ENGINEERS
AUSTRALIA

Responsible AI in Australia:

An engineering perspective

Engineers Australia's submission to the Safe and Responsible AI in Australia discussion paper

July 2023

Responsible AI in Australia: An engineering perspective

The report can be downloaded at engineersaustralia.org.au

Engineers Australia
11 National Circuit, Barton ACT 2600
Tel: +61 2 6270 6555
Email: policy@engineersaustralia.org.au

engineersaustralia.org.au

Contents

Introduction

Engineers make a significant contribution to our society and economy and are vital to overcoming many of the challenges we face. Their expertise can be applied in many areas, including the development, implementation, integration, use and maintenance of artificial intelligent systems and products. The rapid uptake and constantly evolving nature of artificial intelligence means the development and uptake of these technologies has become a significant focus in many industries.

Greater use of AI has been likened by some Engineers Australia's members to the introduction of the calculator. It has the potential to significantly change the professions, increasing productivity and enhancing an engineer's capability to solve complex problems. However, it may also have negative impacts on the profession, by minimising the inherent knowledge engineers possess about science and mathematics and reducing their ability to think critically. Many AI applications already exist, with countless others continuing to emerge in fields directly related to engineering. Engineers, therefore, must embrace AI as a tool to embrace their capability and drive innovation.

While certain engineering disciplines are well known as playing a central role in building and deploying AI systems (such as software engineers), it is essential to recognise the expertise of engineers from all fields is equally indispensable. Fields such as biomedical engineering, chemical engineering, industrial engineering, systems engineering, and sustainable engineering are increasingly utilising AI systems. The use of AI in these areas and more, demands subject matter expertise, to enable these systems to reach their potential and to safeguard against negative impacts.

Engineering is the practical application of mathematics and science. As such, it is responsible for translating current knowledge into practical applications. When it comes to AI systems, engineers have a responsibility to develop and implement the technology in a safe and ethical manner. Engineers, developers and companies implementing AI systems and products, need to do so in an ethical way and hold themselves accountable for their work. This responsibility means those developing AI cannot rely on disclaimers such as "use with caution" to absolve themselves; this is particularly critical when dealing with vulnerable sections of society. By harnessing the potential of AI responsibly and with a deep understanding of their fields, engineers can share the future of the profession and the use of AI to positively impact the community.

About Engineers Australia

Engineers Australia is the peak body for the engineering profession in Australia, constituted by Royal Charter, to advance the science and practice of engineering for the benefit of the community. We are the collective voice of over 115,000 members representing all disciples of engineering.

As Australia's signatory to the International Engineering Alliance multi-lateral accords, Engineers Australia maintains national professional standards, benchmarked against international norms. We evaluate engineering courses against entry-to-practice competencies to determine whether graduates will meet international benchmarks to practise.

The paper is guided by our Royal Charter and Code of Ethics, which states engineers act in the interests of the community, ahead of sectional or personal interests, working towards a sustainable future. As engineers are largely responsible for the development and implementation of AI systems and algorithms, Engineers Australia is exploring the development of an AI Area of Practice, that would support engineers working in AI to uphold standards which are deemed a requirement in this field.

Contact

To discuss the content of this submission further, please contact Michael Bell, Senior Policy Advisor, at mbell@engineersaustralia.org.au or phone +61 8 64214 6321.

The impact of AI on the engineering profession

The engineering profession is not immune from the impact of AI. The impact of AI on the profession is already evident and is expected to grow. AI can fundamentally impact the profession through its ability to:

- Optimise the design process.
- Improve modelling.
- Increase productivity.
- Enhance robotics.
- Extract more meaningful analysis from data.

By integrating AI into engineering systems and processes, industries can benefit from increased effectiveness and efficiency, through reduced costs and the ability to design systems with new levels of capability.

There is an increase in AI systems and technologies becoming tools for engineers in their work. The influence of AI on engineering is expected to be profound. It is already an essential component of information and control technologies across industries. While it has the potential to enhance the reliability and efficiency of systems, it does carry risks. Below are some examples of different areas of engineering using AI.

Transport:

AI enabled Digital twins for transport infrastructure which takes in real-time data to predict efficiencies and reliability of assets. Automated route planning and road access for heavy vehicles is filled with potential for AI applications. Engineers Australia's *Future of Transport* discussion paper outlines how flight plan optimisation and management technologies can allow more aircraft to take-off and land each hour.[1] AI can analyse weather, flight congestion and other data much faster than a human, which saves time.[2] This allows extra capacity to be squeezed out of the system and reduces the need for additional infrastructure.

