responsible use of AI, particularly targeting small to medium businesses that are unlikely to have internal capabilities or resources for AI impact assessment and that may be considering using self-service generative AI applications for business dealings and interactions
- educational resources and self-assessment leading programs for designated senior officers who are to be allocated as responsible for ensuring safe and responsible deployment of AI, particularly focussed upon organisations that do not have internal project management capabilities and resources, or developed prudential/regulatory teams (being most medium to small businesses and not-forprofits)
- support for complementary initiatives to assure safe and responsible deployments of AI in particular contexts. For example, for upskilling/cross skilling of (1) IT professionals to develop their capabilities to project manage AI impact assessment (Australian Computer Society and like professional associations could lead this project); (2) human resource professionals to develop capabilities to inform and manage uses by personnel within organisations of AI applications, particularly uses of generative AI for task assistance in circumstances where the use of particular AI for a particular task is not being project managed into an organisation following AI impact assessment (Australian HR Institute (AHRI) and like professional associations might lead); (3) marketing professionals to develop capabilities to inform and manage uses by personnel within organisations of AI for consumer marketing (Association for Data-driven Marketing and Advertising (ADMA) and like professional associations might lead).

In-depth response

Safe and Responsible AI in Australia

1 Broad approach to regulation

Many submissions to this current consultation will likely express views as to whether Australia should model regulation of AI on the more interventionist and prescriptive proposed EU AI Act[1], or the draft Canadian AI and Data Act[2], or the UK's proposed lighter touch, coordinated but decentralised approach.

We advocate an approach similar to the UK model, but on the basis that:

- regulation should focus upon ensuring that organisations design and implement policies and programs for responsible uses of AI that are appropriate to the organisation, and not the manner in which those policies and programs are implemented
- regulation should incorporate sanctions that create the incentive to ensure that the policies, programs and risk assessment methods are effective and reliably and verifiably implemented
- regulation of uses of AI should be closely targeted to some uses cases of AI, through sector-specific or use-case specific prescriptions or prohibitions that are tailored to address the particular context of use ('data context' or 'AI context')
- economy-wide prohibitions and prescriptions should build from existing statutes and other laws, notably by (1) building out from existing provisions of Australian Consumer Law and the Privacy Act 1988, in order to more clearly address transparency as to limitations of particular AI applications and services, and unlawful discrimination prohibitions under human rights statutes, and (2) further statutory prescriptions addressing AI-assisted targeting of misinformation and disinformation and excessive surveillance and profiling
- specific or use-case specific prescriptions or prohibitions should not be a response to pressure to address perceived existential threats 'from AI', but be targeted to circumstances where other incentives to improve AI-affected decisions made by organisations leave unacceptable risks of harms to humans or the environment
- a number of areas should be considered as priority areas for statutory reform, regardless of whether implementation of risk management assessment and

[1] References to the EU AI Act are to the draft compromise text of 16 May 2023, being the current draft as at 18 July 2023 and available at https://www.europarl.europa.eu/resources/library/media/20230516RES90302/20230516RES90302.pdf.

[2] The draft Artificial Intelligence and Data Act (AIDA), available at https://www.parl.ca/legisinfo/en/bill/44- 1/c-27.

management is mandated in relation to specified categories of AI uses (such as high-impact systems). Priority areas might be:
    - blacklisting (prohibition) of certain manifestly harmful and therefore societally unacceptable uses of AI
    - new documentation, transparency and disclosure requirements applying to offering of AI systems and commercial uses of AI systems and amendments to Australian Consumer Law to ensure coverage of provision of AI applications and services
    - statutory changes to address appropriate allocations of AI liability across supply and use chains[3] and addressing evidentiary burdens of providers and commercial users of AI systems as to safety of use of those systems.[4]
- regulation should be accompanied by programs to upskill and educate Australians and Australian organisations of all sizes on the safe and ethical application of AI.

1.1 Policymaking principles

We suggest that policymaking for regulation of AI follow similar principles to those proposed by the UK government, being:

- safety, security and robustness. The UK Competition and Markets Authority (CMA) in the CMA's response to the UK government's 2023 White Paper stated that organisations in properly functioning markets should "face the correct incentives to determine and implement the appropriate level of security and testing to ensure that their systems function robustly", and the CMA may need to intervene when this incentive is missing, i.e., when "AI use affects a consumer who may not be in a position to assess technical functioning or security of the product"
- appropriate transparency and 'explainability'
- fairness: the CMA noted that this principle should be applied to "the context surrounding the AI system", including data collection, testing and evaluation practices, and not just any underlying algorithms or AI functionality
- accountability and governance of organisations deploying AI

---

[3] The draft EU AI Liability Directive of 28 September 2022 proposes to make it easier for claims to be brought for redress of harm caused by AI systems and the use of AI. The proposal addresses the specific issues with causality and fault linked to AI systems. See https://commission.europa.eu/system/files/2022-09/1_1_197605_prop_dir_ai_en.pdf

[4] Adapting from recommendations of the Australian Human Rights Commission in its Human Rights and Technology Final Report, 2021. "Recommendation 11: The Australian Government should introduce legislation that provides a rebuttable presumption that, where a corporation or other legal person is responsible for making a decision, that legal person is legally liable for the decision regardless of how it is made, including where the decision is automated or is made using artificial intelligence." https://tech.humanrights.gov.au/artificial-intelligence/ai-informed-decision-making

- contestability and redress: the CMA noted that the "opacity of algorithmic systems and the lack of operational transparency" make it hard for customers to "discipline" firms, and stressed the importance for regulators to effectively monitor potential harms and to have the powers to act where necessary.

