Royal Australasian College of Medical Administrators (RACMA) Response to the Australian Government's "Safe and Responsible AI in Australia" Discussion paper

August 2023

CONTENTS

About RACMA

The Royal Australasian College of Medical Administrators (RACMA) is unique as a provider of medical management and leadership qualifications as it is the only specialist medical educator whose programs are recognised for the granting of Specialist Registration in Medical Administration. The Fellowship Training Program offered by RACMA is accredited by the Australian Medical Council and the New Zealand Medical Council. Fellows of our College are recognised as medical specialists.

Our education programs are aimed at equipping doctors with the leadership and management skills needed to influence and lead the Australasian health care systems with the clear aim of improving health outcomes of Australians and New Zealanders.

The value of RACMA is its Members, who strive to lead for change and positive outcomes for all Australians, New Zealanders, and peoples in all parts of the world in which their Members practice, demonstrating their skills in key areas such as system leadership, clinical governance, financial management, workforce management, and professional leadership. RACMA Members fill key roles in all aspects of health, including government, the public. private and not-for-profit sectors. Their roles are diverse, including Chief Executives, Chief Medical Officers, Heads of Departments, Clinical Informaticians and Academics informing, influencing and making key decisions about system clinical governance and quality, including clinical informatics.

Informed Decision-making in Healthcare – Key Principles

A key focus of RACMA's response is enabling healthcare leaders to develop governance structures and processes to inform decisions that manage the risk of AI technology in healthcare settings. This ensures its use results in achieving one or more of the four aims of healthcare being improved population health, patient experience, healthcare provider wellbeing, and cost efficiency, without unacceptable trade-offs.

Australia's AI Ethics Principles

RACMA supports the Australian Government's AI ethics principles in (Ref: Box 3: Australia's AI Ethics Principles" of the Discussion Paper) and those of the World Health Organisation "Guidance on Ethics and Governance of Artificial Intelligence for Health":

1. Protecting human autonomy
2. Promoting human well-being and safety
3. Ensuring transparency, explainability and intelligibility
4. Fostering responsibility and accountability
5. Ensuring inclusiveness and equity
6. Promoting privacy and data governance

(Ref: https://www.who.int/publications/i/item/9789240029200)

In the healthcare context, the practical implementation of these principles in a real-time, time-sensitive healthcare environment, and one that supports patient-centred, team-based, and evidence-informed care, also need to be considered.

A number of key principles, many of which are contained in the discussion paper, inform RACMA's response regarding the use of AI in healthcare. These include:

1. Recognition that AI development is occurring in a relatively unregulated environment, in both the public and private sectors, such that even if the development environment could be regulated, the use and impact of AI technologies cannot necessarily be predicted or controlled by the developer.
2. The development and application of AI technologies exists in a global context, with the development and use extending beyond national boundaries. As such, where possible, Australia should endorse, align with, and promote the highest regulatory standards developed by the international community and global institutions including the WHO, European Union, and countries noted in the discussion paper, that are centred around the public interest to protect and enhance the wellbeing of humanity, society, and the planet. The aim of this is to hold AI developers to agreed global standards, if they are seeking approval for use of their technology in Australia, while also remaining consistent with regulations of different countries. Considering the

impact of federalism on governance in the Australian healthcare system and to ensure consistent standards can be applied and complied with, a national approach, rather than a state and territory-based approach, should be sought.

3. Significant work such as the Australian Alliance for Artificial Intelligence in Healthcare' Roadmap for Artificial Intelligence in Healthcare for Australia is noted. Such guidance needs to be implemented through effective regulatory and governance processes, in consultation with expert stakeholder groups, and training of the health workforce to ensure appropriate use.
4. The use of risk management-based approaches to the use of AI should be supported, recognising the rapid rate of growth in the development of AI techniques will produce known, known unknown and unknown unknown risks and outcomes. As such, existing risk management frameworks should be reviewed with a single framework identified and endorsed. The effective application of such frameworks in exemplar high-risk industries e.g. aviation, should be considered and learned from.
5. In the development
6. Where there is the potential to regulate the use of AI, this is contingent on:
    a. identification of products that use AI technology.
    b. regulatory classification of the risk associated with the product; and
    c. decision makers, regulators and users understanding the appropriate use and impact of AI on the output of such products.
7. In healthcare,
    a. Any AI technology needs to consider and demonstrate benefit for the four aims of health care - improved population health, patient experience, provider wellbeing, and reduced costs, without unacceptable trade-offs.
    b. AI-based healthcare in the form of software as a device must adhere to regulatory requirements (https://www.tga.gov.au/how-we-regulate/manufacturing/medical-devices/manufacturer-guidance-specific-types-medical-devices/regulation-software-based-medical-devices#artificial-intelligence-chat-text-and-language)
    Given the rapid evolution of AI technology, regulatory agencies such as the Therapeutic Goods Administration should continue to engage with the AI stakeholder community and be resourced to maintain currency and relevance of regulatory processes, and ensure they have the expertise and capacity to assess, classify, monitor, and effectively regulate medical devices, software and technology to enable the safe and responsible use of AI.
    c. In healthcare settings, institutional governance committees and processes overseeing procurement of new technologies should be used to source any AI related products, systems, applications or add-in modules to existing products or systems, and such committees must include membership with clinical and technology experts in healthcare applications of AI. Such committees should consider the complete technology lifecycle of inception, development, deployment, maintenance and decommissioning in their assessments (Ref: American Medical Informatics Association Position Paper: Defining AMIA's artificial intelligence principles), as well as business continuity plans to ensure continuity of healthcare provision, when AI-based technologies fail or are deemed not fit-for-purpose.

