RMIA
RISK MANAGEMENT INSTITUTE OF AUSTRALASIA

RMIA Submission

Safe Use of AI

Date 26th July 2023

CONTRIBUTORS:

SIMON LEVY – CHIEF EXECUTIVE OFFICE RMIA
GRAEME SAMUEL AC – MONASH BUSINESS SCHOOL
SAMANTHA NEWLANDS – VP RISK – APPEN AI
CLAYTON SCOTT – EAGLE ONE CYBER
JASON SMITH – EXPERIENCED CRO

Risk-Based Approaches Question #1

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

As we navigate the rapidly evolving landscape of technology, we must address the increasing role of Artificial Intelligence (AI) and Generative AI in our society. Rather than attempting to regulate the intricate process, we must adopt a principled approach that focuses on desired outcomes that focus on desired outcomes. AI, including advanced models such as LLM (Large Language Model), Chat GPT, and Bard, has revolutionised how we interact with information and has the potential to bring about both positive and negative impacts. While existing laws often prioritise the principle of "do no harm," it is essential to acknowledge that AI can extend beyond the boundaries of this principle, necessitating the establishment of robust protections for individuals and communities, which is a risk-based approach supported by principles.

As the genesis of AI dates to the visionary work of Alan Turing in the 1930s, its evolution has witnessed tremendous advancements with the advent of powerful computing technologies and the internet. Today, Generative AI represents the forefront of this progress, pushing the boundaries of innovation even further. Recognising that AI, including Generative AI, is an inherent product of technological evolution, we must refrain from singularly focusing on regulating the process itself.

Instead, our approach should encompass a set of principles and guidelines that prioritise desired outcomes while safeguarding against potential harm. This outcome-focused strategy aims to uphold existing laws that protect individuals and communities, emphasising the principles of "do no harm." We must establish and evolve laws to address emerging challenges, such as privacy breaches, intellectual property theft, and defamation in the context of AI usage.

As Generative AI continues to advance and proliferate, the focus must remain on preventing harm and respecting the boundaries of the law. The fundamental objectives should be to avoid breaches of existing legislation and ensure that the principle of "do no harm" remains intact while acknowledging that AI applications can present novel challenges.

By adopting this outcome-focused approach, we aspire to balance embracing the potential benefits of Generative AI and mitigating its potential risks. As technology evolves, so must our regulatory framework, adapting to the changing landscape and AI capabilities. Doing so can pave the way for a safer, more responsible, and equitable integration of Generative AI in our society, fostering innovation while safeguarding the welfare of all individuals and the community.

The Risk Management Institute of Australasia (RMIA) has made several recommendations for addressing the potential risks associated with artificial intelligence (AI) systems. These recommendations are based on the principles of compliance with existing laws, the "do no harm" principle, and protecting the community.

One of the RMIA's key recommendations is adopting a blended approach incorporating both principles and a risk-based AI approach. This means that organisations should assess the potential risks of AI systems and implement measures to mitigate those risks. The RMIA has outlined several factors that should be considered when evaluating the risks of AI, including the purpose and use of the AI system, the data used to train the AI system, and the potential for bias in the AI system. Organisations should assess the potential risks of AI systems and implement measures to mitigate those risks. The RMIA has outlined several factors that should be considered when assessing the risks of AI, including the purpose and use of the AI system, the data used to train the AI system, and the potential for bias in the AI system.

Another key recommendation from the RMIA is to ensure that AI systems are transparent and explainable. This means that users should be able to understand how AI systems work and how they make decisions. This is important for several reasons, including ensuring that AI systems are not used in a discriminatory way and that users can trust the decisions made by AI systems.

The RMIA has also recommended that organisations take steps to mitigate the risk of misinformation and disinformation being spread through AI systems. This could involve using AI to identify and flag potentially harmful content or working with social media platforms to remove this content from their platforms and take steps to mitigate the risk of misinformation and disinformation being spread through AI systems. This could involve using AI to identify and flag potentially harmful content or working with social media platforms to remove this content from their platforms.

Finally, the RMIA has emphasised the importance of balancing the potential risks of AI systems with the need to protect individual rights. This could involve ensuring that AI systems are designed to respect privacy and freedom of expression. This could include providing that AI systems are designed to respect privacy and freedom of expression.