Perceptions by some organisations that harmful effects of their uses of AI are externalities suffered by others, rather than the organisation, may need to be changed through government action to create new, or changed, negative incentives that cause those organisations to internalise risks of AI harms and then be stimulated to mitigate these risks.

Most of the initiatives that we advocate require reimagining how, why, and by whom, risks of AI harms are evaluated by organisations. These initiatives require more than simple re-tooling or modifications of enterprise risk or project management frameworks and methodologies as today in common use by individuals within organisations that already have developed capabilities to use those methodologies.

1.2 Prescriptions

The Australian government's policy and regulatory responses over the last decade to data security threats illustrates what Australian governments should not do. Safe and responsible AI cannot be assured through a confusing proliferation of policy and legislative requirements, or through conflicting requirements imposed in parallel by multiple responsible regulatory agencies.

Safe and responsible AI does not require a 'super-regulator for AI', or 'super AI rules'. In this respect, assuring safe and responsible AI across Australian organisations is quite different from improving cyber-resilience of Australians. The Australian Government's assessment of incentives and prescriptions for safe and responsible AI should focus upon the contexts in which AI is used by people in processes of decision making by organisations operating in Australia.

Regardless of the operation of the incentives structure in relation to actions of most organisations, some organisations will be irresponsible, or knowingly or negligently commit or sanction illegal harms.

New prohibitions and prescriptions may be necessary to deter or punish irresponsible or bad actors. However, interventions need to be measured and adaptive, given rapid developments in functionality and reliability of AI and the myriad use cases now being trialled for use of generative AI applications.

Where AI regulation is justified, there needs to be consideration as to striking the right balance between:

- **prohibitions** – AI must not be used for a specified AI-affected activity
- **before the event (a priori) prescriptions** – AI may only be used for a specified AI-affected activity if a regulated entity first complies with specified preconditions (for example, conduct of an AI impact assessment)
- **before-the-event requirements** – for transparency, for organisations to develop and implement policies and programs to act responsibly and ensure

safety in organisational uses of AI, nomination of a responsible officer, and due consideration by or for the responsible senior officer of a regulated entity of relevant possible significant risk of harm factors even where the requirement for consideration does not extend to a formal structured process for risk of harms evaluation such as conduct of an AI impact assessment

- **after the event (ex post) legal exposures to damages and penalties**, and the appropriate level of transparency to enable detect-and-respond (remedy) incentives to operate, or to enable detect-and-prosecute.

Detect-and-respond (remediate), and detect-and-prosecute, are of course closely related. Both require (1) detection of an AI harm, (2) an evidence trail for root case analysis and for allocation of accountability, and (3) a relevant party/parties willing and resourced to respond, whether that party is the one causing the harm, a party suffering the harm but able to avoid further occurrence of the harm, and/or the regulator.

A sensible approach to AI regulation is to ask whether rules that restrict or prohibit particular uses of AI, or that mandate application of a particular risk assessment framework or methodology, are justified, or whether 'detect' and 'respond' incentives as adjusted for AI would then provide sufficient incentives to cause appropriate mitigation of AI risks by regulated entities. Many Risks of AI harms can be addressed by getting the detect and respond incentives right, and therefore an upfront restriction or prohibition is not required or justified.

2 Blacklists and assessment tools

While ACS agrees with the government's proposed position that a risk-based approach is a practical and workable solution to AI regulation, as noted through this response we will note below, we don't believe it is the complete solution. Instead it provides a framework for discussion and a model of comfort for AI implementors and the broader community.

2.1 Starting with a blacklist

Initial development of an 'AI blacklist' will address many of the most common concerns held by the public about AI. The blacklist should address:

1. Implementations of AI that create unacceptable levels of risk in critical systems.
2. Applications of AI that would be abhorrent to the broader community (examples might include public facial recognition databases or the use of deepfakes for spreading misinformation. Our response to Question 10 outlines some of those areas.)

The blacklist will create guardrails around AI that will address most of the community concerns and create space for more nuanced approaches in more 'grey' areas of AI implementation.

2.2 Provision of assessment tools and frameworks

Beyond a pure 'blacklist', the Australian Government should also lead in the development of tools and frameworks that will help organisations assess where their particular applications of AI sit within the risk framework.

Ideally, businesses would have access to self-assessment tools that:

1. are somewhat customed to different business sizes and types
2. provides a simple 'gating' assessment to determine whether the application of AI is on the blacklist (and therefore should not be used), is safe for use (that is, presents no notable risks), or is in the 'grey' area where some risks might be present
3. where the application does fall in the grey area, a deeper assessment is presented to determine where the application falls on the risk scale and any requirements or remediations that may be necessary for its use. Certain outcomes may point to the need for an escalation to a full AI impact assessment
4. are designed to enable organisations to continuously self-evaluate and re-evaluate their evolving uses of AI.

The tools need to be appropriate to the audience, being responsible and accountable officers within organisations who need to understand and address the diversity of risks of harms to humans and the environment that may arise from the organisation's adoption and use of AI.

The NSW AI Assurance Framework can serve as a starting point for this tool, having already been successfully implemented and mandated for both NSW departments and suppliers.

ACS would be happy to assist with the development of such a tool and the governance of an AI assessment framework and program. We are already working with ISO and the IEEE on the development of international frameworks.