Medical and healthcare:

Most commonly when the use case of AI is considered in the medical profession, it's use in diagnostics is the first use case. However, the use of AI is growing in the fields of rehabilitation and disability support, and hospital admissions. There are several emerging applications of AI technologies in rehabilitation engineering. This includes providing non-tactile interfaces for clients to engage with technology. An example of this autonomous monitoring of vital parameters and automatic escalation (for example pacemaker monitoring which would extend to intervention and escalation).

Manufacturing:

The fourth industrial revolution (Industry 4.0) incorporates AI with additive manufacturing and other advanced processes. As AI progresses it will be able to replace human involvement and decision-making in various processes by utilising smart technologies, algorithms and digital tools.

As the technology progresses it will become more reliable at handling increasingly complex tasks. We are already seeing AI capable of designing new chemical compounds, including

[1] Bell, M. 'Future of Transport: Discussion Paper' *Engineers Australia* (January 2023) https://www.engineersaustralia.org.au/sites/default/files/2023-01/future-transport-discussion-paper-jan-2023.pdf
[2] Kwon, K. 'AI: how it's delivering sharper route planning' *Aerospace America* (February 2023) https://aerospaceamerica.aiaa.org/features/ai-how-its-delivering-sharper-route-planning/

medications, which can then be manufactured.[3] AI and the development of quantum computing opens new possibilities for advanced research and development, allowing for innovative breakthroughs in various industries.

While not exhaustive, these examples show how AI is already impacting the profession and the opportunities this is creating. However, these examples also identify clear risks to the use of this technology. As engineers, it is our responsibility to navigate the ethical and technical considerations of AI to harness its potential, while ensuring the safety and wellbeing of society.

Summary of recommendations

An AI system is an engineered system. It is defined with inputs and outputs and operations through a programmed, albeit complex, calculation which uses the inputs to generate an output. Due to this, at its core, AI does not have intent (any intent is programmed in) and its behaviour is deterministic. Australia needs to strike a balance in the regulation of AI, to allow the benefits to be realised, while also ensuring the negative aspects of the technology are mitigated. This balance should take a risk-based approach to ensure the safe and responsible use of the technology.

The approach for governing responsible and ethical AI should incorporate elements of voluntary actions, guidelines, information and education, and legislative. Any regulations should be focused on areas where AI technology is considered high-risk and requires public protections. Regulations should also focus on control systems before bans are considered. The primary focus of regulation should be on the implementation and use of the systems; however, developers should also be held responsible for the outcomes in some circumstances, to ensure they operate as expected.

- The definition of AI should be more aligned to an AI system being an engineered system which generates outputs, such as content, forecasts, recommendations or decisions, based on learning data and given inputs and/or parameters. It is a system which is designed to operate with varying levels of automation.
- The definition for large language models should be change to '…is a type of generative AI that specialises in generation of symbolic sequences such as Human-Like text.'
- A more focused approached should be taken to identify risks and options to manage them. Government should engage industry and academia through taskforces, to focus on specific tangible risks and challenges. This will allow greater oversight of what regulatory action is needed.
- When it comes to non-regulatory initiatives, Engineers Australia recommends a broad approach should be taken which combines regulations and obligated good practice with information campaigns, training, guidelines and frameworks.
- Engineers Australia further recommends developing and promoting industry guidelines for ethical and responsible AI practices.
- When it comes to governance, while general principles can be agreed across sectors, it is recommended that the governance approach is more sector specific. This will make allowance for the needs and risks of different industries.
- Engineers Australia recommends the government establish senior engineering roles with expertise in AI within government to oversee, guide and advise on AI systems.

[3] Field, H. 'The first fully A.I.-generated drug enters clinical trials in human patients' *CNBC* (29 June 2023) https://www.cnbc.com/2023/06/29/ai-generated-drug-begins-clinical-trials-in-human-patients.html#:~:text=The%20first%20fully%20A.I.%2Dgenerated%20drug%20enters%20clinical%20trials%20in%20human%20patients,-Published%20Thu%2C%20Jun&text=Insilico%20Medicine%2C%20a%20Hong%20Kong,fibrosis%2C%20a%20chronic%20lung%20disease.

