Response to The Australian Government Department of Industry, Science and Resources: ‘Safe and Responsible AI in Australia’ Discussion Paper

Digital Health CRC (DHCRC) Ltd

The Digital Health CRC Limited connects government, academia and industry to accelerate the implementation and translation of evidence-based digital health technologies that solve the most pressing healthcare challenges. We are co-funded through the Commonwealth Government’s Cooperative Research Centres (CRC) Program, and by our Participant organisations.

Research Australia, the national alliance for the health and medical research and innovation sector, recognises the expertise of the Digital Health CRC in relation to AI in healthcare and is pleased to support this submission.

Definitions

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

Yes. The DHCRC agrees with definitions used in the paper based on the ISO definitions.

By way of a comment, at the internationally acclaimed Medical and Health Informatics congress, known as MedInfo held in Sydney from 8 to 12 July, many of the presentations involving application of AI in healthcare were highly experimental, involving significant use of ‘synthetic data’. The ISO detailed definitions of AI referenced in the Discussion paper do not define the term synthetic data, yet as the European Data Protection Supervisor notes ”Synthetic data is gaining traction within the machine learning domain.”

https://edps.europa.eu/presspublications/publications/techsonar/synthetic-data_en

Given the legislative, cultural, and technical challenges in timely access to big data sets in Australia, we can expect the use of synthetic data to be a feature of AI development for some time. The GDPR recommend that a privacy assurance process should be applied to synthetic data sets to ensure that data subjects can’t in fact be identified in a synthetic data set. We suggest consideration should be given to including a definition for synthetic data on the basis that many discussions around governance of AI can expect to reference the term.

The research firm Gartner estimates that, by 2030, synthetic data will overtake actual data in training AI models. IBM Blog https://research.ibm.com/blog/what-is-synthetic-data accessed 17 July 2023.

Definitions include:

European Data Protection Supervisor (Tech Champion Robert Riemann) “Synthetic data is artificial data that is generated from original data and a model that is trained to reproduce the characteristics and structure of the original data.”

OECD: “Usually generated by computer simulations, including data collected through reinforcement learning. Synthetic data allow for simulation of scenarios that are difficult to observe or replicate in real life (e.g., a car accident) or are otherwise too expensive to collect at scale (e.g., millions of miles of driving time for self-driving cars). They include most applications of physical modelling, such as music synthesisers or flight simulators.” OECD Framework for the Classification of AI Systems February 2022.

Potential gaps in approaches

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

Whilst there are valuable initiatives relevant to AI listed in Attachment A, the protracted time frames involved and the focus on Australian Government (vs States and Territories and Private Sector) initiatives mean there are a range of potential risks that may go unattended, particularly in health and social care.

The DHCRC supports the 2021 Australian Alliance for Artificial Intelligence in Healthcare (AAAiH) roadmap for AI call for the development of a National AI in Healthcare Strategy to support and encourage collaboration and strategic leadership.

A healthcare specific AI strategy would provide a cohesive approach to the design, development, testing and delivery of AI whilst ensuring alignment with many of the existing initiatives listed. For example:

- The ACCC review referenced in Attachment A is focused on large platforms. Many apps in use and/or available to healthcare consumers and providers have been developed by developers, often inhouse or in partnership with local universities, on smaller platforms. We assert such development should be in scope for the Digital Platform Services Inquiry to ensure that any new regulatory regimes applying to the larger platforms does not send the wrong message to an AI industry looking to supply digital platform services into healthcare by way of smaller platforms that have not been subjected to appropriate controls or supports.
- Similarly, if not already in scope, the joint ACCC, ACMA, eSafety Commissioner and OAIC Digital Platform Regulators Forum should have risks from use of AI in healthcare in scope and for that scope to be extended to include other levels of government. A national AI in healthcare strategy could provide the vehicle for extrapolating the findings from the Regulators Forum so they can be applied to health specific initiatives in a timely manner, given timelines for addressing potential risks from AI need to be short. For example, to inform effective governance of use of generative AI/LLMs in healthcare the views and concerns of these respective agencies may be important to healthcare developers and procurers of AI.
- The roadmap for AI spells out a number of recommendations with a 1-3 year timeline, several of which would significantly extend the scope of the regulatory approaches in train, for example the setting of minimum standards for cybersecurity for apps used in healthcare.
- There is significant procurement of AI by healthcare providers at the state, local government and in the private sector suggesting there is an immediate need to define clear implementation guidelines and licensing mechanisms to enable legally enforceable responsible use of AI systems in healthcare. We note however that the IP Australia AI Working Group does not include the Department of Health and Aged Care and we recommend that they be invited to join.