The RMIA's recommendations provide a framework for addressing the potential risks of AI systems responsibly and ethically. This approach should be guided by a set of principles that safeguard against breaching the law and prioritise the principle of "do no harm." When considering the use and deployment of AI technologies, evaluating the potential harm they may cause to society and individuals is crucial. Furthermore, the risk-based approach must account for inherent biases in the data life cycle and consider the implications of AI concerning misinformation, disinformation, and compliance with legislation, such as the racial discrimination act.

Principles of a Risk-Based Approach:

1. Compliance with Legal Frameworks: Any AI system deployed must comply with the law of the Commonwealth. This principle ensures that AI technologies do not violate existing legal statutes and regulations and helps protect individuals and society.
2. Do No Harm: The "do no harm" principle requires that the deployment of AI systems considers potential risks and minimises any adverse impact on society and individuals. Harm encompasses various aspects, including carefully defined social, economic, psychological

(offence harm), and physical harm. This principle ensures that AI technologies are designed and used responsibly (ethical deployment).

3. Protecting the Community: A risk-based approach should aim to prevent harm to the community and its members. This entails identifying potential harm and implementing measures to mitigate those risks. Laws and regulations exist to protect the least sophisticated community members, and AI deployment should not undermine those existing protections.

Considerations for a Risk-Based Approach:

1. Evaluation of Purpose and Use: When assessing AI technologies, it is essential to understand their intended purpose and use thoroughly. This includes identifying potential benefits and risks associated with their application. Understanding the context in which AI systems will be utilised helps shape appropriate risk mitigation strategies.
2. Data Sourcing and Bias: Consideration should be given to the data used to train and operate AI systems. The entire data life cycle, from collection to processing, should be examined for inherent biases that may result in discriminatory outcomes. Robust strategies should be in place to address and mitigate such biases.
3. Mitigating Misinformation and Disinformation: The risk-based approach must include measures to identify and counteract the dissemination of misinformation and disinformation through AI systems. Proactive steps should be taken to minimise the negative impact of AI-generated content on public discourse and ensure compliance with relevant laws, such as 18C of the racial discrimination act.
4. Balancing Risk and Rights: When assessing AI technologies, balancing their potential risks, and protecting individual rights is crucial. The risk-based approach should ensure that significant risks and potential rights violations do not overshadow AI's benefits.

In addition to the RMIA's recommendations, it is also essential to consider the following points when addressing the risks of AI:

- The pace of technological change: AI is a rapidly developing field, and the risks associated with AI systems are constantly evolving. This means it is essential to be vigilant and review risk mitigation strategies regularly.
- The complexity of AI systems: AI systems can be extraordinarily complex, and it can be challenging to understand how they work and identify potential risks. This means having a team of experts who can assess the risks of AI systems and develop appropriate mitigation strategies is crucial.
- The potential for misuse: AI systems can be misused for malicious purposes, such as spreading misinformation or discrimination. This means that it is vital to have strong security measures in place to protect AI systems from being hacked or misused.

When guiding the use of AI, several principles can be considered to ensure responsible and ethical deployment:

1. Fairness and Equity:

Ensure that AI systems are designed and deployed in a manner that treats all individuals fairly and avoids bias or discrimination. Strive to minimise the impact of algorithmic biases and ensure equitable access and outcomes for all users.

2. Transparency and Explainability:

Promote transparency by ensuring that the processes and outcomes of AI systems are explainable and understandable to users and affected individuals. Provide clear explanations for the decisions made by AI systems, enabling accountability and fostering trust.

3. Accountability and Responsibility:

Establish mechanisms to hold individuals and organisations accountable for AI systems' development, deployment, and outcomes. Encourage responsible AI governance, including clear lines of responsibility, appropriate training, and adherence to ethical guidelines.

4. Privacy and Data Protection:

Protect individual privacy and ensure compliance with applicable data protection laws and regulations. Safeguard personal data used by AI systems and provide transparent data handling practices, including obtaining informed consent and anonymising or minimising data where appropriate.

