Supporting responsible AI: Discussion Paper

As a trusted government partner and corporate citizen, Transurban Limited, on behalf of itself and the Transurban Group (**Transurban**), welcomes the opportunity to provide feedback on the Supporting Responsible AI Discussion Paper (**Discussion Paper**).

Transurban uses technology like automatic incident detection systems and smart sensor monitors for things like debris and stopped vehicles, and to alert traffic control centres to potential issues. Transurban takes this technology to the next level by feeding the data into a purpose-built AI platform that learns from previous incidents and recommends how our crews should respond. The platform prioritises the safety of drivers and incident response teams, often recommending measures like speed reductions or lane closures to keep everyone safe while the incident is managed.

Transurban also supports the Responsible AI Index, which measures and tracks how well Australian organisations are designing and implementing AI systems, with a view to fairness, accountability, transparency, and impacts on people and society.

Our feedback on specific questions is set out below.

Definitions

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

The key definitions used in this discussion paper are broad and seemingly vague, for example, the definitions of "Machine learning" and "Generative AI". The definitions as they are currently drafted are not entirely reflective of the complexities of the subject matter. Transurban suggests the following definitions which provide further substance:

"**Machine learning** is a subfield of artificial intelligence (AI) that focuses on the development of algorithms and statistical models that enable computers to learn from data and improve their performance on a specific task without being explicitly programmed. In traditional programming, developers write explicit instructions to perform a task, but in machine learning, the computer learns patterns and relationships from data and uses that knowledge to make predictions or decisions"; and

"**Generative AI** is a subset of artificial intelligence that focuses on the creation of new content, such as images, videos, music, or text, by training models to generate realistic and original outputs. Unlike traditional AI, which primarily focuses on problem-solving or pattern recognition, generative AI aims to generate new and creative content that resembles what a human might produce".

<u>Potential gaps in approaches</u>

2. What potential risks from AI are not covered by Australia's existing regulatory approaches? Do you have suggestions for possible regulatory action to mitigate these risks?

Australia's regulatory approach needs to apply to both the organisations creating and training the AI models, as well as the organisations applying those models. At the training stage, regulation needs to ensure training of the models minimises bias, entrenches ethics and is, to a reasonable degree, explainable. It also needs to ensure the models are trained lawfully, which requires consideration of existing regulations including those relating to privacy and intellectual property. At the same time, it's important that any regulatory burden at the training stage does not prohibit small to medium sized organisations being able to compete in the creation of AI models.

At the application stage, Australia's regulatory approach needs to recognise that some organisations applying AI technology will be the AI creators themselves, but a large proportion will be organisations who have procured the AI technology from third parties and are then applying and/or customising it to their own business and customers. The regulatory mechanisms should account for this distinction and ensure that regulations apply to both creators and consumers of AI technology.

In addition to AI-specific regulation, consideration needs to be given to existing laws and regulations and whether they need to be updated to ensure they apply to scenarios where an outcome is driven by an AI model, rather than a human decision. For example, could two AI models collude to fix pricing in contravention of the Competition and Consumer Act?

Any regulation of AI needs to consider the underlying incentives of the technology and the organisations building and applying it, as well as the principles that the regulations are trying to uphold, so as to ensure that AI is used responsibly and ultimately for the benefit of individuals, society and the environment. On the other hand, any such regulation needs to ensure Australia remains competitive with the rest of the world when it comes to AI, and advanced computing, which will most likely require alignment with international approaches.

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

Part of supporting responsible AI practices is to ensure that consumers, users and customers of AI products have a right to request details about, or where appropriate, challenge how decisions were made by AI products using data provided by them. This allows for an opportunity for such individuals to understand how decisions were made, including how their data was used during the decision-making process. Such right of review may require escalation processes if not sufficiently dealt with by the organisation making the decision.

Transurban acknowledges that the complexities and technical intricacies of AI technology may make explainability difficult. In relation to a consumer's right to challenge the basis of a decision, often the issue with a decision will be the quality and accuracy of the data points used to reach that decision, rather than the mathematical mechanics underlying the AI model. It is reasonable to expect that AI creators and organisations who apply the AI models would provide an individual

information about the data they hold and how it was used in the AI model to arrive at a certain outcome.

Australian governance measures and initiatives should take into account the fact that the creation of an AI system is based on data points and algorithms. There will be scenarios where an individual may not have the applicable data points to fairly receive an AI-generated decision. This is demonstrated by an example of using an AI product designed in Australia to assist with the rental crisis and filtering applicants for tenancies. If a recent immigrant or refugee were to arrive in Australia and has the means to pay for a rental property but no applicable rental history or credit history, they may not be considered fairly by the AI product. Importantly, AI governance measures:

- should ensure that individuals subject to an AI model are not unfairly treated due to a lack of available data; and
- must be adaptive to change and contain improvement processes to ensure that biases and discrimination are eliminated from AI products as much as possible.