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.

As stated in our response to question 2, we support the AAAiH call for a National AI in Healthcare Strategy to provide strategic governance and leadership in directing a three-year strategy.

Responsibility for delivering the strategy should rest with the Department of Health and Aged Care, with Digital Health and the Therapeutic Goods Administration assuming joint responsibility. There is no need to establish a new separate regulatory and oversight organisation.

The DHCRC also believes the AI industry in Australia would benefit from access to clear licensing mechanisms to enable legally enforceable responsible use of AI systems in healthcare.

We see merit in the development of a suite of exemplar licensing templates that enshrine the safe, ethical, and trustworthy use of AI into practical licenses that can give teeth to compliance through legislation vs reliance upon self-regulation.

Developers and deployers have expressed mixed confidence in the quality and experience of legal advice and support in managing the risks and opportunities of AI systems in healthcare, suggesting training and awareness needs within the legal profession regarding understanding of AI and merit in the collaborative development of tools. We therefore see merit in the development of templates /

explainers that help clarify aspects of existing legislation that may be impeding AI developers or AI adoption or influencing overly risk averse legal advice.

We believe AI developers are receptive to the use of tools to assist in the development and adoption of data licensing agreements that help navigate the complex territory of data exchange between a licensor and licensee. The Rapid Response Information Report of 24 March 2023 addressing Generative AI (referenced in the Discussion Paper) refers to some of these tools such as the Montreal Data License Generator. This uses a questionnaire to generate intellectual property licensing terms that can be attached to datasets to govern its distribution. Contractor, D et Al Behavioral Use Licensing for Responsible AI, ACM International Conference Proceedings Series 2022 https://dl.acm.org/doi/abs/10.1145/3531146.3533143 Accessed 18 July 2023.

Deployers of AI in healthcare are already using a range of behavioural use licences and associated tools as part of their procurement guidelines. Most recently, 7 leading AI tech developers and providers in the US have prominently pledged to adhere to self-imposed codices of conduct. Use of behavioural use licensing and voluntary self-regulation frameworks are not seen as a replacement for legislation. Some examples include data licensing agreements, datasheets ensuring transparency of data lineage and ontology, watermarking of software, and watermarking of AI generated content. Those health and social care service providers new to the deployment of AI would benefit from ease of access to such templated tools.

The above recommendations complement the AAAiH AI Roadmap recommendation to develop best practice industry standards for AI developers and users to comply with regulatory and legislative requirements, work that can be channelled through the National AI Centre's Responsible AI Network.

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.

The DHCRC supports the 2021 Australian Alliance for Artificial Intelligence in Healthcare (AAAiH) roadmap for AI call for the development of a National AI in Healthcare Strategy to support and encourage collaboration and strategic leadership.

The roadmap defines a series of pragmatic, strategic and regulatory goals.

Funding and coordination of the activities outlined in detail in the roadmap for AI in healthcare report should be directed through the Department of Health and Aged Care and include public and private healthcare stakeholders to ensure there is a common, cohesive approach to the ethical and safe use of AI in healthcare.

The lack of cross-jurisdictional governance arrangements for accessing healthcare data at scale is one of the top obstacles cited by the AI industry. Government-led action at the national level is required to support AI development in Australia.

The DHCRC notes the recent release of the Australian Commission on Safety and Quality in Health Care (ACQSHC) report which advocates for a One Stop Shop (National Clinical Trials Front Door) to deliver a nationwide interconnected, rapidly responsive, streamlined, and intuitive cross-government platform to fast-track trial commencement and patient recruitment. When designed, built, tested, and implemented this would replace other national and jurisdictional systems. Such cross-jurisdictional, national, approaches are also required to provide the consistency of approach sought by AI developers.

Responses suitable for Australia

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

Policymakers globally are exploring various avenues to regulate AI technology and ensure its responsible and ethical use. Each approach has its strengths and weaknesses, with the EU's AI Act being considered the most influential and potentially setting a standard for other regions including APAC. We reference their core features, resulting pros and cons below:

1. A legally binding AI treaty by the Council of Europe:

Pros: A treaty requiring signatories to protect human rights and democracy, including potential moratoriums on risky AI technologies.

Cons: Individual ratification and implementation by each country may take years, and some countries might opt out of certain elements.

2. OECD AI principles:

Pros: Nonbinding principles adopted by OECD countries to guide AI development and policy initiatives globally.

Cons: The OECD's main mandate is economic growth, not AI regulation, requiring individual countries to translate economic principles into ethical policies.

3. Global Partnership on AI (GPAI):

Pros: Founded to encourage international research and cooperation on responsible AI and inform global policies.