5. Robustness and Reliability:

Develop AI systems that are robust, reliable, and resilient to ensure their consistent performance across different scenarios. Conduct rigorous testing, validation, and risk assessments to identify and mitigate potential failures or biases.

6. Human-Centric Design:

Prioritise the well-being and interests of individuals and society in designing and deploying AI systems. Consider human values, needs, and preferences, and involve diverse perspectives throughout the development process to avoid harmful or dehumanising applications of AI.

7. Societal Impact and Benefit:

Strive to maximise the positive impact of AI on society, including economic growth, improved services, and addressing societal challenges. Conduct thorough assessments of the potential social, economic, and environmental impacts of AI systems, and ensure they contribute to the betterment of society.

8. Ethical Use and Governance:

Uphold ethical standards and principles in the development and use of AI. Foster a culture of responsible AI governance, including regular ethical reviews, stakeholder engagement, and adherence to codes of conduct or industry-specific guidelines.
9. Continuous Learning and Improvement:

Foster a commitment to ongoing learning, improvement, and collaboration in AI. Encourage sharing of best practices, lessons learned, and research findings to address emerging challenges and enhance the responsible use of AI.

It is important to note that these principles serve as a starting point, and their application may vary based on specific contexts, industries, and cultural considerations.

Risk-Based Approaches Question #2

What are the main benefits or limitations of a risk-based approach? How can any limitations be overcome?

A risk-based approach to the safe use of AI is a systematic process for identifying, assessing, and mitigating the risks associated with AI systems. This approach can help organisations to ensure that AI systems are used safely and responsibly.

The benefits of a risk-based approach to AI safety include:

- Identifying and mitigating potential risks: A risk-based approach allows for a systematic assessment of potential risks associated with AI systems. By identifying these risks early on, appropriate measures can be taken to mitigate or manage them effectively.
- Focusing on critical areas: Prioritising risks based on their potential impact helps allocate resources and attention to the most critical areas. This ensures that efforts are concentrated where they are most needed, reducing the likelihood of harmful outcomes. This ensures that actions are concentrated where they are most needed, reducing the likelihood of harmful outcomes.
- Flexibility and adaptability: A risk-based approach provides the flexibility to adapt risk management strategies as added information becomes available or the AI system evolves. It allows for continuous monitoring and adjustment to address emerging risks effectively.
- Encouraging responsible development: A risk-based approach enables developers to incorporate safety considerations into designing and developing AI systems by emphasising risk assessment and management. This promotes responsible and ethical practices.
- Incomplete or incorrect risk assessments: Risk assessments may be limited by the availability of data, biases, or limitations in the assessment methodologies used. This can lead to incomplete or incorrect identification and evaluation of risks associated with AI systems.
- Adherence to the principles of ISO31000 – the standard approach to a risk framework: The absence of a standardised approach for risk assessment and management in AI can lead to inconsistencies and variations in how risks are addressed. This can make it challenging to compare or evaluate the effectiveness of risk management strategies across different AI systems. This can make it challenging to compare or evaluate the effectiveness of risk management strategies across other AI systems.

To overcome these limitations, the following steps can be taken:

- Collaboration and expertise sharing: Stakeholders, including experts from various domains, academia, industry, and government bodies, should collaborate to develop standardised frameworks and share best practices for risk assessment and management in AI.
- Continuous evaluation and improvement: Regular evaluation of risk assessment processes and methodologies is crucial to identify and rectify any shortcomings. Incorporating feedback from users, experts, and the broader AI community can help refine and improve risk assessment practices.
- Accountability and transparency: Establishing clear lines of accountability for risk assessment and management is essential. Both developers and end users should share responsibility for conducting risk assessments and ensuring the safe use of AI systems.
- Independent validation and auditing: Independent third-party validation and auditing of AI systems can help verify algorithms' veracity, integrity, and safety. This can provide an additional layer of assurance and build trust among users and the public.

In summary, a risk-based approach to AI safety offers benefits in identifying and managing potential risks. However, limitations such as incomplete assessments and lack of standardisation must be addressed through collaboration, continuous improvement, accountability, and independent validation to ensure safe and responsible AI systems.