2.3 Incentives to self assess

In addition to the framework, as noted above in section 1.2, the government will be required to provide incentives to organisations to apply the framework. These would likely be some combination of 'carrot' (tax and procurement incentives, for example) and 'stick' (mandates and regulatory penalties). Again, the UK can serve as a model, with its emphasis upon the creation and operation of a bundle of positive and negative incentives for organisations that are providing or using AI in their operations.

3 Enhancing organisational DNA

While regulatory incentives and access to assessment tools will be critical to the safe transition to AI, ACS believes a broader program of education is needed to ensure that the safe and ethical use of AI becomes part of organisational DNA. Ensuring that organisations – from large business to SMEs – have the required understanding to properly govern their application of AI will be imperative. Thanks to generative AI, we have already seen how quickly use of AI can become pervasive, and the implementation is not always "managed" – many individuals might start self-directed use of AI tools (as we have seen with the incredible spread of the use of large language models) without guidance, presenting real risks to their organisations and their customers.

Safe and responsible uses of AI for myriad tasks across diverse organisations cannot be reliably assured by modifying roles and responsibilities of currently designated privacy professionals, or by the frameworks and tools that they use to conduct privacy impact assessments.

Safe and responsible uses of AI will also not be reliably assured by organisations:

- beefing up current second or third line of defence functions or capabilities
- re-purposing privacy officers as AI officers
- outsourcing AI assessment to large professional services consultancies for episodic review of 'AI projects'.

Assurance of safe and responsible uses of AI needs to become part of the DNA of each organisation – public and private, business and not-for-profit, large and small – and consistently and reliably applied in the course of each organisation's business-as-usual processes.

We would suggest a good analogy is current workplace health and safety rules, which many organisations have made a fundamental part of their operations, including worker onboarding and dedicated organisational governance.

3.1 Requirements and incentives

Organisations should be required to develop and implement policies and programs to act responsibly and ensure safety in organisational uses of AI.

As a minimum, organisations should be required to prepare an annual plan setting out what they propose to do about ensuring safety in organisational uses of AI, including specification of reasonable precautions that the organisation is putting in place.

This mandate could be supported by mandated transparency requirements: for example, to publish risk of AI harms policies, and overviews of risk of AI harms programs.

There would need to be associated meaningful legal exposures for organisations, and their directors and their senior officers in the event that the organisation:

- did not develop and oversee reliable implementation and operation of policies and programs

- did not take reasonable precautions to mitigate reasonable foreseeable risks of AI harms
- did not comply with transparency requirements.

The legislative requirements would not be prescriptive as to the processes by which risk management is implemented and conducted within the organisation. The enforced self-regulation approach thereby enables flexibility in how organisations address particular categories of types of AI risks. For example, risk assessment and mitigations for availability of generative AI as a task assistant for myriad tasks performed by various staff members across an organisation could be quite different to risk management of single-task AI managed into an organisation following a structured project management process.

In addition to enforceable responsibilities of organisations and their directors and their senior officers (and meaningful legal exposures) under an enforced self-regulation model, we suggest a requirement that each organisation that is operationally using AI designate a senior responsible officer with responsibility to implement policies and programs to ensure safety in organisational uses of AI.

Ideally, that person will care about ensuring safe and responsible AI, and will have skills, authority, knowledge and access to practical tools to assure safe and responsible AI.

Of course, that person would not act alone: assurance of safe and responsible AI requires a multidisciplinary and cross disciplinary team approach that adapts existing ways of doing things and supplements existing risk frameworks, methodologies and tools. It's only through the combined effort across organisations – building it into the DNA – that safe and responsible AI implementation can be assured for Australians.

- Ends (Q&A to follow) -

Responses to questions posed in paper

Q1 **Do you agree with the definitions in this discussion paper? If not, what definitions do you prefer and why?**

The definitions on page 5 of the Discussion Paper are a useful guide to discussions as to when risks of harms arise from use of inference engines, including algorithmically-enabled automated decision making, use of foundational models and generative AI applications.

In considering possible AI harms and incentives to assess and mitigate relevant risks of harms, a broad definition should be used. As Robodebt illustrates, many significant harms to humans or the environment may flow from inadequate governance of use of hardcoded advanced data analytics (algorithmic) systems to inform or otherwise affect human decision-making or produce automated outcomes.

These uses therefore need to be brought within the ambit of new frameworks and methodologies for assessment of governance and associated controls for inference-assisted decisions, regardless of whether advanced data analytics (algorithmic) systems are considered as 'AI'.

It is generally not possible to determine in advance whether employing a technology, whether hardcoded advanced data analytics (algorithmic) systems, ML foundational models, or generative AI applications, moves a decision context across the gating threshold at which a desktop review is inadequate and at which a more comprehensive, data context-specific, risk assessment should be conducted.

Certain use cases may be considered sufficiently high risk of harms that they should be prohibited. In some of those uses cases, the risk factors that cause the use case to be so high risk of harms that cannot reasonably be expected to be mitigated relates to the lack of 'inside the box' explainability that is a characteristic of some, but not all, ML, or the error limitations of current generation generative AI.

For these use cases it may be useful to distinguish ML foundational models, and generative AI applications, from deployment and use of hardcoded advanced data analytics (algorithmic) systems.

In this regard, we commend the UK approach. Instead of attempting to define "AI", and then imposing generic requirements to AI as defined, we suggest that the Australian government should focusing on setting economy-wide expectations for the development and use of AI and empower existing regulators to issue guidance and regulate the use of AI within their remit.