d. Research and use of AI technologies involving the use of participant, patient or health system data, needs to be predicated on key data management principles such as the Australian Privacy Principles, and the "Five Safes" data governance and risk management framework, considering the domains of Safe people; Safe projects; Safe settings; Safe outputs; Safe data. (https://www.aihw.gov.au/about-our-data/data-governance/the-five-safes-framework)
e. "Human in the loop" approaches that ensure AI systems are "guided, communicated, and supDervised by human expertise," are critical for maintaining safety and quality in healthcare services (Ref: Sezgin E. Artificial intelligence in healthcare: Complementing, not replacing, doctors and healthcare providers. Digital Health. 2023 Jul;9:20552076231186520.)
f. For any AI-based health information system, there must be clear labelling, including instructions for use that detail:
    i. the intended use of the AI technologies e.g. developed for consumer versus healthcare professional users.
    ii. a description of how AI is used in the respective device or application particularly with respect to contribution to decision making e.g., clinical-decision support, prediction, and/or final decision-making in health care.
    iii. the specific healthcare context and cohorts in which the application has been validated.

8. When an AI-based system is used, the output should provide information to the healthcare provider as to whether the use of that system has been validated in the healthcare context and/or patient cohort, and cite supporting evidence.
9. Education about the risks and issues and benefits of AI technology should be made a prioritised focus of initiatives to increase digital health workforce literacy and capacity, and that this education include healthcare application developers, regulators, decision makers and users. This includes education of professional services review bodies e.g., AHPRA, Coroners etc, who review adverse outcomes and causative agents, so that their knowledge is current, and to ensure vendors are also held accountable for the impact of their technologies on healthcare provision.

QUESTIONS TO ADDRESS

Definitions

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

Response:
In general, we support the use of standardised terms that are defined by recognised authoritative sources such as those used in the discussion paper. For the drafting of any health-care specific regulations, policies, and procedures, defining the application of AI in healthcare, a similar approach should be taken. Given the general use of the term “Artificial Intelligence”, clear and current definitions of the type of AI are also critical to ensure the accurate interpretation and application of these definitions are used in regulatory and governance processes, particularly with respect to risk.

Potential Gaps in Approaches

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?

Response:
As stated in our key principles, there needs to be recognition that AI development is occurring in a relatively unregulated environment, or at a rate greater than regulation can be developed, in the public, not-for-profit and private sectors. Even if the development environment could be regulated, the use and impact of AI technologies cannot necessarily be predicted or controlled by the developer. As such a principles-based risk management framework, that can be readily applied, rather than specific legislation over specific technologies or uses, may be more likely to be effective. Collaborative risk-based frameworks for assessing AI risks for individuals, organisations and society, such as those noted in the discussion document, including the US National Institute of Standards and Technology AI Risk Management Framework, and that have been adopted by several countries worldwide and could be readily applied in Australia. This is particularly important in high-risk environments such as healthcare where a first do no harm approach needs to be assured. Just like in other areas of healthcare, the primacy of ethical principles including beneficence, non-maleficence, autonomy and justice, is critical particularly where other considerations such as cost-reduction and profitability might be prioritised benefits of implementing AI-based technology. As such, the importance of context-specific risk assessment is critical. Criteria around ‘risk level’ would need stricter with more transparent parameters. In healthcare, higher-risk settings for AI should applied, especially with algorithm-assisted decision making. For example, the recent media coverage of AI-based products to generate patient records in healthcare settings has been classified as ‘Medium risk’ by developers, despite the fact that this could lead to safety and quality risks and adverse events. It is also important to note that triggering the use of such risk assessment frameworks is firstly contingent on transparent disclosure and labelling of the use of AI in a product.

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.

Response:
Significant work has been already carried out by groups such as the Australian Alliance for Artificial Intelligence in Healthcare, however they have not been effectively implemented or communicated to stakeholders including the application developers, healthcare organisations, providers and consumers. The Government should leverage existing national and international resources, avoid duplication of work, and focus on disseminating endorsed authoritative information to key stakeholders, so that it can be applied in practice.