Human intelligence and intervention remain critical for validating algorithms, avoiding undesirable outcomes, and maintaining accountability.

The RMIA recommendations for organisations that are developing or using AI systems include:

- Adopt a risk-based approach to AI safety.
- Collaborate with other stakeholders to develop standardised frameworks and share best practices for risk assessment and management in AI.
- Continuously evaluate and improve your risk assessment processes and methodologies.
- Establish clear lines of accountability for risk assessment and management.
- Consider independent validation and auditing of AI systems to verify algorithms' veracity, integrity, and safety.

By following these recommendations, organisations and individuals can help to ensure that AI systems are used safely and responsibly.

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

The safe use of artificial intelligence (AI) is a complex issue that requires a comprehensive and inclusive approach. The Risk Management Institute of Australasia (RMIA) has made several recommendations for ensuring the safe and responsible use of AI, including:

- Adopting a risk-based approach means assessing the potential risks associated with AI systems and implementing measures to mitigate those risks. The risk-based approach should be tailored to the specific sector or industry and flexible enough to adapt to the rapid pace of technological change.

- Review existing regulatory framework: if an overarching framework is considered, it should be comprehensive and apply to all sectors, regardless of size, AI maturity, or available resources. The regulatory framework should be designed to promote AI's safe and responsible use while ensuring that innovation is not stifled. It is noted that developing a new regulatory framework is not recommended, and Regulations can breed complacency.

- Fostering a risk-aware culture means organisations should create a culture where risk management is essential to business. This includes ensuring that everyone in the organisation, from the CEO (Chief Executive Officer) to the front-line employees, is aware of the risks associated with AI and how to mitigate those risks. This includes ensuring that everyone in the organisation, from the CEO to the front-line employees, is aware of the risks associated with AI and how to mitigate those risks.

- Support mechanisms: Organisations should provide employees with the resources and training to effectively understand, assess, and manage AI-related risks. This includes providing access to training materials, tools, and experts. To effectively understand, evaluate, and manage AI-related risks. This includes providing access to training materials, tools, and experts.

The RMIA also emphasises addressing cross-sector concerns, such as modern slavery, as an example of why additional regulation is not required.es addressing cross-sector concerns, such as modern slavery as an example of why regulation is not required.

In addition to the recommendations made by the RMIA, it is also essential to consider the following points when ensuring the safe use of AI:

- The pace of technological change: AI is a rapidly developing field, and the risks associated with AI systems are constantly evolving. This means it is essential to be vigilant and review risk mitigation strategies regularly.

- The complexity of AI systems: AI systems can be overly complex, and it can be challenging to understand how they work and identify potential risks. This means having a team of experts who can assess the risks of AI systems and develop appropriate mitigation strategies is crucial.
- The potential for misuse: AI systems can be misused for malicious purposes, such as spreading misinformation or discrimination. This means that it is vital to have strong security measures in place to protect AI systems from being hacked or misused.

By considering the RMIA's recommendations and the other points outlined above, organisations can help to ensure that AI is used safely and responsibly.

The following are some additional points that could be included in a more professional discussion of the safe use of AI:

- The importance of transparency and accountability: Organisations should be transparent about how they use AI and accountable for the decisions that AI systems make.
- The need for ethical guidelines: There is a need for ethical policies to govern the development and use of AI. These guidelines should be based on fairness, non-discrimination, and privacy. These guidelines should be based on justice, non-discrimination, and privacy.
- The importance of public engagement: The public should be engaged in developing and using AI. This will help ensure that AI systems are designed and used in a way that benefits society.

Risk-Based Approaches Question #4

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

A risk-based approach is essential for managing the potential risks associated with AI systems. This approach is essential for ensuring that AI is developed and deployed safely and responsibly.

The elements outlined in Attachment C—impact assessment, notices, human-in-the-loop/oversight assessments, explanations, training, and monitoring and documentation- provide a comprehensive and proportional approach to addressing potential AI risks. They cover crucial aspects from the initial stages of AI system development through deployment and ongoing monitoring. By incorporating these elements into AI governance frameworks and policies, you can ensure that the risks associated with AI are adequately assessed, managed, and mitigated.