The UK Government's 2023 White Paper[5] scopes proposed activities by reference to two "characteristics of AI", being adaptivity and autonomy. The 'adaptivity' of AI can make it difficult to explain the intent [object] or logic of the system's outcomes: AI

[5] https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper

systems are 'trained' – once or continually – and operate by inferring patterns and connections in data which are often not easily discernible to humans. Through such training, AI systems often develop the ability to perform new forms of inference not directly envisioned by their human programmers.

The 'autonomy' of AI can make it difficult to assign responsibility for outcomes. Some AI systems can make decisions without the express intent or ongoing control of a human. This broad characterisation is intended to future-proof proposed regulatory frameworks for AI against new technologies and to confer sufficient discretions upon individual regulators, who might then issue guidance to regulated organisations setting out their expectations about the use of AI within the regulator's remit.

This approach allows flexibility, but also creates a risk of lack of comprehensiveness of scope and coverage and inconsistency between regulators. AI-affected activities within many Commonwealth, State and Territory government agencies might fall outside the scope of existing regulators' remits.

Q2 **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?**

Disclosure and transparency

Many AI applications and services commercially offered in Australia originate from outside Australia. Australian regulation can reasonably address transparency and disclosures offered by offshore providers to organisations using these applications and services in Australia.

Data sheets, system cards, and model cards made available by some providers of commercially offered AI applications and services are today of different levels of transparency and frankness in disclosure of known limitations, comprehensiveness and quality.

The problem reflects still nascent standardisation as to supplier disclosures for AI systems and AI services. This is partly a transitional issue that is already being addressed in the competitive AI market. However, regulated requirements for transparency and disclosure may assist.

Transparency and disclosure requirements might be applied economy-wide to commercial offering of an AI product or service that is intended for use by a customer, whether a business customer or other organisation, or a consumer or other end user.

These requirements could take the form of new provisions in Australian Consumer Law and analogous requirements in sector specific laws, such as the Corporations Act, that ensure economy wide coverage of such provisions.

Disclosure might be required of suspected errors, bias or other limitations that should reasonably be anticipated to materially affect reliability of outputs of the AI product or service when used by or for a customer for such tasks, or other uses as ought reasonably be anticipated by the commercial provider of an AI product or service.

Disclosures should be made promptly and with sufficient prominence, and disclosures should be capable of being understood by a non-expert reader of these disclosures.

They might include warnings or cautions in data sheets, system cards, model cards or other sufficient prominent instructional or explanatory material or as to tasks of other uses for which the AI product or service is not intended by the commercial provider to be fit for purpose.

Disclosures might include checklists, tools or other assurance aids and recommendations as to evaluation and assessment by the customer. However, warnings and cautions, and recommendations as to evaluation and assessment by the customer, must not be unfair to the customer, having regard to the nature of the AI product or service and the relative knowledge, skills and capabilities of the provider and the customer to anticipate, assess and mitigate likely risks of harms to humans or the environment caused by uses of AI affected decisions.

Gaps in consumer law

There are gaps in coverage of Australian Consumer Law, and in particular operation of the definitions of "consumer", "goods" and "services" as used in the ACL, and the consumer guarantees under the ACL. Many of these issues of gaps in coverage, and similar issues as to appropriate allocation as between providers and customers of responsibility and accountability to anticipate, assess and mitigate risks of harms, also arise in relation to deployment of internet of things consumer ("smart') devices and IoT device enabled internet connected ('smart') services.

The draft EU AI Act proposes further notice requirements that would apply both to:

- providers placing AI systems on the EU market or putting AI systems into service in the EU, and
- users of AI systems.

Providers and users of AI systems would have new transparency obligations vis-à-vis affected individuals, subject to limited exceptions.

Relevant provisions of the draft EU AI Act remain under active negotiation, but their current form would require providers to ensure individuals are informed that they are interacting with an AI system. If an AI system generated 'deep fakes', the user of the AI system would be required to disclose this. Users of an emotion recognition system or a biometric categorisation system would be required to inform affected individuals.

Providers and users of generative AI would be subject to additional transparency requirements, including disclosing that the content was generated by AI and preventing the generative AI from generating illegal content. Providers of generative AI would be required to publish a description of copyright material used for training the foundational model.

We recommend consideration of further notice requirements similar to these EU proposals, while noting that:

- the categories of AI systems and relevant uses that should be subject to such requirements should be the subject of further consideration,
- notice fatigue of users is already a well-recognised problem of data privacy regulation.

Users should not be unreasonably burdened with further notices with any expectation that users should read and engage with such notices in order to protect themselves from harms that were reasonably foreseeable to the drafter of the notice. Notice to

users should not be allowed to become a substitute to AI system providers and users exercising reasonable diligence to protect affected individuals from AI harms, by those providers and users taking risk mitigation measures that are reasonably within their capabilities.

Biometric data gathering

We also note heightened risk of harms through many uncontrolled and opaque uses of AI enabled facial recognition to enable identification of individuals. We commend the proposal for an AI facial recognition model statute as made in the UTS Human Technology Institute's report of September 2022.[6]

Q3 **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.**

The broader response paper attached to this submission outlines a number of non-regulatory responses, including:

- higher level guidance and guidelines as to safe and responsible deployments of AI
- coordination between regulators to ensure commonality of approach
- support for further development and modification of enterprise risk and operational risk frameworks
- support for development of best practice in data sheets, system cards, and model cards for AI applications and services
- support for development of international and Australian standards for AI impact assessment, including sector-specific, application-specific and task-specific standards
- support for development of better understanding within organisations of the respective roles of executives and professionals, in assuring safe and responsible use of AI by organisations.
- an information campaign and publication of explanatory materials about safe and responsible use of AI, particularly targeting small to medium businesses that are unlikely to have internal capabilities or resources for AI impact assessment
- educational resources and self-assessment leading programs for designated senior officers
- support for complementary initiatives to assure safe and responsible deployments of AI in particular contexts.

