UNIVERSITY
OF WOLLONGONG
AUSTRALIA

Response to the Australian Government's Discussion Paper on *Safe & Responsible AI*

Date: 26 July 2023

Prepared by: Dr. Yves Saint James Aquino, Emma Frost, and Prof Stacy Carter

Australian Centre for Health Engagement, Evidence and Values,
University of Wollongong

ISSUE

We welcome the opportunity to respond to the Department of Industry, Science and Resources' *Safe and Responsible AI in Australia Discussion Paper*. The discussion paper provides a detailed overview of domestic and international governance mechanisms for Artificial Intelligence (AI). In addition, the paper identifies key opportunities for AI to improve economic and social outcomes, as well as some potential risks and harms associated with AI. The paper seeks advice on how to mitigate the potential risks of AI, and identifies gaps in the existing domestic governance landscape to support the responsible and safe development and adoption of AI in Australia.

OUR WORK ON HEALTHCARE AI

The Australian Centre for Health Engagement, Evidence and Values (ACHEEV) includes a team of researchers focused on *Data, AI and other health technologies*. Our healthcare AI research has been funded by NHMRC Ideas Grant 1181960, NHMRC CRE 2006545, and grants from the University of Wollongong. Our work focuses on the ethical, legal and social implications of the use of machine learning in healthcare.

We have conducted a national survey of public values and attitudes about healthcare AI,[1,2] in-depth interviews with experts,[3,4] dialogue groups with healthcare consumers,[5] and a Community Jury.[6]

A Community Jury is a deliberative democratic process, designed to produce recommendations for policymakers. Following best-practice methods, we used random recruitment and stratified selection to identify 30 Australians to participate. The jurors were diverse and closely matched to the Australian population on gender, age, ancestry, level of education, urban/regional/rural place of residence, and state or territory of residence (all states and territories were represented). The jury worked together for 24 hours, over 18 days, including three days face to face. The jury was asked to make recommendations on the following question: "Under what circumstances, if any, should artificial intelligence be used in Australian health systems to detect or diagnose disease?"

Our response below is based on our study findings, and focuses on health contexts as our area of specialisation. Health contexts demonstrate the urgency of questions regarding governance. Healthcare and public health are high-risk and high-stakes areas for any application of AI and automated decision-making

---

[1] "Measures of Socioeconomic Advantage Are Not Independent Predictors of Support for Healthcare Ai: Subgroup Analysis of a National Australian Survey." *BMJ Health & Care Informatics* (2023). https://informatics.bmj.com/content/30/1/e100714

[2] "The Adoption of Artificial Intelligence in Health Care and Social Services in Australia." *Journal of Medical Internet Research* (2022). https://doi.org/10.2196/37611.

[3] "Practical, Epistemic and Normative Implications of Algorithmic Bias in Healthcare Artificial Intelligence: A Qualitative Study of Multidisciplinary Expert Perspectives." *Journal of Medical Ethics* (2023). https://jme.bmj.com/content/early/2023/02/22/jme-2022-108850

[4] "Utopia Versus Dystopia: Professional Perspectives on the Impact of Healthcare Artificial Intelligence on Clinical Roles and Skills." *International Journal of Medical Informatics* (2023). https://doi.org/10.1016/j.ijmedinf.2022.104903.

[5] Australian women's judgements about using artificial intelligence to read mammograms in breast cancer screening. *Digital Health* (in production) DOI 10.1177/20552076231191057

[6] For more details, visit the Community Jury on Healthcare AI page https://uow.info/TAWSYN_JURY

(ADM). In addition, Australia is behind relative to other countries in developing both technology and strategies for safe and responsible implementation of healthcare AI and ADM.[7]

COMMENTS ON THE DISCUSSION PAPER

Definitions

We suggest removing the phrase "without explicit programming" from the definition of AI (page 5), as the intent on page 6 appears to be to include expert systems, which are AI applications that involve explicit programming. Across the document, we suggest focussing on both AI and ADM, since ADM can be equally harmful and raise similar ethical and safety concerns.

We note that because the paper did not make a clear distinction between regulatory and non-regulatory initiatives, responding to Guide Questions 2 and 3 was sometimes challenging. In the discussion below, we have assumed that non-regulatory initiatives are "voluntary mechanisms", as defined on page 4 of the paper.

