STRATINNOVA

Subject:	Safe and Responsible AI in Australia
Document:	Survey Responses - STRATINNOVA
Recipient:	DISR - Department of Industry, Science and Resources
Contact:	Rohan Fernando
Email:	rohan@stratinnova.com
Mobile:	+61 407 976 385
Information:	Public
Date:	26 July 2023

Subject: Safe and Responsible AI in Australia

Document: Survey Responses – STRATINNOVA

Recipient: DISR - Department of Industry, Science and Resources

Contact: Rohan Fernando

Email: rohan@stratinnova.com

Mobile: +61 407 976 385

Information: Public

Date: 26 July 2023

Responses to the survey are as follows:

Definitions

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

1.1. The definitions of AI Technologies and AI Applications provided are a good start, however being strictly correct, these all relate specifically to types of AI technologies, including the proposed Applications provided in the Definitions. That is LLM, MfM and ADM are all really types of AI technologies, and these can each be used for an extremely diverse range of different AI Applications.

1.2. In order to identify AI Applications, it may be additionally helpful to add a definition of Intelligence itself, and then create an extremely large list of Capabilities that are enabled by Intelligence. I suggest considering these Capabilities as different dimensions of Intelligence. The Capabilities or dimensions of Intelligence fundamentally determine the different types of AI Applications that can be performed.

1.3. This approach will be helpful with AI, its various Applications, and associated Benefits and Risks, as AI's ability to learn, combine, and perform more Capabilities will grow over time. This is because AI is based on digital technology that continually advances in performance, and Computer information processing rates increase yearly.

1.4. In the 'Mainstream Science on Intelligence' reference paper published in 1997 by Linda S Gottfredson at the University of Delaware, that includes 52 signatories, all experts in Intelligence and allied fields, it says Intelligence is a very general mental Capability that, among other things, involves:

1.4.1. the ability to reason,

1.4.2. plan,

1.4.3. solve problems,

1.4.4. think abstractly,

1.4.5. comprehend complex ideas,

1.4.6. learn quickly and learn from experience.

1.5. Capabilities are dimensions of Intelligence - Intelligence can be observed and described in terms of Capabilities of an entity, irrespective of its specific embodiment, that exist across many different dimensions. While there is no single definitive and exhaustive list, there are some well recognized dimensions of Intelligence, of which most Humans simultaneously possess multiple, such as the following examples:

1.5.1. Adaptive Intelligence - the ability to learn and adapt to new and changing environments.

1.5.2. Bodily-kinesthetic Intelligence - the ability to control and coordinate one's body movements.

1.5.3. Creative Intelligence - the ability to generate and implement new and innovative ideas.

1.5.4. Emotional Intelligence - the ability to understand and manage one's own emotions, as well as the emotions of others.

1.5.5. Empathetic Intelligence - the ability to accurately sense the emotions of others and correctly conceptualize what someone else is feeling and thinking.

1.5.6. Financial Intelligence - the ability to understand and manage financial matters effectively.

1.5.7. Leadership Intelligence - the ability to understand and influence group dynamics and achieve goals through effective leadership.

1.5.8. Linguistic Intelligence - the ability to use language effectively, both in written and spoken form.

1.5.9. Logical-mathematical Intelligence - the ability to reason logically and solve mathematical problems.

1.5.10. Planning Intelligence – the ability to dynamically conceive and temporally construct logically interlinked functional components and higher order processes within a sequential system in order to achieve an objective.

1.5.11. Political Intelligence - the ability to understand and navigate political systems and processes.

1.5.12. Strategic Intelligence - the ability to plan and execute strategies to achieve specific goals.

1.5.13. Systems Intelligence – the ability to understand the functional linkage of components and processes.

1.5.14. Technological Intelligence - the ability to understand and use technology effectively.

1.6. These are just a mere few example Capabilities available and used by Intelligence within Humans.

1.7. Some of these Capabilities are already possible with AI, and there is a potentially unlimited number more.

1.8. It is already self-evident with commercial AI tools available now, such as GPT-4, Bard and Midjourney, that AI can perform some different types of Capabilities significantly better than Humans.

1.9. High Intelligence effectively uses more dimensional Capabilities

1.9.1. The number of dimensions of Intelligence are theoretically unlimited, and presently only constrained by an entity's ability to computationally process information rapidly, and use this information processing to produce an Intelligent outcome that achieves an objective.