[6] UTS Human Technology Institute, *Facial recognition technology: Towards a model law*, September 2022, https://www.uts.edu.au/human-technology-institute/projects/facial-recognition-technology-towards-model-law

Q4 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.

We commend the UK's lighter touch, coordinated but decentralised approach, as described in the UK Government's Policy paper *A pro-innovation approach to AI regulation* of 29 March 2023[7] and *AI Regulation Policy Paper* of 18 July 2022[8], with the modifications that we describe in section 1 of our attached response.

The UK Government proposes creation of central functions to support the multi-regulator, decentralised frameworks, including by:

- developing a central monitoring, evaluation and risk assessment framework
- creating a central guidance to businesses looking to navigate the AI regulatory landscape in the United Kingdom
- offering a multi-regulator AI sandbox, and
- supporting cross-border coordination with other countries.

While no announcement has been made, the UK Government's Office for Artificial Intelligence, a unit within the UK Department for Science, Innovation and Technology, may take on some of these central functions. The UK Government currently addresses regulatory coordination through activities of the Digital Regulation Cooperation Forum (DRCF) and the Centre for Data Ethics and Innovation (CDEI).

Q5 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?

While Australia should take its own approach, as noted to our response to Question 3, the UK's approach aligns with what we believe should be the Australian model.

Q6 Should different approaches apply to public and private sector use of AI technologies? If so, how should the approaches differ?

No. We are not aware of any reason why other assessment and management of Risks of AI harms by public sector agencies should be less than assessment and management by other Australian organisations.

Q7 How can the Australian Government further support responsible AI practices in its own agencies?

[7] https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper

[8] https://www.gov.uk/government/publications/establishing-a-pro-innovation-approach-to-regulating-ai/establishing-a-pro-innovation-approach-to-regulating-ai-policy-statement

The development of AI governance skills will be critical for government agencies – as much or more so than private organisations. Right now, there is a severe AI skills shortage within both the public and private workforces. Encouragement of agencies to engage in upskilling for AI (possibly through the APSC's Digital and Data Professional development streams) will be a good first step.
Such education should not just focus on technical applications of AI, but in the proper governance and risk assessment for AI system. The Robodebt Royal Commission report highlights the needs for these skills and accountabilities within the public service.

Q8 **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.**

This is covered in our response to Question 2 above and in the attached paper.

Q9 **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.

This is covered in our response to Question 2 above and in the attached paper.

Q10 **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?

There are certain technologies or applications of technology that should present uncontrollable risks or outcomes that warrant bans. We imagine that most Australian citizens, for example, would find the work of an organisation like Clearview AI (which created and sold access to an enormous global facial recognition database) abhorrent.

However, discussions as to blacklists can readily become politically contentious as to edge cases (why was X included, but not Y?), and the objectively assessed mitigation effects of risk measures and assurance controls may not be broadly understood. In our view, blacklisting of particular categories of AI applications should only be considered in relation to uses of AI that are of such extreme risks of harms to humans or the environment as to be unacceptable to Australian society regardless of whether safeguards and assurance controls are reliably and verifiably applied.

Inclusions in the proposed blacklist for the draft EU AI Act remain controversial. The list was significantly extended by Members of the European Parliament in May 2023, and it is unclear whether the extended list will be the final list. The list as then adapted includes:

• systems that deploy subliminal or purposefully manipulative techniques or exploit people's vulnerabilities

• systems used for social scoring (classifying people based on their social behaviour, socio-economic status, personal characteristics)

• use of real-time remote biometric identification systems in publicly accessible spaces

• post-time remote biometric identification systems, with the only exception of law enforcement for the prosecution of serious crimes and then only after judicial authorisation

• biometric categorisation systems using sensitive characteristics (eg., gender, race, ethnicity, citizenship status, religion, political orientation)

• predictive policing systems based on profiling, location or past criminal behaviour

• emotion recognition systems in law enforcement, border management, workplaces or educational institutions

• indiscriminate scraping of biometric data from social media or CCTV footage to create facial recognition databases and by so doing violating human rights and right to privacy.

We should note that there is likely to be a strong correlation with comments made in respect to the current review of the Privacy Act. Many of the immediate concerns around AI are related to the gathering, retention and use of personal information. It will be tempting for businesses to keep personal data on hand to 'feed the AI machine', and a set of strong and workable laws around privacy will mitigate many of the issues people might have with AI.

Q11 What initiatives or government action can increase public trust in AI deployment to encourage more people to use AI?

Education programs, particularly targeted at small businesses, can alleviate many of the issues. An information program modelled on the ACCC's Scamwatch, for example, can be used to educate people on the safe application of AI.

Please see our response to question 3 above for specific recommendations.

Q12 How would banning high-risk activities (like social scoring or facial recognition technology in certain circumstances) impact Australia's tech sector and our trade and exports with other countries?

If an activity is worthy enough to warrant a ban, then we would argue that its impact on trade should be irrelevant. Australian companies should not be encouraged or

enabled to engage in activities that are immoral or damaging so long as it is done "over there but not here."

Q13 What changes (if any) to Australian conformity infrastructure might be required to support assurance processes to mitigate against potential AI risks?

No response

Q14 Do you support a risk-based approach for addressing potential AI risks? If not, is there a better approach?