This also leads into an important conversation regarding data sharing and data sovereignty in an international landscape and how that interacts with the use of AI globally.

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

AI governance is overdue in Australia and AI must be regulated. There should be cohesiveness and coordination across Australia, and any AI specific regulation should apply at the federal level. Coordination mechanisms could include legislative standards, ethical guidelines, codes of conduct and practical procedures, methods and tools.

Using the example of generating "deepfakes" to provide financial services or legal advice, which may cause deceit by encouraging people to rely on such services/advice, legislative changes may be required to contemplate such scenarios. For example, expanding identity fraud legislation and subsequent penalties to apply if there is a risk of an individual using someone else's personal information via the generation of "deepfakes" to obtain a benefit, without the subject individual's consent.

However, as noted above, any regulation of AI in Australia must not become a barrier to Australia's competitiveness and must ensure Australia remains able to take advantage of emerging technologies without undue delay. Coordination between government agencies using AI in their own activities, regulators and policy makers could also help ensure Australia's regulatory response matches pace with the technology.

Responses suitable for Australia

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

No response.

<u>Target areas</u>

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

Implementing a risk-based approach to AI governance should account for the different inherent risks that public and private sector use of AI technologies pose. Please see further comments below regarding the risk-based approach.

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

No response.

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

No response.

9. Given the importance of transparency across the AI lifecycle, please share your thoughts on:

a) where and when transparency will be most critical and valuable to mitigate potential AI risks and to improve public trust and confidence in AI?

As outlined above, there appear to be two regulatory touchpoints applicable:

- at the AI training stage; and
- at the application point.

First, transparency is critical to mitigate any innate bias that exists during and at the point the AI technology is being trained. There must be an ability for creators to explain the process in which outcomes are derived.

Secondly, transparency is required to enable consumers the ability to test and challenge any automated decision making, ensuring that such decision-making and actions based on AI models are appropriately reached and not via prohibited means.

Transparency will be required from AI-model creators and the organisations that market and sell the AI products to increase public trust and confidence in using AI products.

b) mandating transparency requirements across the private and public sectors, including how these requirements could be implemented.

No response.

10. Do you have suggestions for:

a) whether any high-risk AI applications or technologies should be banned completely?

AI applications that generate content in the form of video, images, text, dialogue etc. are highly risky (for example, "deepfakes"). Scammers can target vulnerable people and use these products to financially deceive in a broad range of industries, such as financial services or law. This has serious consequences for an individual, financially but also emotionally. "Deepfakes" can also be used to personally attack individuals, and to sow political discord. These high-risk AI applications and technologies require stringent mechanisms that clearly specify what content is AI-generated or human generated.

Transurban supports consideration being given to laws that require some mechanism that clearly states what is AI-generated content. This could take the form of a disclaimer or a nationally recognised watermark, that must be provided anytime AI-generated content (such as video and/or text) is produced, so as not to mislead or deceive a consumer. Consideration could also be given to the imposition of offences and penalties for the removal of any disclaimer or nationally recognised watermark.

Other laws and regulations may also need to be updated to apply to the "deep fake" application of AI technology. For example, where "deepfakes" are used to provide financial services or legal advice, which may cause deceit by inducing people to rely on such services/advice under the guise of a real person providing those services/advice. This may require changes such as expanding identity fraud legislation and imposing penalties if there is a risk of an individual using someone else's personal information via the generation of "deepfakes" to obtain a benefit, without the subject individual's consent.

b) criteria or requirements to identify AI applications or technologies that should be banned, and in which contexts?

AI applications or technologies that may be used to target vulnerable people (children, the elderly etc.) should be restricted or regulated to mitigate risks to those people. The responses of such groups to AI applications and technologies could be more problematic if that technology can purport to be an adult or other authority figure (such as a teacher).

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

History suggests that the uptake of AI will only increase. Transurban doesn't see a requirement for the government to actively encourage more people to use AI at this stage. This could change as time goes on, if for example dispersed uptake is particularly disadvantaging older generations of Australians.

Implications and infrastructure

12. 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?

Banning forms of social scoring that are determined to be high-risk are unlikely to negatively impact Australia's technology sector. It could be argued that most people do not want to be monitored 24/7 in their daily life for the questionable benefit of improving technology. There are high risks that these "social scores" can be manipulated or gamed.