1.9.2. There are certainly various levels of Intelligence observed within different entities, and this can best be defined by the ability to learn, perform, and most effectively combine different Capabilities. For example, it can be considered that some non-Human biological Intelligent entities possess unique Capabilities such as 'Radar Navigated Flight Intelligence' – the ability to naturally fly, and simultaneously emit and sense ultrasonic frequencies for spatial echo-location and real-time flight navigation. Bats possess this specific dimensional Capability, however bats do not possess Financial Intelligence, or Technological Intelligence, because they are entities that have not yet adapted to a Computational information processing level required to learn and utilize these Capabilities, whereas Humans have.

1.9.3. The levels and numbers of dimensions of Intelligence across all entities in all its forms and associated variable Capabilities is perhaps only limited by the technology on which it operates. In the case of bats, Humans, and most other entities, this technology is a biological information processor that uses a neural network architecture. In the case of AI, this technology is a computer information processor(s) that use a neural network architecture.

1.9.4. The underlying biological neural network information processing architecture of any entity provides the ability to learn, pattern match, pattern differentiate, process, recall, and respond to input information, so that an entity can produce outputs that enable it to survive and grow in a given environment. One thing seems certain, the raw ability to simultaneously process different types of input information, in addition to extremely large quantities of input information, and then use this to produce output information, take action, and achieve some useful goal, is absolutely central to the Capabilities and associated benefits of Intelligence. An entity that possesses and can adaptively mix and use more dimensional Capabilities has a portfolio of distinct advantages that can be used to survive and grow in most environments, versus less Capable entities.

Potential gaps in approaches

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

2.1. The Canadian approach of categorizing AI risk levels from Low (level I) to Very High (level IV) and associated Bill C-27 Part 3 seems a totally appropriate starting point. Similarly, the EU appears to be taking a sensible approach to AI risks and regulation.

2.2. A very important issue associated with AI Risks is that AI is a technology that has an ability to cause Harm at an extremely high scale of Impact at a single Individual level. Reading through the Canadian Bill C-27, I'd suggest the penalties for Individuals may be too lenient, and Australia might consider making the penalties for serious Harm caused by AI extremely harsh, so as to act as a very strong deterrent.

2.3. I'd suggest the paper presently does not address National Security, Global Security, and Existential risks of AI. The first two are at least Very High (level IV), and the third is in a category of its own, as Catastrophic (level V).

2.4. Nowhere in this document are Catastrophic risks of AI mentioned or addressed. That's probably OK for this paper as it is a public document and it's not helpful to frighten people, however these Catastrophic risks are very real and must be addressed as a matter of priority by Government. Given a lifetime career working with a wide range of very advanced digital technologies, automation systems, and AI systems, I'd suggest these have the potential to occur much more quickly than is generally expected. Mitigative steps really need to occur immediately in a well-considered, coordinated, and professional manner, as a Federal Government initiative.

2.5. The list of specific Risks associated with AI is too large to define and document as there are potentially trillions of permutations and combinations in the world. It is perhaps more useful to analyze and list and group different types of AI Harm to Humans, Environment, Economy, Government, Financial Systems, Food Supplies, Manufacturing, Transportation, Critical Infrastructure, Cybersecurity, etc, and then match these against the existing and emerging Capabilities of AI as it develops with time. This may help to write laws that encapsulate and appropriately control the Capabilities and how they are permitted to be applied to various Applications across Australia.

2.6. Analyzing AI in terms of potential types of Harm and then regulating preventative measures is a logical approach. By way of analogy, a Gun is both useful and lethal, and there are laws that control access and limit the potential to cause Harm, and there are penalties associated with misuse and levels of Harm. The same approach can be used for AI, but note that AI will increasingly develop Capabilities and be able to do anything a Human can do... and potentially more in the future.

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.

3.1. It seems the non-regulatory issues with AI are quite manageable and although it can be anticipated that outlier risks and Harm may pop up in this area, these can be addressed on a case-by-case basis.

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.

4.1. I'm presently not involved in Australia's AI governance across government, so cannot make a useful comment.

4.2. I would welcome the opportunity to become involved and provide some assistance.

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?

5.1. The EU AI Act, Canadian, and Chinese approaches all appear to be quite comprehensive. There would be benefits to harmonizing Australia's approach with the best of various international approaches.

5.2. Given how incredibly advanced China is in both the development and applications of AI, I'd suggest it would be useful to consider how they are approaching AI Risks, regulation and the control of Harm. The Chinese Government is managing an enormous population, and their relatively strong approach to the management of citizens will provide some unique insights into how their government measure and manage the risks of AI within their own nation.