Yes. Risk assessment should be part of the 'mix'. It's not a case of either/or, but a matrix of responses to AI. A risk-based approach, in the style of the EU legislation, provides a useful tool for evaluating AI applications and whether they should be restricted or blacklisted.

However, principles-based governance, training and education also has a role to play.

Each AI-enabled, or AI-assisted, decision requires consideration of decision provenance and outcomes: the interaction of people, processes and technologies that effect, or affect, that decision. Most organisations operating in Australia that are implementing AI within the organisation will not be developers and suppliers of AI solutions. Typically, an organisation operating in Australia will be tailoring a third-party AI application or service and using it:

- more commonly, to inform or otherwise aid people within an organisation to perform a decision-making task, or
- much less commonly, to enable a fully automated (self-actuating) outcome.

Senior executives and managers within organisations may not see a business process (decision chain) within the organisation as an 'application of AI', or even as significantly affected by AI.

Most organisations, and particularly those that are not large businesses, do not have internal competencies to reliably translate higher level 'ethical AI' principles into practical business decisions. Most organisations are not large businesses that have experience, settled procedures, internal capabilities and resources to reliably evaluate a third-party AI application or service for fitness for purpose for reliance in a particular business context, and assess the quality of data inputs used by the AI provider to train that AI.

Frameworks, methodologies and tools for assessment of AI are the essential translational layer between:

- higher level 'ethical AI' principles, and
- practical business decisions that effect safe and responsible uses of AI.

This translational layer need not be complex or highly structured, depending upon capabilities and willingness of organisations:

- to recognise that there are relevant risks,

- to understand the nature of risk of harms, and
- to allocate responsibility to address and mitigate those risks, and thereby demonstrate accountability and trustworthiness.

One danger for organisations to navigate when implementing AI is that because AI is a novel technology, the temptation is to focus too much on the technology, and not the relevant humans using that technology in a particular context to make a particular decision.

If the context can be appropriately circumscribed and evaluated for that context, and the skills of relevant humans reliably pre-assessed, risks of uses of AI can be substantially mitigated.

For example, each neighbourhood florist must water cut flowers and plants and each day changes where watered pots sit within a store. Each florist store manager manages a known slip and fall hazard zone. Their management of that zone is a combination of inherent human competency, procedures (eg. promptly mop up split water), and simple checklists.

Management of that zone is usually effective to mitigate those risks, because the risks are familiar and well understood by non-specialist risk managers. Each airline pilot manages a more complex work environment, aided by rigorous context (cockpit specific) training, structured checklists, and knowledge as to critical dependencies upon other people conducting other checks and exercising safety-related responsibilities.

The airline pilot requires checklists to aid learned knowledge and skills. The airline pilot does not require a complex structured risk management framework or methodology.

Building capability

Design and deployment of:

- advanced data analytics (data and algorithms) to inform or assisted decision-making by humans,
- task-specific AI/ML applications, and
- multi-purpose generative AI

should cause organisations to consider change control, and address management of deployment and use of a new technology.

Some of these organisations will have capabilities and experience to apply enterprise or technology risk frameworks, methodologies and tools. This experience may have been applied by the organisation to deployment of earlier technologies, such as cloud platform enabled integration of diverse data sets, use of social media, COVID accelerated take-up of video-conferencing, and COVID and post-COVID remote access by personnel to an organisation's information systems and trade secret and sensitive data.

Typically, an organisation addressing a new technology should promptly consider:

- whether new gating criteria are required – whether the new technology will be taken up by the organisation, and if so, who should be permitted to do what, using the new technology

- whether prohibitions (no-go zones) need to be created, either to ensure that the technology is only used by the organisation's personnel for lawful purposes, or to ensure that there are not significant harms to the organisation or to others (customers, suppliers, citizens and stakeholders)
- design and implementation of appropriate guardrails that ensure that personnel that are authorised to use the new technology to perform particular tasks operate only within those guardrails.

In other words, an organisation's governance framework needs to be fit for purpose to enable an organisation to promptly determine, in response to the new technology, how to adjust each of the three interrelated elements of people, processes and technology, in order to ensure that having regard to the interaction of the three elements, the organisation's use of the new technology is lawful and does not cause significant harms to the organisation or to others (customers, suppliers, citizens and stakeholders).

Within an organisation, introduction of a new technology will often require significant change management led by the human resources team, because the human element of the socio-technical decision chain within the organisation is so critical in ensuring safe and responsible use.

Assuring implementation of safe and responsible AI for Australian citizens requires organisations:

- to understand what is 'safe and responsible AI',
- to internalise and address risks of harms to others, to the extent that those harms are reasonably attributable to the organisation's provision, deployment or use of AI.

Organisations are as diverse as the tasks, processes and decisions for which they will be using AI. Within these diverse organisations, significant change management will be required to implement safe and responsible AI.

In larger organisations with more mature risk management frameworks, new technologies are typically passed through a project initiation, evaluation and project management process. Often these processes are managed by technology professionals that are skilled in applying either an enterprise-wide risk framework, or a technology focussed framework, and associated methodologies and tools.

Frameworks, methodologies and tools typically have been developed for use in organisations where structured assessment and management of enterprise, operational and technology risk. This is not the case for the large majority of Australian businesses and social enterprises that are now implementing AI applications and services.

This fact creates a challenge for Australian policymakers. A majority (by number) of Australian organisations are unlikely to develop capabilities to implement complex structured frameworks, methodologies and tools for management of AI risks within the next three to five years. Many of those organisations will implement and use AI applications and services policy within that period. Further, for most organisations addressing risks of AI harms will be quite different from managing technology disruptions in the past.