Consultation questions

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

The definition of AI in the discussion paper is more closely aligned to an AI system, rather than artificial intelligence. AI is the capability of machines and systems to acquire knowledge and apply it to perform intelligent actions, simulating human intelligence in specific applications. This includes large language models.

AI, machine learning and data science are all interconnected disciples. These use mathematical and computer science models to process data and produce either insight, action, more data or more advanced AI capabilities. These models, including large language models, computational models and semantic-based data models, can become intricate and result in the inability for humans to understand them. As such, these models' become algorithms that may have unpredictable consequences.

Any definition needs to recognise that AI is developing quickly into emerging technologies that we do not yet fully understand. Any definition needs to be broad enough to encompass these.

It is recommended the definition be more aligned to an AI system being an engineered system which generates outputs, such as content, forecasts, recommendations or decisions, based on learning data and given inputs and/or parameters. It is a system which is designed to operate with varying levels of automation.

In addition, the definition for the application of Large Language Models (LLMs) is not broad enough. Stating that it specialises in the generation of 'Human-like' text no longer matches current capabilities. LLMs are capable of learning and generating many symbolic forms of language including computer programs in multiple programming languages and even binary machine code and machine to machine protocols. It is recommended the definition change to '…is a type of generative AI that specialises in generation of symbolic sequences such as Human-Like text.'

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?**

The discussion paper covers the risks and options to manage these risks well at a high level. However, it is recommended that a more focused approach is taken in consultation with industry and academia, to focus on specific tangible risks and challenges. Only then can regulatory action be considered. It is also important to have a plan to review the landscape and known risks on a regular basis.

Comprehensive regulatory frameworks will be essential to ensure the responsible and ethical development and implementation of AI systems and products. Even where existing regulation provides protection, it is unclear whether these will be robust enough to cover the breadth of AI's capabilities. There are likely to be regulatory gaps for ensuring the resilience of infrastructure, supply chains and communities from the impact of any adverse outcomes from AI causes.

Privacy and ownership rights for example will require regulators to consider the protection of individuals' likeness, including their face, voice and other biometric markers. This is particularly the case with the advancing 'deep fake' technology. Generative AI systems, such as deep fakes, are based on observation of humans and their data to gain knowledge and understanding and train the AI models. Ownership also relates to the outputs of AI, such as generative AI. What are the copywrite implications, currently providers believe they own the output of generative AI systems.

There is also a need for clarity regarding the collection, storage and use of human data in AI systems. Individuals need to be informed about when, how and what data is being collected, the nature of the collected data, and the implications of their interactions with AI systems. Risks arise from the quality and purpose of the data fed into AI systems. For example, deep fakes can generate convincing fake content that can be misused. Regulatory measures can address these risks by ensuring transparency about fake or AI-generated content and exploring authorship transparency for accountability.

There will also be niche areas which will need to be considered. An example of this is the use of artificial intelligence in warfare, particularly automated targeting systems. Defence applications of AI currently require human intervention at some level, however there are emerging scenarios where this is not preferred. This is primarily due to issues around response times. Having 'human boundaries' in high-risk areas will be crucial to retain society's confidence in how these systems are developed and used.

Furthermore, as AI progresses, there is a need to consider artificial forms of life such as robots imbedded with AI systems, looking and behaving similar to humans. It will also have an economic impact, particularly how the evolving technology will impact employment and create job displacement.

For regulation to be effective, there needs to be an individual or body corporate held accountable. In some cases, such as the collection and protection of data, this is more straightforward. However, less straightforward examples may require developers to be held to account. This is the case as autonomous systems become more sophisticated and take independent action. An example of this would be autonomous cars. Should the user of the autonomous vehicle or the developer be liable? This is complex and requires careful consideration. Drawing on the definition above, AI generates outputs based on learning data and made decisions based on that data and patterns.

To mitigate this risk, a framework needs to be urgently developed to ensure transparency in the way AI is developed and the data it learns from. Regulation can also be implemented to require ongoing monitoring and evaluation of AI systems.

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.**

A comprehensive approach is recommended which would involve a combination of regulation, information campaigns, training, guidelines and frameworks.

Providing information on the responsible use of AI and the risks of misuse would support broader knowledge and a common understanding of the technology. This would support a more constructive use of AI systems and products and allow more informed discourse.