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.

Response:
Existing authoritative national organisations that have collaborative arrangements internationally, and in the public and private sectors, should be invested in and inform government policy. The aim should be to standardise AI governance, strengthen and reinforce the role of global institutions to regulate AI development internationally, reduce complexity, and support compliance, enabling safe implementation in Australia. As stated in the key principles, a national approach, rather than individual state and territory-based approaches, would support standardised governance, implementation and compliance across the Australian healthcare system.

Responses Suitable for Australia

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?

Response:
The implementation and application of a risk management framework to all AI applications is a matter of priority to support the identification, classification and labelling of any medical advice and software that use AI and the output of such systems to inform healthcare providers and consumers decision making. In this regard, we support legislative approaches such as those taken by the European Union through the Artificial Intelligence Act.

Target Areas

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

Response:
In accordance with the WHO AI in healthcare ethical principle of ensuring inclusiveness and equity, regulatory and governance approaches that 'facilitate emerging technologies rather than hinder innovation' should be applied equitably across all sectors, or there is a risk the public and not-for-profit sectors will face greater hindrance and much less agility, with private sector players greatly outpacing in innovation. Such an equitable approach may benefit society, given safe and successful implementation of AI in the public sector could be considered a public good. Support for application of AI governance principles including risk assessment and standardised assessment tools should be included as part of technology programmes across the public sector, including introduction of new technology and improvements to existing technology. While there may be challenges to applying a common approach to private organisations operating beyond national borders, we would still advocate for such an approach.

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

Response:
Standardisation of AI governance principles including those specific to healthcare, across all levels of government and, public, private and not-for-profit organisations, similar to existing effective models of ethics governance e.g. NHRMC governance of ethics in human research and supporting educational initiatives for the implementation and management of such processes, may be effective.

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.

Response:
Given the unknown potential risks, a risk management framework, such as those identified in the discussion paper and as proposed above, and taking a likelihood and consequence approach is probably the most applicable to managing the risks of AI in healthcare.

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?

Response:
RACMA support approaches informed by the Australian Alliance for Artificial Intelligence in Healthcare' Roadmap for Artificial Intelligence in Healthcare for Australia, and the American Medical Informatics Association Position Paper defining artificial intelligence principles. Of key importance for transparency is the need for plain-language, concise and clear labelling where AI is in use and available at the time of use of the technology. For example, if chatbots or automated response systems are in use, this could be declared at the time of first interaction with the system.

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

Response:
Through existing authoritative bodies that collaborate with the private, public, and not-for-profit sectors, nationally and internationally, labelling standards needs to be agreed upon for when AI is in use.

Do you have suggestions for:

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

Response:
Any healthcare AI application that contravenes the WHO ethics guidelines, specifically those adversely affecting autonomy, well-being and safety, inclusiveness and equity, privacy and data governance particularly if using data collected for healthcare but used for other purposes. requires assessment of whether it constitutes an unacceptable risk and consideration of bans as per the EU Artificial Intelligence Act. Such assessments require an understanding of the impact of the type of Artificial Intelligence used given the diversity, intended and potential uses.

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

Response:
In general, automated decision-making applications used in high-risk industries or contexts e.g., healthcare, transportation, essential services, and involved in real-time decision-making affecting the health and wellbeing of people, are key criteria. Criteria are likely to be context specific and there is a role for ongoing authoritative and regulatory bodies to have processes for the continual review of such criteria based on evolution of AI technology and recognition of evolving risks. To enable the consideration and application of restrictions on development or use, including bans, a high priority needs to be placed on regulations, policies, and procedures for identifying, reviewing, and classifying high-risk AI, using such frameworks as being developed by the EU, such that a first do no harm approach can be taken to approvals.

In healthcare, applications that enable autonomous decision-making without human input in high-risk settings, particularly in real-time, may also constitute an unacceptable risk warranting restriction of use. Risk assessment frameworks need to clearly define assessment of threshold risks, consider uncertainties about the impact when such technologies become embedded, and how restrictions could be enforced when there may be a lack of transparency about the transition point between human and AI-based decision making. This illustrates the importance of understanding the role of the “Human in the loop” in AI systems for maintaining safety and quality of AI-enabled healthcare.

With regard to responsible AI being a voluntary/self-regulation or be mandated through regulation, due to the significant risks involved, a stricter cautionary mandated approach should be considered in healthcare. This may also provide a pathway for greater public acceptance and trust in AI going forward.

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

Response:
As per the American Medical Informatics Association (AMIA) position paper, the two aspects to trust in the case of AI are:

- the organisation deploying and operating the AI must be transparent, responsible, and accountable, and
- the AI system itself and its data and output must be verifiable.

This implies several principles for organisation (Benevolence, Transparency, and Accountability) and for the AI (Explainability, Interpretability, Fairness, Dependability, and Auditability).