Target areas

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

6.1. It may be prudent to apply precisely the same regulatory approaches for AI to the private sector and public sector. The public sector directly engages with every citizen of Australia in one way or another, and consequently has the potential to incorporate AI Applications that could possibly cause the most widespread Harm and impact. The private sector will largely follow the public sector's lead, in so far as, what the private sector can get away with in AI Applications. I'd suggest the public sector must lead by example with the use of AI and following AI regulations, and use its force of law to very firmly realign any business in the private sector that actively attempts to deviate from the AI regulatory path.

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

7.1. Clear guidance on acceptable AI usage, including extremely strict laws for very high risks, through to recommended guidelines for generally low risks, that are very well communicated, with appropriate training of government personnel, and particularly IT staff and contractors, will be quite helpful.

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.

8.1. Generic solutions may best be applied to high AI risks, and technology-specific solutions may be better suited to specific Application risks elsewhere. High AI risks with high scale impacts primarily stem from access to specific classes of training data, and its applications, that is only available to groups with specialized knowledge.

8.2. It would be prudent to establish very broad and extremely strong regulatory controls on these generic types of AI training data and AI models, with extremely severe penalties for any breaches.

8.3. These are not so much technology issues, but rather more related to actionable information and its potential for deliberate misuse in the wrong hands.

8.4. Eg. Strictly controlling authorized access and use of Drug Design data sets, and associated pre-trained AI models might be a good thing to do. This type of training data or trained AI model weights, floating around in the general public, could become extremely harmful to millions of people worldwide.

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

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

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

9.1. A suggested approach to consider is to establish an Australian Federal Government AI Management Agency (GAIMA) immediately, that all suppliers of AI hardware, software, and application services in Australia are required to comply with. This approach might seem unusually extreme right now, but it will be enormously helpful in protecting Australian citizens quite soon as AI computing power increases.

9.2 Implement mandatory AI interaction and usage requirements to advise Humans using AI on the following:

9.2.1 RAIT - Ranked AI Type - to advise Humans on the AI model type they are interacting with, as officially classified by GAIMA.

9.2.2 RAICM - Ranked AI Computation Maximum - to advise Humans on the maximum AI Computational Capabilities they are interacting with, as officially classified by GAIMA.

9.2.3 HTAIR - Human to AI Ratio - to advise Humans on the ratio of Human to AI information content they are interacting with, as officially certified and classified by GAIMA. ie. 100% Human, 100% AI, or some range in-between such as Human response with AI support from RLHF (Reinforcement Learning Human Feedback) and associated amounts of Rewards and Penalties that an AI has received. There is a way to measure how much information an AI provides is purely AI generated, and how much information comes directly from Human feedback. This is done by tracking the number of times an AI is given a reward or a penalty for a particular response. The more rewards an AI receives, the more likely it is that the response was purely AI generated. The more penalties an AI receives, the more likely it is that the response came directly from Human feedback. For example, if an AI is asked to provide a summary of a factual topic, and it provides a response that is accurate and informative, it is likely to be given a reward. If an AI is asked to create a story, and it provides a response that is creative and engaging, it is also likely to be given a reward. However, if an AI is asked to provide a response to a question, and it does not know the answer, and it simply copies and pastes the answer from a website, it is likely to be given a penalty. So, by tracking the number of rewards and penalties an AI receives, it is possible to get obtain an indicative measure of how much information the AI provides that is purely AI generated, and how much information comes directly from Human feedback. Importantly, the amount of information an AI provides that is purely AI generated is constantly changing as it learns and improves. Therefore, an AI is able to generate more and more information on its own. However, an AI may always need some input from Humans in order to learn and improve.

9.2.4. LAISID - Licensed AI Supplier Identification - to advise Humans on the identity of the legally licensed AI supplier they are interacting with, as officially classified by GAIMA.

9.2.5. LAIA - Licensed AI Applications - to advise Humans on the complete set of AI applications they are able to legally access from a legally licensed AI supplier, as officially classified by GAIMA.

10. Do you have suggestions for:

a. Whether any high-risk AI applications or technologies should be banned completely?

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

10.1. There are numerous actions that could be implemented to protect Australians from high-risk (& catastrophic risk - level V) AI applications and technologies, and it is suggested that these could be implemented under the supervision of GAIMA. Again, the following list may seem totally outlandish and extreme right now, however neural networks on which AI is based have extraordinary information processing capabilities that are yet to unfold within developing digital technologies. The following will make sense very soon.

10.2. Implement Computer processing AI calculation per second (CPS) rate limits:

10.2.1. establish AI system Computer processor development technology limits, with processor types and quantities applied to any single AI application requiring GAIMA approval.

10.2.2. potentially restrict or outlaw quantum Computing for both AI model learning and AI model runtime (post-learning) applications.