If such technologies were to be implemented, there must be an ability to "opt out" or challenge a score and each individual must be afforded that right. Individuals should have the right to understand why a decision has been made about them, and whether a decision has been made by automated decision-making. There must be a mechanism for human review, human oversight and ultimately, the ability for humans to override any AI decision.

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

No response.

Risk-based approaches

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

Transurban supports a risk-based approach. It is a reasonable, sensible way to address AI governance and regulations. Transurban supports combining a risk-based approach with the ability to make specific regulations in relation to high risk activities as they emerge over time (for example in relation to "deep fake" technology, referred to above).

As outlined above, Australia's regulatory approach needs to apply at the point of training the AI model, and in the application of those models. A risk-based approach, with say 3 levels (low, medium and high) alone, is limited in its application for all possible scenarios. A better approach is a "matrixed" approach, where the likelihood of a risk and subsequent impact is assessed against a scoring system of low, medium and high. This approach is more suitable to AI by contemplating a risk scenario concerning the use of AI, the likelihood of a problem arising and assessing the impact of that problem. The proposed regulation of three risk levels alone to measure the impact of errors is too narrow. The additional tool of using a matrixed approach to assess the likelihood of an error and its subsequent impact will provide more flexibility and adaptability as the technology develops.

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

As outlined above, Transurban supports a risk-based approach, however the approach is limited if it does not appropriately consider the impact and likelihood of risks of AI technology on individuals, such as the detectability, reversibility, financial and emotional impact of an error.

The draft risk management approach for AI outlined in Box 4 of the Discussion Paper describes 'Medium Risk' as "high impacts that are ongoing and *difficult to reverse*". The phrase "difficult to reverse" is vague. How is this difficulty measured? From a financial impact lens? Or the level of practical difficulty to reverse or detect an error or mistake?

Using the example of billing services, it would be difficult to assess a risk level using the parameters in Box 4 where an AI-generated outcome results in a customer being incorrectly charged. Say that an AI model incorrectly charges customers of an organisation based on an incorrect data input, and no customers raise the incorrect charge with the organisation. The false charge would not be difficult to reverse per se, (i.e. simply refunding the customer the overcharged amount), however if this error were widespread across thousands of customers and difficult to detect, what risk category would apply? Would this scenario be considered "difficult to reverse" on account of detectability rather than reversibility?

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

Most organisations are based on risk management practices, therefore a risk-based approach is the best approach. There may be differences in the maturity of the risk-based approach across different sectors, however considering the variability and breadth of AI generally, Transurban believes a risk-based approach is appropriate.

17.   What elements should be in a risk-based approach for addressing potential AI risks? Do you support the elements presented in Attachment C? Attachment C is found on page 40 of the discussion paper.

Transurban broadly supports the elements contemplated in Attachment C, however the section on training requires more consideration. The proposal calls for providing 'adequate employee training in the design, function and implementation of the AI', however the field of AI is advancing at such a rapid pace that training may become outdated at a rate that exceeds the industry standard. While many training providers offer certifications and courses on AI training, there is no widely agreed and accepted qualification for working within the field of AI.

Will consideration be given as to what constitutes appropriate AI training qualifications and courses? This includes consideration as to the content that is covered in such training, i.e. technical and ethical training. Most training providers currently offer technical and ethical courses separately.

Further, more consideration needs to be given to the requirements of publishing impact assessments "for greater transparency". If organisations are required to publish the intricate details of their impact assessments about their AI models, how would this impact commercial confidentiality principles? It would not be practical to require organisations to publish details of their AI commercial models, however it would be reasonable to request that organisations provide:

1. sufficient summaries of AI model impact assessments to the general public in an easy-to-understand manner; and
2. confidential impact assessments to technical AI experts for peer review.

Any obligations to publish this information should be tied to the risk assessment of the particular AI technology.

18.   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?

In any organisation, managing AI risk assessments is best suited to sit in the first instance within the data and analytics team/data office, with escalation for risk assessments that exceed a predetermined threshold (for example, medium to high risk). This escalation point, such as to a specialist AI committee, should comprise of members from diverse areas such as data and analytics, privacy, risk, ethics, customer and public affairs. In many organisations, this could align with existing processes relating to data governance and privacy.

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

LLM is a tool and not an outcome on its own. It is essential that risks are measured for applications that use LLM and not the LLM per se, except to the extent of identifying and removing bias in the training of the model.

20.   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?

A risk-based approach for responsible AI should be mandated through regulation and applied to both public and private organisations, as well as both developers and deployers.