Cons: Some experts criticize its low profile and lack of publications since its launch.

4. The EU's AI Act:

Pros: A sweeping regulation aiming to regulate high-risk AI usages, holding bad actors accountable, and becoming a global de facto AI regulation.

Cons: Controversial prescriptive elements, extensive lobbying by big tech aiming to dilute ethics principles, and a lengthy legislative process are potential challenges.

5. Technical industry standards:

Pros: Industry standards help companies comply with regulations and build products that work across multiple jurisdictions.

Cons: Standards are often general and may require translation for specific industries, posing a burden for small businesses.

6. United Nations' AI ethics framework:

Pros: A voluntary AI ethics framework adopted by UNESCO and member countries, allowing global influence on AI policy.

Cons: Sincerity in following ethical guidelines is questioned, and the UN's track record in tech coordination is mixed.

Source: M. Heikkilä, 'Our quick guide to the 6 ways we can regulate AI', May 22, 2023, MIT Technology Review.

Target areas

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

Different approaches should NOT apply to public and private sectors use of AI technologies in healthcare.

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

The Department of Health and Aged Care can show leadership with other Australian Government agencies by supporting the development of a National AI Strategy for Healthcare and by resourcing the Department (Digital Health and TGA) to deliver on this three-year roadmap in collaboration with National AI Centre's Responsible AI Network.

The roadmap purposely spells out a tight timeline (3 years) because of the concerns amongst users, developers, and researchers of AI that ethical and sustainable use of AI in healthcare requires urgent action under each of 8 pillars outlined in the AI in Healthcare Roadmap for Australia.

The DHCRC encourages the Health Economics and Research Division to support both internally and externally focused research to quantify the economic benefits, costs and indicators of AI enabled healthcare in national health priority areas. Many of the Australian and international reports describing the opportunities that use of AI will have in healthcare cite economic or productivity benefits. Some of these have been extrapolated from other service industries and have yet to be verified for healthcare.

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-specific solutions to assessing and mitigating risks of AI are to be avoided. There are no circumstances in which a technology-specific solution is better than a use-case specific solution.

We advocate for a 'Precision Regulation' approach that establishes rules to govern the deployment of AI in specific use-cases and does not regulate the technology itself. Precision regulation carries a high level of adaptability - new AI technologies and applications and the capabilities and performance of existing AI systems evolve at lightening-speed without generating the evidence base for risk mitigation at the same pace. Only regulation that focuses on outcomes rather than technology will be able to keep up and adapt to changing conditions quickly and efficiently. However, it needs to be pointed out that the nature of an AI tool cannot be entirely separated from the intent of using it. Every AI algorithm inherently (before any deployment, solely through the way the model is trained and tested) can contain features that could cause harm when applying it, regardless of, and sometimes even against the intent of use.

Therefore, smart AI regulation, while primarily focussing on risk-based use-case and outcomes assessments will also consider the type of AI algorithm a priori and to an extent that covers the risks related to building it in the first place.

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

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

As algorithms and AIs become ever more embedded in people's lives, there is a growing demand for transparency around when an AI is used and what it is being used for. That means communicating why an AI solution was chosen, how it was designed and developed, on what grounds it was deployed, how it is monitored and updated, and the conditions under which it may be retired. There are four specific effects of building in transparency: 1) it decreases the risk of error and misuse, 2) it distributes responsibility, 3) it enables internal and external oversight, and 4) it expresses respect for people. Transparency is not an all-or-nothing proposition, however. Regulators need to find the right balance with regards to which degree of transparency to ask from which stakeholders. Contrary to the assumptions about transparency by many organisations, transparency is not something that happens at the end of deploying a model when someone asks about it. Transparency is a chain that travels from the designers to developers to executives who approve deployment to the people it impacts and everyone in between. Transparency is the systematic transference of knowledge from one stakeholder to another - the data collectors being transparent with data scientists about what data was collected and how it was collected and, in turn, data scientists being transparent with executives about why one model was chosen over another and the steps that were taken to mitigate bias, for instance.

Source: R. Blackman, B. Ammanath, 'Building transparency into AI projects', June 20, 2022, Harvard Business Review.

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

We have developed a world-first comprehensive ethical framework for the responsible design, development, and use of generative AI technology in health and medicine (published in The Lancet, reference [1] below) alongside implementation guidelines to apply this framework through a risk-based approach (also published in The Lancet, reference [2] below):

[1] https://www.thelancet.com/journals/ebiom/article/PIIS2352-3964(23)00077-4/fulltext

[2] https://www.thelancet.com/journals/ebiom/article/PIIS2352-3964(23)00237-2/fulltext

While the developed framework is primarily applicable to health and medicine its core principles can be extrapolated to a broad field of other areas of use:

Source: https://newatlas.com/computers/case-code-of-conduct-ai-healthcare/

This framework addresses risk identification and mitigation guidelines with respect to the core ethical principles of responsible AI development and use including transparency and 'explainability', and MLOPs best-in-class engineering practices as defined by regulators such as the FDA and TGA.