10.2.3. for every AI application establish a maximum calculations per second (CPS) upper limit for any AI system, including single AI system and multiple aggregated AI systems. Perhaps with CPS kept at 1% of Computational Hard Limit of 1 Human Brain as defined by Kurzweil (ie. 1 Human Brain Upper limit = 20 quadrillion CPS x 0.01% = 0.2 quadrillion CPS).

10.3 Implement AI energy access limits and immutable Emergency Stop systems:

10.3.1. limit the amount of energy that can be supplied to any single AI system.

10.3.2. implement mandatory emergency stop on power supply systems for every AI system including multiple serialized fail-safe mode including uninterruptible remote access and immutable E-stop control by GAIMA.

10.4 Implement AI system networking limits:

10.4.1. limit the number of AI systems that can be interconnected together at any time.

10.4.2. multiple interconnected AI systems limits must not in aggregate exceed the CPS limits for a single AI system.

10.4.3. AI systems cannot batch process and offload CPS functions to another remote AI System or other general Computer resources such as High Performance Computing datacenters without authorized GAIMA approvals.

10.4.4. mandatory fail-safe emergency network link breakage on every AI system communications interface, with no exceptions, including all wired and all wireless communications links of any type.

10.5 Implement AI system input information limits:

10.5.1. limit the different types of input information (eg. different sensor types) that can be supplied to a single AI system.

10.5.2. limit the total number of inputs permitted to be connected simultaneously to a single AI system.

10.5.3. communication networking of different AI systems with different types of input information must be approved by GAIMA.

10.5.4. communication networking of different AI systems that produce an increase in total aggregate input numbers, must be approved by GAIMA.

10.5.5. mandatory fail-safe emergency input information disconnection on every single AI system.

10.6 Implement AI system output information and control limits:

10.6.1. limit the different types of output information that can be supplied from a single AI system.

10.6.2. limit the different types of output information that can be supplied from multiple AI systems that are connected together through communication networking.

10.6.3. limit the number of total outputs permitted to be connected simultaneously from an AI system.

10.6.4. limit the number of total outputs permitted to be connected simultaneously from multiple AI systems that are connected together through communication networking.

10.6.5. communication networking of different AI systems that produce an increase in total aggregate output numbers, must be approved by GAIMA.

10.6.6. mandatory fail-safe emergency output information disconnection on every single AI system.

10.7 Implement AI model access limits:

10.7.1. official classification of all AI models including all new classes of AI developed (eg. Transformer), including submission of model algorithms for review and licensed approval by GAIMA.

10.7.2. registration of all AI models with GAIMA including new classes.

10.7.3. licensing of AI model usage on every AI system.

10.7.4. safety and ethics training and certifications on AI model usage.

10.7.5. annual auditing and certification re-approvals by GAIMA. (eg. similar to ISO / NATA testing for laboratories).

10.8 Implement AI application monitoring with a Supervisory Control and Data Acquisition (SCADA) system linked to GAIMA:

10.8.1. realtime reporting of AI applications on every AI system available to GAIMA.

10.8.2. realtime reporting of AI usage levels on every AI system available to GAIMA.

10.8.3. realtime reporting of AI model types used on every AI system available to GAIMA.

10.8.4. realtime reporting of AI Computer processing levels on every AI system is available to GAIMA.

10.8.5. realtime ability to shutdown any AI system from GAIMA using fail-safe controls.

10.9 Implement AI application controls:

10.9.1. nuclear AI application controls.

10.9.2. biological AI application controls.

10.9.3. chemical AI application controls.

10.9.4. robotic AI application controls.

10.9.5. military AI application controls.

10.9.6. general microprocessor and AI microprocessor development controls.

10.9.7. neuromorphic Computing system and software development controls.

10.9.8. quantum Computing system and software development controls.

10.10 Implement AI model generation, learning, processing, and optimization limits:

10.10.1. strictly regulate and potentially outlaw the use of quantum Computers for AI model error optimisation and AI learning.

10.10.2. strictly regulate and potentially outlaw the use of neuromorphic processors for AI model error optimisation and AI learning.

10.10.3. strictly regulate and potentially outlaw the use of genetic algorithms for AI model evolutionary development on high performance Computing systems.

10.10.4. develop and institute a global equivalent to the International Traffic in Arms Regulations (ITAR) controls, for technologies including neuromorphic processors, quantum Computers, and emerging optical Computing.

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

11.1. Many people are already fearful of AI, so it is important to provide the public with education to raise awareness of the many incredible benefits that AI can actually provide.