Training and education initiatives should be implemented. This includes training in the design, implementation, operation, monitoring and maintenance of AI systems. Training should upskill users and developers on issues (such as biases) and showcase how human involvement can be implemented in the technology. In addition, it is also important to provide education about the need to utilise AI as a tool, not a replacement for human critical thinking and problem solving. This is particularly important in education institutes, where students might miss out of tacit learning if they rely too heavily on AI, such as generative AI using large language models.

Another initiative to promote greater collaboration and inclusion in the development of AI tools. This means giving people from all kinds of backgrounds, genders, ethnicities and religions an opportunity to contribute to the development and maintenance will help to mitigate the risk of bias and incorrect outputs. Developers of AI systems need to keep this front of mind when deciding the data these systems will learn from.

In addition, more work is needed on standards development, both in Australia and in the international domain to provide better principles and frameworks for managing risk and opportunities presented by AI. Some industry specific work is already being done in this area, such as the IEEE standards to advance trustworthy AI systems.[4]

Developing and promoting industry guidelines for ethical and responsible AI practices is also recommended. This will help business understand the importance of responsible AI and encourage them to adopt these practices. In addition, the government could provide funding for research and development of projects specifically focused on developing ethical and responsible AI technologies.

While non-regulatory initiatives are encouraged in certain circumstances, there is a strong need for regulation depending on the risk profile of the AI system to ensure the responsible and ethical use of AI. This regulation will be vital to protect against potential misuse of this technology.

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.**

There is a need for a coordinated approach to AI governance. Firstly, goals should be established. These should include:

- Ensuring AI has quality assurance.
- Identifying the limitations of each generation of AI.
- Detecting and eliminating discrimination and biases in AI.

While general principles can be agreed across sectors, it will be challenging to have a single coordinating body. It is recommended governance approaches be sector-specific, and consider the needs and risks of different industries.

Areas where broader governance could apply would be around individual protections and privacy. This is important as it would help address potential risks and provide safeguards and privacy rights, helping to increase trust in the use of this technology.

It is further recommended government bring together policymakers (particularly with STEM backgrounds) subject matter experts, academia and legal and industry representatives to collaborate on regulatory requirements. This will avoid conflicts in how the technology is governed and provide a unified voice on AI governance.

As a starting point, government should create a dedicated taskforce with the remit to establish and execute a framework for collaboration between industry, academia, and government with a goal to promote research and development of AI technologies that align with ethical and responsible practices.

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?**

Work is being done to consider how countries and organisations are approaching AI. The Future of Life Institute investigates the challenges associated with AI development which need to be addressed.[5] It is recommended this research is reviewed.

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

[4] 'IEEE GET Program: GET Program for AI Ethics and Governance Standards' *IEEE* (accessed 14 July 2023) https://ieeexplore.ieee.org/browse/standards/get-program/page/series?id=93
[5] Yelizarova, A. 'Global AI Policy' *Future of life institute* (16 December 2022) https://futureoflife.org/resource/ai-policy/

The basis of approach for both public and private sectors should adhere to similar underlying principles. However, due to the different nature of the sectors, there are considerations which will alter the approach. A risk management approach should be taken, and in areas where the decisions have greater influence and impact on outcomes, more transparency should be required.

Another distinction is AI's use in industrial applications such as manufacturing. Industry 4.0 technologies encompasses transformative technologies like AI. This is often deployed to improve productivity of the production process. In this case, the emphasis is on reliability and safety as opposed to privacy and confidentiality of personal information, and therefore a different approach to regulations are needed.

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

The key to support responsible AI practices is education and ensuring those who are involved have the right qualifications, knowledge and competency. Engineers Australia recommends the government establish senior engineering roles with expertise in AI within government to oversee, guide and advise on AI systems. Having these specialists within government will assist in being able to ensure the system is operating as it is intended, but more importantly, are able to advise on solutions to fix any issues which arise.

In addition, strong administrative guidelines and processes should be established. This includes establishing:

- Robust processes for the validation of data to be used by AI algorithms.
- Independent validation and verification of the system and adherence to any relevant legislation.
- Independent legal review to ensure any AI system is compliant with the legislation being administered by the agency.
- Protocols to ensure transparency through checks and balances, including independent auditing of application outcomes.

Once confidence is built, there will be a need for a regulatory body to continue monitoring the application and can act as an Ombudsmen for the technology. Another option would be to implement dedicated AI governance units within each agency to oversee the implementation and enforcement of responsible AI practices and to encourage an open approach to collaboration between agencies.

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.

Technology evolves rapidly, risk assessment and mitigation will most likely need to be an ongoing exercise. Technology specific solutions are more effective. This is because AI involves a spectrum of capabilities which, depending on the use, require specific management.