Assurance by organisations that their uses of AI are safe and responsible requires a team approach bringing together different disciplines. Many possible AI harms arise because outputs from AI or hardcoded algorithmic data analysis are statistically based, so outputs are not universally reliable for use in the broad range of contexts in which decisions based upon those outputs may be made.

Decision context should affect assessment of whether earlier links in the decision provenance chain are sufficiently robust to be reliable inputs for a particular decision.

An informed understanding of the quality and reliability of each link in the chain of data inputs, people, processes, and technologies used to create AI/algorithmically enabled output, and then apply that output in a way that affects, or makes, a particular decision (an outcome), is crucial to ensuring that an AI/algorithmically assisted decision is appropriately reliable for the reliance that is placed upon it.

Evaluation of the AI/algorithmic links within this chain of decision provenance is important, but only part of an evaluation of the quality and reliability of decision provenance that needs to be made by an organisation responsible for an AI/algorithmically assisted decision.

'Statistical errors' may be addressed by a 'human in the loop': organisational reliance upon appropriately skilled humans to review the outputs and detect and override insufficiently robust results.

In addition, information technology risk frameworks and methodologies have been developed over three decades. AI project frameworks and methodologies are nascent, less developed and standardised, and therefore less understood than standardised information technology project frameworks.

In the last two years we have also seen emergent, albeit still work-in-progress, best practice exemplars for AI project assurance frameworks, such as the NSW AI Assurance Framework[9], national and international AI assurance standards and work by entities such as the Turing Institute, Ada Lovelace Foundation, World Economic Forum, Gradient Institute and CSIRO.

These exemplar AI/algorithmic assessment frameworks and methodologies are generally designed for deployment within a system of a project initiation and approval, where appropriately skilled and experienced individuals evaluate the suitability, safety and legality of a proposed implementation of AI.

This 'gating and penning' process reduces risk of AI being inappropriately deployed. Effective 'gating and penning' requires an organisation to ensure that:

- there is a gate,
- the gate is manned by humans with appropriate skills
- candidates for assessment are identified early and required to pass through the gate
- a suitable assessment framework reliably and rigorously applied for each proposed use of AI that is 'within the pen'

[9] https://www.digital.nsw.gov.au/policy/artificial-intelligence/nsw-artificial-intelligence-assurance-framework

- evaluation within the pen ensures that AI outputs are fit for purpose, having regard to the likely reliance that will be placed upon them, such that decisions enabled or affected by AI outputs (i.e., outcomes) reasonably reflect the quality and other provenance of the AI outputs
- 'out of the pen', real world outcomes from uses of AI outputs are assessed for fit to expectations,
- proper change evaluation controls are applied before any subsequent changes in data inputs, data processes, AI/ML/algorithmic functionality or uses of outputs are made,
- adverse consequences suffered by others from uses of the AI by an organisation are not treated as externalities and ignored by the organisation, or left for some party to address.

However, emergent AI project assurance frameworks have a number of limitations:

- They are new and therefore unfamiliar.
- They are quite complex, and typically require multi-disciplinary input in order to be done well. Typically, an experienced project manager is required to manage the process. The project manager will require the skills, experience and conferred authority to obtain, manage and evaluate input from a diverse range of stakeholders, typically including data scientists, algorithmic/AI engineers, operational process specialists, human resource personnel, prudential and regulatory risk advisors, privacy professionals and legal counsel.
- Oversight governance personnel need to be also appropriately familiar with AI risks and harms assessment. In understanding how use of AI affects decisions made by or on behalf of an organisation, each decision context affected by the use of AI needs to be considered, having regard to the decision chain: the links of people, processes (including rules and policies) and technologies that make up a chain that leads to an AI affected decision.
- When AI affected decisions are irresponsible or unsafe, harms can be caused at scale and velocity. Because AI tools are now available as self-serve applications available to all organisations, even smaller organisations can quickly cause harms at scale.

As noted in the attached response, ACS believes that the Australian Government can lead in the development and diffusion of these frameworks across the nation, including to smaller businesses that might not have the skills to execute a high-level impact assessment. ACS would be happy to help in the development and governance of a common assessment framework and toolset.

Q15 What do you see as the main benefits or limitations of a risk-based approach? How can any limitations be overcome?

This is covered in our response to Question 14 above and in the attached paper.

Q16 Is a risk-based approach better suited to some sectors, AI applications or organisations than others based on organisation size, AI maturity and resources?

Yes. This is covered in our response to Question 2 above and in the attached paper.

Q17 What elements should be in a risk-based approach for addressing potential AI risks? Do you support the elements presented in Attachment C?

The elements summarised in Attachment C are relevant but incomplete, partly because the focus in Attachment C is upon assessment of the AI system itself, rather than the decision context that is being affected by the AI.

We discuss the appropriate scope for transparency and disclosures, and incentives for transparency and disclosures, in section the attached paper and in our response to question 2.

Q18 How can an AI risk-based approach be incorporated into existing assessment frameworks (like privacy) or risk management processes to streamline and reduce potential duplication?

This is covered in our response to Question 14 above and in the attached paper.

Q19 How might a risk-based approach apply to general purpose AI systems, such as large language models (LLMs) or multimodal foundation models (MFMs)?