However, it is important to note that the risk landscape surrounding AI is continuously evolving. Therefore, it is recommended to periodically review and update these elements to align with the latest advancements, emerging risks, and regulatory frameworks.

In conclusion, a risk-based approach and the elements outlined in Attachment C provide a solid foundation for addressing potential AI risks. By adopting such an approach, organisations and individuals can promote AI's safe and responsible use while maximising its benefits.

The risk-based approach to AI is a dynamic and evolving framework. As AI technology continues to develop, new risks will emerge. It is, therefore, necessary to periodically review and update the elements of Attachment C to ensure that it remains effective.

Table 1: Blended Risk-Based and Principles-Based Approach to Managing Potential AI Risks in the Australian Industry

Elements	Description
Risk Assessment	Conduct thorough risk assessments to identify potential risks associated with AI applications.
Ethical Principles	Establish a set of overarching ethical principles that guide the development and deployment of AI systems- such as "do no harm".
Risk Prioritisation	Prioritise identified risks based on their severity and potential impact on individuals, society, and businesses.
Risk Mitigation Measures	Implement specific measures to mitigate identified risks, such as bias detection and mitigation algorithms.
Continuous Monitoring	Continuously monitor AI systems to detect and address emerging risks and ensure ongoing compliance with principles.
Compliance Mechanisms	Develop mechanisms to ensure organisations comply with ethical principles and risk mitigation requirements.
Industry Collaboration	Foster collaboration among industry stakeholders, academia, and regulatory bodies to share best practices and insights.
Regulatory Guidelines	Develop guidelines and standards that outline the expected ethical conduct and risk management practices.
Research and Education	Invest in research and education to enhance understanding of AI risks and promote responsible AI practices.
Accountability	Establish mechanisms for accountability, including reporting requirements, audits, and consequences for non-compliance.

Risk-Based Approaches Question #5

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 an AI risk-based approach into existing assessment frameworks and risk management processes can help streamline and reduce potential duplication. Here's how various aspects can be addressed:

1. Parties Developing, Building, or Adding Domain Knowledge:

When developing or building AI systems, the involved parties must comprehensively understand the domain or industry they are operating. This knowledge should encompass the technical aspects of AI and the specific risks and considerations relevant to the industry. Collaborative efforts between AI developers and domain experts can ensure that AI systems are built with a deep understanding of the risks and potential impacts on the industry.

2. Companies Buying and Plugging AI Solutions:

Organisations that purchase AI solutions from third-party vendors must assess the associated risks, thereby taking responsibility for their supply chain. This includes evaluating the reliability and integrity of the purchased AI systems and their potential impact on their existing risk management processes. Due diligence should be performed to ensure that the AI systems align with their risk management objectives and comply with relevant regulations and privacy requirements.

3. Data Governance and Techniques:

Effective data governance is crucial for managing AI risks. Organisations must establish robust data governance frameworks covering data quality, security, privacy, and ethical considerations. This involves defining data governance policies, ensuring compliance with relevant regulations, implementing appropriate data protection measures, and continuously monitoring and auditing data usage. Data anonymisation, differential privacy, and secure data-sharing protocols can help mitigate privacy risks associated with AI systems.

4. AI Data Cycle:

The AI data cycle involves data collection, pre-processing, training, validation, deployment, and monitoring. Each stage presents risks that need to be addressed within existing risk management frameworks. These risks include biased or discriminatory data, data breaches, algorithmic vulnerabilities, model drift, and unintended consequences. Integrating AI-specific risk assessments and controls at each data cycle stage can help proactively identify and mitigate potential risks.

5. Incorporation into Existing Frameworks:

To ensure the incorporation of AI risks into existing frameworks, it is essential to have a holistic and 360-degree view of risks. This includes assessing financial and non-financial risks associated with human-led and AI-driven activities. Risk management frameworks should be expanded to accommodate AI-specific risks, such as algorithmic biases, cybersecurity vulnerabilities, unintended consequences, and ethical concerns. This integration ensures that AI risks are treated as an integral part of the overall risk landscape.