Generic solutions to the risks of AI are generally more valuable when addressing broader systemic issues that affect multiple AI systems, such as bias and fairness in algorithmic decision making or accountability mechanisms for AI systems. Sector- or use-specific solutions would be better for more specific use cases. For example, different applications, such as autonomous vehicles and industrial manufacturing processes, require unique approaches to validation, verification, and understanding safety limitations. Generic solutions to the risk of AI would be most valuable in areas such as cybersecurity architecture, to protect data.

When it comes to standards, some generic standards such as ISO 31000 Risk Management, ISO 31010 Risk Management – Risk assessment technique and ISO 31050 Guidance of managing emerging risks to enhance resilience will be applicable. However, specific standards for AI are also likely to be needed. There will also need to be resilience standards which provide principles for responding to risk management failures and rapid changes of AI capability.

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?**

Increasing transparency in when and how AI is used will help to build confidence and increase user trust. Transparency is critical at the point of data collection, including what data is being collected, how and where it will be stored, who will have access and how it will be used. It should also be clear as to if the data will be used to train future AI systems. Greater transparency should be required when the systems are used to assist in making decisions related to individual circumstances.

It should also be disclosed when an interaction is with a human or when it is output from an AI system. For example, if a human is chatting online, make it clear whether they are chatting with another human or getting AI generated messages. If there is a combination of human and AI generated messages (e.g., human moderation of the AI generated message) this should also be disclosed.

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

Firstly, before applying any mandates to transparency, it needs to be defined in the context of AI systems. Transparency to one, may be different to others.

Once a definition is established, it would be beneficial to mandate the disclosure of what data was used to train the AI system. It is particularly important to know if the data was open source or if the AI system is only using internal data. It is also important to know the data stamp of the most recent data used in the training model. For example, if data from the last 12 months is excluded, then the outputs may not be accurate to current situations.

In the example of disclosing whether an interaction is with a human or AI system, it would be as simple as an automated disclose similar to what we hear when we are calling an organisation and are told the call 'will be recorded for training and monitoring purposes'.

10. Do you have suggestions for:

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

The difficulty with banning AI applications is the range of different uses. For example, Generative AI can have extremely negative outcomes if used for malicious purposes. If you take this as the criteria, there are many applications where the technology should be banned, however this is also the case with many different products available today. Through dedicated taskforces, greater engagement is recommended with experts in risk as well as AI systems to understand use cases and criteria before high-risk systems are banned. Ongoing and active monitoring of high-risk applications and technologies could also be done by a dedicated regulator, with the assistance of a certification scheme. It should also be noted, banning technologies will be difficult with many available for download via the internet through various channels.

At the minimum, high-risk AI applications should have a requirement for 'human in the loop/oversight assessments' when used for decision making in high-risk areas. Sector specific governance models could establish frameworks/ For example in areas such as judicial decision making and finance.

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

More context of the individual use cases is needed to be able to establish criteria as to if an AI application or technology should be banned. AI technology and application are evolving quickly, and it will be difficult to propose bans without a thorough, and ongoing risk assessment. At a high level, AI applications which cannot be proven to be safe (based on a developed criteria) could be banned.

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

History shows that countries which leverage emerging technologies, benefit from them the most. Training and education will play a significant role in increasing public trust. Currently many people don't know how to use these various technologies, or what they are/can be used for. Some initiatives to consider include:

- Clear regulations: We should ensure that the regulations governing the development, deployment, and use of AI systems are clear and easy to understand.
- Explanations and transparency: Where practical, AI systems should be designed to provide clear explanations for their decisions, outcomes, and actions. This will help users understand why an AI system made a particular decision or recommendation, increasing trust in the technology. The field of "Explainable AI" could be a good area into which government can supply research funding.
- Privacy protection: Strong data protection laws and regulations help build trust among users who are concerned about their personal information being used by AI systems without their consent.
- Ethical guidelines: These can help ensure that AI systems are developed and deployed in ways that promote social good and do not cause harm to individuals or society.
- Public education: An investment in public education campaigns to increase awareness and understanding of AI technology will help build trust among users who may be hesitant to use AI systems due to lack of knowledge or misinformation about the technology.

For higher risk use cases of AI, a certification scheme will help with increasing trust.

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?