11.2 Provide example use-cases where AI has been used to provide benefits to the public and a wide range of industries. This will serve to foster greater trust in AI and the government's role in ensuring it is being kept safe. It is important to demonstrate and message that AI risks are being controlled.

11.3 Provide examples in how government is strictly regulating AI to keep the public safe, in both public and private/commercial applications.

11.4. This needs to be long-term media campaign.

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?

12.1. It is suggested that high-risk AI applications will most typically be used to Harm, coerce, and exploit people and businesses. Therefore, it seems logical that banning high-risk AI applications in Australia is likely to be perceived favorably by other countries, as it will mean Australian citizens and businesses are generally more trustworthy than countries that do not impose such high-risk AI bans. The more professionally, ethically, and reasonably that AI is deployed, used, and supervised and managed through regulation in Australia, the better it will be for both Australia and its international trade partners.

12.2. It can be anticipated that countries that do not impose controls on high-risk AI will see their AI hardware, AI software, and AI products and AI services fully banned by other nations. By way of example, consider the bans imposed on some computer networking technology suppliers because they have been deeply technically analyzed and are considered to incorporate 'backdoor' cybersecurity risks. The same types of extremely firm international trade bans will apply to high-risk AI systems as they come to light.

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

13.1. The existing Standards and Regulatory bodies in Australia provide an excellent portfolio of capabilities for mitigating many of the potential AI risks. The issues and risks with AI may emerge very quickly and unexpectedly as AI gains increasing dimensional Capabilities, so this conformity infrastructure will need to be extremely well informed on AI issues and risks, and then be adaptive to change and imposing new recommendations and controls that Australia can use effectively. Speed of response will be very important.

13.2. As previously mentioned, a suggested approach to consider is to establish an Australian Federal Government AI Management Agency (GAIMA) immediately, that all suppliers of AI hardware, software, and application services in Australia are required to comply with. The level of conformity and compliance that is imposed by GAIMA could be directly related to the AI risk level.

Risk-based approaches

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

14.1. A risk-based approach makes sense. This could be analyzed in terms of different types of Harm that AI can cause, as suggested above.

14.2. Breadth and depth of Harm will both need to be considered.

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

15.1. A risk-based approach is logical.

15.2. The only real limitation with a risk-based approach for AI, is that it is prudent to fully expect AI will progressively become more Intelligent than most, if not all, Humans.

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

16.1. It will be important to focus on sectors, applications and organizations that have the most potential to cause Harm, particularly at widespread scale. eg. Government Services, Critical Infrastructure, Emergency Services, and Safety Critical Systems used in various Industries, such as Transport and Traffic Management.

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

17.1. The approach outlined in Appendix C covers most issues, however Automation in particular can have very significant impacts on Humans that are not always clearly obvious at first.

17.2. Automation typically has truly zero concern for Humans, as Automation has no comprehension of its Human impact, so it is critical to envelope Automated systems inside very thoroughly designed and robust protective external systems and barriers to appropriately protect Humans. This is true for both hardware and software-based Automation.

17.3. Consider developing and instituting a regulatory framework and processes for providing Humans with a strictly defined methods for recourse and contesting the outcomes and impacts of Automated systems.

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?

18.1. There are already numerous existing laws in place, many of which are designed to protect Humans from Harm. Therefore, it is important to analyze existing legal policies, eg. Australian Privacy Principles, and determine ways in which the Capabilities of AI can pierce through the protections that these policies and laws provide, and can enable harm.

18.2. Where this occurs, legal amendments will be required as a matter of urgency.

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

19.1. LLMs, MFMs, and several other AI models are just variations of the same underlying fundamental neural network technology.

19.2. It is perhaps more useful to focus on the Capabilities that various AI models enable, and design a suite of risk-based approaches to controlling the harm that can be caused by these various Capabilities.

19.3. All AI models have a baseline level of risk associated with their Computational information processing rates and volumes, which is defined by the entire computing infrastructure they employ. Section 10 above provides some suggestions on how a risk-based approach could be implemented. However, I'm not suggesting this will be easy to achieve given the wide range of commercial and global competitive pressures that exist with AI right now. Nevertheless, section 10 is really about Existential level V risks.

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?

20.1 It will be prudent to be extremely skeptical that voluntary self-regulation of AI will be able to professionally and correctly manage the diversity of risks of AI, and in particular, limit the Harm of totally unexpected and extremely 'Intelligent' but highly infantile solutions produced by AI.

20.2 It is prudent to apply enforced regulation to public and private organizations.

20.3 It is prudent to apply enforced regulation to developers and deployers of AI.

Please let us know if you would like any further clarification on these survey responses.

Yours faithfully,
Rohan Fernando
STRATINNOVA