There are at least three relevant levels at which risks from general purpose AI systems need to be assessed:

1. The underlying foundation models and the data inputs used to fuel those models. The developers of these foundational models need to have appropriate incentives to make fair disclosures as to limitations of those models, so providers of generative AI applications built upon those models may consider the reliability of the foundational model, assess the reliability and safety of the application, and in turn make appropriate disclosures and ensure that their generative AI application offering is safe and complies with law.
2. Provision of the generative AI application built upon a foundational model.
3. Use of a generative AI application for a particular user-determined task.

We have already made specific suggestions as to transparency and disclosure requirements to enable risk assessment at levels 1. and 2.: see in particular our response to question 2 above.

As to 3., we note that there are a number of reasons why generative AI applications are 'fast fashion' for many prospective users:

- generative AI applications are readily available at low cost
- they can be accessed by anyone within many organisations without first needing the organisation to buy, the IT department to deploy, or the boss or the HR team to approve. This 'ungated' use of AI within organisations is sometimes referred to as use of 'shadow AI' or 'stealth AI'
- they are easy to play with
- their outputs are compellingly useable, even when unreliable
- they readily demonstrates their usefulness as an aid to performance of many tasks, even when unreliable as used as an aid for those tasks.

Many would-be users will experiment, including in their own time and without use of an organisation's IT resources. Warnings and cautions issued by organisations should be expected to be ignored by some would-be users within those organisations. Bans or controls should be expected to be circumvented by many users.

Provision and use of generative AI applications therefore amplify some of the categories of AI harms as discussed elsewhere in this paper. Take-up of generative AI applications will likely expand the range of risks of harms, notwithstanding providers of generative AI applications also improving reliability of these services and disclosures as to limitations in reliability of these services.

Take-up will also be fuelled by developing generations of general purpose AI that enable 'air-guarding' of prompting data from the data corpus of a provider of the AI model or functionality, thereby enabling deployments in many data contexts where under current laws regulated data sets could not be used either to train the underlying foundational model, or to prompt the generative AI.

Many Australian organisations that do not have capabilities to implement complex structured risk management frameworks or methodologies will be implementing generative AI to assist non-technical humans to perform myriad tasks.

Larger organisations that have structured functional teams will usually only introduce a new technology after risk assessment and with an associated change management program led by the human resources team.

That program is likely to include changes to policies and process documentation; changes in oversight and internal review processes; re-designation of roles and responsibilities of staff members; new training; new instructional materials; new warnings and 'no-go zones', and so on.

Because addressing the human element in socio-technical decision chains within the organisation is critical in ensuring safe and responsible use of AI, the risks of unintended and unanticipated AI harms are much greater for the majority of organisations that do not have a project evaluation program, a change management program, or a human resources team.

This creates an important policy challenge in addressing widespread use of generative AI, given material prevalence of unanticipated errors within AI outputs that are compellingly presented as ready for use (what has been popularly described as "confident bullshit").

Generative AI applications therefore create novel governance risks of unknowing amplification of disinformation and misinformation, as well as opportunities for deliberate use for disinformation and misinformation.

Consider one well-publicised recent example. Steven Schwartz, a New York attorney with over 30 years of post-admission experience, represented Roberto Mata in an action against Avianca Airlines for injuries sustained from a serving cart while on the airline in 2019. At least six of the submitted cases by Schwartz as in a brief to the court of the Southern District of New York court "appear to be bogus judicial decisions with bogus quotes and bogus internal citations," said Judge Kevin Castel in a May 2023 order.[10]

Judge Castel ordered Schwartz and his law firm to pay $5,000 for submitting without checks a brief with fake cases and then standing by the research. As one response to this case, a federal judge in Texas is now requiring lawyers in cases before him to certify that they did not use artificial intelligence to draft their filings without a human checking their accuracy.

One view might be that this example illustrates a transitional problem, and not a lacuna in regulation. Regardless of any view as to the small penalty imposed, no sensible lawyer would wish to suffer the reputational damage flowing from global reports as to the lawyer's failure to understand and mitigate the misinformation risks of reliance upon an LLM.

However, the difficulty is that analogous inappropriate reliance upon erroneous outputs from generative AI may arise in many of the myriad tasks for which generative AI is now being used by individuals without those individuals knowing to exercise appropriate caution as to the possibility of such errors and responsibility to take appropriate steps to mitigate such risks.

A risk management approach, coupled with an enforced self-regulation model, should be applied to general purpose AI services, designed to address the challenges associated with:

- lack of organisational control over how it is likely to be introduced into and used in many organisations for a myriad of tasks
- the role of individuals within those organisations in determining when and how general purpose AI services are used as a task assistant, and the best ways to ensure those individuals exercise appropriate restraint and care

[10] The cases, generated by ChatGPT, included Varghese v. China South Airlines, Martinez v. Delta Airlines, Shaboon v. EgyptAir, Petersen v. Iran Air, Miller v. United Airlines, and Estate of Durden v. KLM Royal Dutch Airlines. Neither the judge or nor the defence lawyers could find reports of these judgements: they did not exist, although generated by ChatGPT.

- the key role that transparency can play in building awareness of risks and capability to mitigate risks.

Q20

Should a risk-based approach for responsible AI be a voluntary or self-regulation tool or be mandated through regulation? And should it apply to:

a. public or private organisations or both?

b. developers or deployers or both?

See our discussion paper on the enforced self-regulation model as a way to change organisational DNA and enable flexibility as to the processes by which organisations adopt a risk-based approach for responsible AI.

A risk-based approach for responsible AI should be applied by both developers and deployers of both foundational models and algorithmic decision-making systems and generative AI applications built upon those foundational models or algorithms; as well as organisations, both public or private, that are users of third party supplied foundational models and algorithmic systems.