Australia's domestic technology sector will need to incorporate these technologies into their products and services or will risk becoming obsolete by international players. Banning high-risk activities may help to protect Australian consumers from potential harm. It also may help increase trust in the tech sector and lead to increased demand for Australian products and services. However, to ensure Australia remains internationally competitive, bans should only be considered if they meet a pre-defined criteria and control systems are not available.

Bans could also undermine existing AI applications currently used in both private and public sectors. It is crucial to establish and modify existing regulations to safeguard privacy and individual rights in relation to AI, machine learning and data science within the country. Any potential bans should also be differentiated between use and research. It may be acceptable for a technology to still be used for research purposes.

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

The process of developed AI and machine learning models often results in a set of parameters that cannot be derived through alternative means. This means a level of trust in the training of the model is required.

To address this, greater transparency is required on the input training models and techniques, and the resulting AI models be provided with the algorithms for use. This will allow the user to understand the inherent limitations of the model. This could be achieved through standards or other forms of compliance and/or assurance processes.

Improved oversight and regulation of AI development and deployment, including the creation of new regulatory bodies or the strengthening of existing ones to ensure that compliance with new regulations can be effectively measured is also recommended.

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

Due to the spectrum of capabilities AI can possess, a risk-based approach is the best approach. However, there are a lot of other levers which should be combined with a risk-based approach. These can include:

- Overarching ethical guidelines.
- Sector specific standards
- Technology specific regulation
- Industry body guidelines and controls.

Due to the evolving nature of AI, any risk-based approach needs to be closely monitored. This will help with identifying unanticipated and unintended consequences in a controlled manner.

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

A risk-based approach to AI offers several advantages. By allowing AI to reveal its own risks, it enables us to anticipate and prepare for new issues which inevitably will arise. This will help us stay at the forefront of innovation while being prepared for the unanticipated challenges.

However, a risk-based approach will need a set of underlying standards and guidelines which outlies how to assess risks effectively. Without these, the approach lacks direction. Overcoming the limitations in a risk-based approach requires a better understanding of AI and how it works. This again emphasises the importance of education in this area. Lastly, there may need to be an AI-based compliance technology to monitor the suggested certification and compliance of AI-based products and systems.

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?

Yes, and there are many factors which should be considered. Some critical services which require additional focus would be:

- Critical services will need additional focus:
- Medical Services
- Legal Services
- Banking and Finance
- Construction
- Transport (autonomous vehicles and aviation)
- Defence
- Critical infrastructure (water, power, communications)

An iterative approach to deploying new systems would help to mitigate some of the risks in these sectors.

17. 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 presented in attachment C are supported, however, we note that for “Human in the Loop” we should not discount this as a viable methodology when operating at scale. Oversight by humans can still be effective by 'sampling' the AI output such that the sample is statistically representative of the scaled output.

Any approach needs to be open to change as we learn more about different AI capabilities. These elements should be reviewed and updated on a regular basis.

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?**

Incorporating AI risk-based approaches into existing frameworks can already be achieved in areas such as privacy legislation. Existing regulatory frameworks might be useful for short term needs to regulate AI development, however there is a need for a new certification and compliance regime that combines safety and security needs for a resilient society, while still promoting innovative ways to develop and use AI.

To enhance existing frameworks, risk mitigation should be added. By adding these aspects into the assessment framework, a more comprehensive approach to managing AI risks can be achieved within existing frameworks.

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

Risk based approaches for general purpose AI systems need to focus on 'use of technology' rather than the specifics of the technology. This approach would aim to improve their performance and reliability by addressing the risks associated with the data used to train the algorithm and cybersecurity to protect the system. The approach should involve incorporating real-time data sources and implementing security measures to ensure accurate and secure operations.

To assist in this approach, a new risk-based certification and compliance regime is needed.

20. **Should a risk-based approach for responsible AI be a voluntary or self-regulation tool or be mandated through regulation?**

The approach to regulation in AI should include aspects of both obligated good practice measures and mandated regulations. Regulation should be in place where the technology is considered high-risk and requires an element of public protection (such as the medical sector, transport sector etc.)

And should it apply to:

a) **public or private organisations or both?**

It needs to apply to both public and private organisations. This will balance innovation with the imperative to protect public interests and individual's rights.

b) **developers or deployers or both?**

Primarily it should be focused more on the deployment side, however, developers should be held accountable for the outputs and should be required to have validation and verification (testing) processes associated with AI algorithms. This is the same for any product or service when it is used for its intended purpose.