6. Competency of Risk Management Professionals:
Risk management professionals, including Chief Risk Officers (CROs), must be well-versed in the potential risks and challenges AI systems pose. They should possess the necessary knowledge and skills to evaluate and address AI-related risks. Regular training and education programs should be provided to enhance their competency in managing AI risks. Collaboration between risk management professionals and AI experts can also help bridge the gap between technical understanding and risk management expertise.

7. Privacy and Responsibility by Design:
Privacy considerations should be integrated into the design and development of AI systems from the outset. Privacy by design principles, such as data minimisation, purpose limitation, and user consent, should be followed to mitigate privacy risks. Responsible AI practices should also be incorporated into AI development processes, including transparency, explainability, fairness, and accountability. This ensures that AI systems are built with privacy and ethical considerations as core components.

8. Likelihood Changes and Regulating the Process:
Adopting AI may lead to changes in the likelihood of specific risks. As AI technologies evolve, regulations should be dynamic and adaptable to address emerging risks effectively. However, there is a risk of complacency when relying solely on regulations. Organisations and regulatory bodies should encourage continuous risk assessment and vigilance to adapt to evolving AI risks. Regular evaluations and updates of risk management processes are necessary to stay ahead of potential threats.

In addition to the RMIA's recommendations, there are several other considerations that organisations should keep in mind when incorporating AI into their operations. These include other considerations organisations should consider when incorporating AI into their operations. These include:

- The use of AI should be aligned with the organisation's overall risk appetite. Organisations should carefully consider their tolerance for risk before adopting AI and ensure that the benefits of using AI outweigh the risks of organisations.
- The organisation should clearly understand the potential benefits and risks of using AI. This includes understanding the specific risks associated with the AI systems the organisation is considering using.
- The organisation should have a plan for managing the risks associated with using AI. This plan should include steps to mitigate the risks and procedures for responding to incidents. The risks associated with using AI. This plan should include measures to mitigate the risks and procedures for responding to incidents.
- The organisation should regularly review and update its risk management plan as AI technologies evolve. The risks associated with AI are constantly evolving, so organisations need to review their risk management plans regularly to ensure they are still effective. Organisations can help ensure they use AI safely and responsibly by taking these steps.

Here are some additional details about each of these considerations:

- Alignment with risk appetite: Organisations should carefully consider their tolerance for risk before adopting AI and ensure that the benefits of using AI outweigh the risks. For example, an organisation that is highly risk-averse may want to adopt AI systems that are less complex and have a lower risk of failure. For example, an organisation that is highly risk-averse may want to adopt AI systems that are less complex and have a lower risk of failure.
- Understanding of the benefits and risks: Organisations should clearly understand the potential benefits and risks of using AI. This includes understanding the specific risks associated with the AI systems the organisation is considering using. For example, some risks associated with AI systems include bias, discrimination, data breaches, and unintended consequences.
- Plan for managing risks: Organisations should have a plan for managing the risks associated with using AI. This plan should include steps to mitigate the risks and procedures for responding to incidents. For example, organisations can mitigate the risk of bias in AI systems by using techniques such as data anonymisation and differential privacy and procedures for responding to incidents. For example, organisations can mitigate the risk of bias in AI systems by using techniques such as data anonymisation and differential privacy.
- Regular review and update of risk management plan: The risks associated with AI are constantly evolving, so organisations must review their risk management plans regularly to ensure they are still effective. For example, organisations may need to update their plans to address new risks or to reflect changes in their risk appetite.

In summary, incorporating an AI risk-based approach into existing assessment frameworks and risk management processes requires comprehensive domain knowledge, diligent evaluation of third-party AI solutions, robust data governance practices, addressing risks across the AI data cycle, expanding existing frameworks to cover AI risks, enhancing the competency of risk management professionals, integrating privacy and responsibility by design, and continuously evaluating and adapting to changing risk landscapes.

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

A risk-based approach can be applied to general-purpose AI systems, including large language models (LLMs) and multimodal foundation models (MFMs), to ensure their safe and responsible use. Here are some key considerations regarding the application of a risk-based approach to these models:

1. Different uses of LLMs and MFMs: LLMs and MFMs serve as foundational tools that enable various applications and solutions. LLMs, such as GPT-3, are primarily used for generating text based on input prompts. At the same time, MFMs combine multiple modalities like text, images, and audio to perform tasks like image captioning or speech recognition.