Two critical actions are clear labelling of technologies using AI, and education initiatives to inform healthcare providers and consumers about when and how AI is in use. These will support autonomy of individuals to make a choice about how such technology is influencing their decisions.

Practically, improving public trust and confidence in an AI-based 'black box' will require concerns about equity, ethics, governance, accountability, and the future-of-work, as well as human-centric approaches to be addressed as part of AI implementations.

While the public will embrace and trust what they perceive as very 'low-risk' AI, as we move across that continuum, the many issues around AI regarding bias, sources of information, trustworthiness, consent, values, transparency will need a system of trust to be built. A clear pathway will also be required for AI literacy, education and advocacy in AI technology and providers of those technologies.

Only a consultative, holistic, collaborative, transparent approach that takes into consideration all stakeholders including users, providers, developers, governing bodies could lead to public acceptance and trust.

Implications and Infrastructure

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?

Response:
There is a difference between the capability of a technology e.g., facial recognition, and the application of a technology e.g., social scoring. Once a technology capability has been achieved, it is likely to be very difficult to control its use, hence the role of approaches such as export restrictions. However, in a globalised environment, with technology extending beyond national boundaries, it is unlikely that Australia banning certain technologies is going to meaningfully impact on relationships with other countries. Australia's support for global initiatives to limit certain applications, through international collaboration and leadership, may have greater importance if Australia wants to restrict the introduction of certain technologies to Australia that could have adverse impacts. It should be noted it is important to regularly monitor this approach as the technology evolves.

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

Response:
A national approach to AI in healthcare would help reduce duplication and resourcing, and support healthcare leaders to implement standardised governance standards and processes in Australian public, private and not-for-profit healthcare institutions.

Risk-based Approaches

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

Response:
Given the rate of growth and application of AI technology, a risk management framework applicable to the healthcare context, that considers the consequences and likelihood of adverse impacts, is most likely to assist in the identification, classification, and governance of AI technologies in healthcare.

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

Response:
An internationally aligned risk management framework based on agreed principles is most likely to be applicable across the range of applications and can evolve with new and emerging risks. This will require education about the technology for those applying frameworks. There will also need to be regular ongoing review of such frameworks, to ensure governance evolves with AI technology and risks to maintain relevance and currency.

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

Response:
Given AI in healthcare should be considered an intervention, a risk-based approach, often used in healthcare is applicable, is well understood and appropriate. It may a be useful approach for educating healthcare providers and leaders, that is applicable across the range of organisation sizes from a GP practice to a healthcare network, the range of consumer to medical grade AI-based systems that are continually evolving, and the increasing resources invested in digital transformation of health.

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

Response:
The elements of impact assessment, notices, role of humans in oversight, explanations, training, monitoring and documentation, are all important from a governance perspective, and relevant as a general approach, but may be difficult to apply in complex AI applications and may not be sufficiently context specific. In healthcare, decision making is complex with potential significant impact on health outcomes that may be uncertain. However, AI systems can create an unjustifiable sense of certainty without understanding the factors upon which a result, guidance or decision is based. Human factors phenomena such as automation bias, the over-reliance on automated aids and decision support systems that mistakenly overrides correct decisions, may result in unintended harm. As such risk-based approaches considering implementation and risk mitigation strategies need to be considered according to the respective context.

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?

Response:
For healthcare applications, we support the inclusion of AI-specific risk assessment frameworks in existing established healthcare privacy and governance frameworks identified in the discussion paper, particularly where it applies to use of patient, research participant and/or provider data. In defining how AI healthcare applications use and present data, information and results, principles such as “Meaningful Use” could also be utilised.
To ensure healthcare providers and consumers are informed when AI technologies are being used and how these affect the information being provided, principles outlined by the European Union and the American Medical Informatics Association should be adapted to the Australian health system at the outset and as a priority.

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

Response:
Healthcare and decision making is heavily dependent on the understanding and interpretation of information, including sources of data, and is highly contextual. As such, the same risk-based approach should be applied to LLM and MFM AI-based technologies as for other applications used in healthcare. Training and education are key elements, such that healthcare providers and consumers can understand the sources of information that LLMs and MFM are basing their output on, with the risk-based approach needing to incorporate the transparency with which AI-based systems display this source information.

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?

Response:
In public, private, and not-for-profit healthcare organisations where AI-based technologies are being used in healthcare provision and operational support, mandatory risk-based approaches should be used for procurement and implementation processes.

b) developers or deployers or both?

Response:
Developers and deployers of AI-based healthcare applications need to have considered how applications could potentially be used and be clear in the intended use of the application and take into consideration the same risk-based approach taken by healthcare providers and consumers using that technology. As such there needs to be a standardised, nationally endorsed risk management approach, aligned with international risk management frameworks.