10. Do you have suggestions for:

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

There should not be bans for specific types of AI technology. However, AI technology that does not satisfy quality and performance standards for specific use cases should be banned. Such standard violations can occur for example through introduction of bias/discrimination, data privacy violations and other infringements on ethical design and deployment standards. However, as laid out above, the relative weighting of such ethical factors needs to be determined as part of a risk assessment specific to individual use case scenarios. There should not be an application-agnostic ban for any type of AI technology but there could be bans of certain applications of AI (see response to Q.10b).

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

The use of AI or other technologies should be restricted or banned if they violate or imperil the exercise of human rights, do not conform to ethical principles or regulations or would be introduced in unprepared or other inappropriate contexts. For example, many countries lack data protection laws or have inadequate regulatory frameworks to guide the introduction of AI technologies. The claim that certain basic moral requirements must constrain and guide the conduct of persons can also be expressed in the language of human rights. Human rights are intended to capture a basic set of moral and legal requirements for conduct to which every person is entitled regardless of race, sex, nationality, ethnicity, language, religion, or any other feature. These rights include human dignity, equality, non-discrimination, privacy, freedom, participation, solidarity, and accountability. Machine-learning systems could advance the protection and enforcement of human rights (including the human right to health) but could undermine core human rights such as non-discrimination and privacy. Human rights and ethical principles are intimately interlinked; because human rights are legally binding, they provide a powerful framework by which governments, international organizations and private actors are obligated to abide.

Source: World Health Organisation, 'Ethics and Governance of AI for Health', 2023 Guidance.

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

- Education
- Watermarks
- Look up websites, portals
- Data sheets
- Demonstrations of value-add
- Visible adoption of AI by governments to assist government work

Implications and infrastructure

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

The DHCRC supports the use of existing functions within the Department of Health and Aged Care currently responsible for the regulation of the quality of therapeutic products in health to include responsibility for the establishment of national assurance processes to mitigate against potential AI risks in healthcare.

Risk-based approaches

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

Yes, we do strongly advocate for a risk-based approach to assessing and mitigating AI risks and are currently developing a risk-assessment framework for the use of AI in health and medicine.

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

Hallmarks of risk-based AI regulation are agility, adaptability, and clarity. Weighing risk-against benefit-profiles risk-based regulation focuses on the outcomes of the use of AI and not on the technology itself. This approach empowers developers, providers, users, and regulators of AI technology to engage in meaningful factual discussions about real immediate risks rather than perceived hypothetical risks and to prevent risk assessments that miss the mark either by overseeing risks or by misrepresenting risks as a result of having to shoehorn individual use cases into prescriptive and rigid AI technology classifications. Of course, a coherent logical line needs to be drawn from the capabilities of AI to the results of using it.

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?

A risk-based, precision regulation approach is sector-agnostic and should be applied in all fields.

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

Yes – we would suggest these go further for healthcare e.g., with use of watermarks.

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?

As per PIAs these are either embedded within a PIA or in addition to and enshrined in a national standard for healthcare, promulgated by the TGA and Digital Health Divisions. Australia should not have seven different approaches.

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

We have developed a world-first comprehensive ethical framework for the responsible design, development, and use of generative AI technology in health and medicine (published in The Lancet, reference [1] below) alongside implementation guidelines to apply this framework through a risk-based approach (also published in The Lancet, reference [2] below):

[1] https://www.thelancet.com/journals/ebiom/article/PIIS2352-3964(23)00077-4/fulltext

[2] https://www.thelancet.com/journals/ebiom/article/PIIS2352-3964(23)00237-2/fulltext

While the developed framework is primarily applicable to health and medicine its core principles can be extrapolated to a broad field of other areas of use:

Source: https://newatlas.com/computers/case-code-of-conduct-ai-healthcare/

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

A risk-based approach for responsible AI will to be mandated through regulation. Self-regulation is commendable but cannot replace legally binding AI regulation. Accountability is key. It should apply to public and private organisations, developers and deployers. Note that we uphold this statement even in light of the recently and prominently displayed pledge for self-regulation by 7 leading AI tech developers in the US White House.

And should it apply to:

a. public or private organisations or both? Both

b. developers or deployers or both? Both