2. Using LLMs (Large Language Models) as a base: Organisations often build upon LLMs or MFMs (multimodal foundation models) to create institutional models tailored to their specific needs. This approach allows customisation while leveraging the underlying capabilities of the LLM or MFM (multimodal foundation models).

3. Risk assessment at the development stage: Conducting a comprehensive risk assessment specifically for the LLM itself may not be feasible due to its general-purpose nature and wide range of potential applications. However, developers should consider potential risks associated with using LLMs, such as biases in training data, privacy concerns, or security vulnerabilities.

4. Questions about transparency and IP (Intellectual Property): Users of LLMs should engage with developers to seek transparency regarding the model's limitations, potential biases, and any known risks. Intellectual property (IP) considerations are also important, as users should understand the ownership and licensing terms associated with the LLM they are utilising.

5. Awareness of LLM changes: Users should be aware of any updates or changes made to the LLM by the provider. Changes in the underlying model, training data, or data sources can impact the output and performance of the LLM, which may have implications for the users' specific applications.

6. Supplier dependencies: Users must understand if their LLM provider relies heavily on a single data set provider. Overreliance on a specific data source may introduce risks such as biased or incomplete information, potentially impacting the outcomes of the LLM's performance.

7. Adapting to advancing technology: As AI technology advances, users must be adaptable and keep updated with the evolving landscape. Knowledge of the potential risks and impacts of using LLMs and MFM enables users to make informed decisions and mitigate potential harms.

8. Becoming a jack of all trades: Given the rapid advancement of technology, individuals and organisations should strive to become knowledgeable in multiple domains and understand the potential risk outcomes associated with AI applications. This multifaceted understanding allows for more comprehensive risk management and responsible decision-making.

In summary, responsible use of AI involves considering the risks associated with LLMs and MFMs, engaging with developers to address transparency and IP concerns, staying informed about updates and changes in the models, being aware of supplier dependencies, and maintaining a proactive approach to adapt to evolving technology. By establishing standards and encouraging users to ask relevant questions, the government can foster the responsible and safe use of AI technologies.

Risk-Based Approaches Question #7

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?**

Regarding voluntary versus mandated regulation, the RMIA recommends that a risk-based approach for responsible AI be mandated through law rather than relying solely on voluntary or self-regulation. While voluntary standards and codes of practice have their place in promoting best practices and encouraging responsible behaviour

Existing regulatory bodies such as the Australian Federal Police, Crime Commission, APRA (Australian Prudential Regulation Authority), ASIC (Australian Securities & Investments Commission), and CASA (Civil Aviation Safety Authority) already deal with breaches of the law and oversee compliance and enforcement within their respective domains. However, considering AI's unique challenges and opportunities, these regulatory bodies should conduct a comprehensive gap analysis to identify areas where existing regulations may fall short in addressing AI-specific risks. If such gaps are identified, they should be rectified through appropriate regulatory measures.

Both public and private organisations should be subject to this regulatory framework.

Regarding the roles of developers and deployers, the regulatory framework should apply to both parties. Developers are responsible for creating AI systems, and they play a crucial role in ensuring the design and development processes adhere to ethical principles and mitigate potential risks. Deployers, conversely, are responsible for implementing and operating AI systems. They need to ensure the appropriate use of AI, including considering potential biases, transparency, and accountability.

A risk-based approach should also encompass continuous review and adaptation. As AI technology evolves rapidly, it is crucial to have an ongoing process of evaluating and updating existing regulations to keep pace with emerging risks and challenges. This process should involve collaboration with experts, stakeholders, and the wider community to ensure that regulations remain effective and responsive to the dynamic AI landscape.

In conclusion, while voluntary standards and codes of practice, such as ISO (International Standards Organisation) 27701 for privacy, have their place in promoting responsible AI, a risk-based approach to responsible AI should be mandated via existing regulations that should apply to both public and private organisations and developers and deployers.

By establishing clear guidelines, accountability mechanisms, and enforcement measures, we can mitigate potential risks, ensure responsible AI practices, and protect the interests of individuals and society.