Potential gaps in approaches

Governance of AI should seriously consider not only risks to individuals but also risks to society or communities as a whole. Current frameworks tend to focus on the former. For example, equity is a particular concern that tends to be missed in regulation that focuses on risks to individuals. Equity was emphasised in the discussion paper. Equity was also a strong emphasis in the Community Jury recommendations. We note that the discussion paper appears to take a narrow view, limiting equity concerns to bias due to non-comprehensive datasets (page 8). Non-comprehensive datasets are a significant issue, but are only a part of a larger problem of systemic and structural inequities that can be reinforced or entrenched by AI. Our work with AI experts has shown that some experts do not recognise bias as a problem requiring attention, suggesting the need for regulation.[8] In any case, bias cannot be addressed only via technical solutions: this also requires social, ethical and political expertise. **Any regulation of AI should make equity and minimisation and monitoring of bias a central priority.**

The Community Jury strongly recommended education and communication about healthcare AI for both members of the public and practitioners. A minority of medical Colleges have begun educating professionals (e.g. RANZCR), but to our knowledge information and education for the Australian public has been limited to date. Information campaigns should be led by independent bodies who have no commercial interests in the development and deployment of AI. Commercial interests can threaten trust, integrity and duties of care.[9]

Responses suitable for Australia

In our view, the Australian government should consider a horizontal approach[10] to the governance of AI. This creates a comprehensive framework that covers impacts of AI across industries, and could be underpinned by a human rights approach, as is underway in the European Union (EU). Australia has the benefit of a portfolio of prior work by the Australian Human Rights Commission which could inform such an approach. A horizontal approach to AI regulation would need to interleave with existing regulatory mechanisms, some of which are identified in the discussion paper (page 10). A horizontal approach has the benefit of setting standards across all domains and levels of government.

The Community Jury began their recommendations by specifying that there must be a charter for AI in the Australian health system and services, managed by an independent decision-making body. This recommendation was domain specific because the jury was asked to think about health applications of AI. It is possible that the Jury's intent could be addressed via a broader charter for AI to harmonise regulatory mechanisms across all sectors or domains.

Nonetheless, a horizontal approach should be complemented by sector- or industry-specific regulatory mechanisms. Sectors and industries vary in their norms, relationships and tasks. Findings across our studies show gaps in governance of AI in the healthcare context. For example, the Software as Medical Device (SAMD) approach in healthcare has made a vital contribution to governance of healthcare AI, but SAMD is

---

[7] "We Need to Chat About Artificial Intelligence." *Medical Journal of Australia* (2023). https://doi.org/10.5694/mja2.51992.

[8] "Practical, Epistemic and Normative Implications of Algorithmic Bias in Healthcare Artificial Intelligence: A Qualitative Study of Multidisciplinary Expert Perspectives." *Journal of Medical Ethics* (2023). https://jme.bmj.com/content/early/2023/02/22/jme-2022-108850

[9] "The Tangled Web of Medical and Commercial Interests." *Health Expectations* (2007): 1. https://doi.org/10.1111/j.1369-7625.2007.00432.x..

[10] "The Regulation of Artificial Intelligence." *AI & SOCIETY* (2023). https://doi.org/10.1007/s00146-023-01650-z..

limited. It focuses on risks to individuals, and lacks guidelines or mechanisms to address societal risks, such as the risk that AI systems discriminate against or perform poorly for marginalised groups. Sector-specific regulation should take into account perspectives of the professional and public stakeholders within that sector to understand the priorities, assumptions and concerns specific to that sector. For example, our interviews with experts showed the complexity of identifying which of the tasks in healthcare work are suitable for automation.[11]

Australia lacks regulatory guidance for non-locked or adaptive machine learning algorithms for healthcare applications. The US Food and Drug Administration (FDA) has made efforts towards regulating AI-enabled medical devices by proposing a "Predetermined Change Control Plan",[12] which is applicable to devices that will involve modifications implemented both manually and automatically. Our work has found that adaptive healthcare machine learning is a significant and as-yet unaddressed concern in a range of jurisdictions which needs to be urgently addressed.

Target areas

The overarching governance of AI for **public and private sectors** should not differ. It is possible for private sectors to have industry-specific regulation or guidance (e.g. codes of ethics), but this regulation should not be in conflict with nor weaker than any regulation that applies to the public sector.

Regarding generic versus technology-specific risk mitigation. Broad ethics principles, such as transparency, should be embedded in the design and deployment of any type of AI technology. However, some types of AI technology raise specific concerns that require technology-specific solutions. For example, copyright issues are raised in varying degrees depending on the datasets used to develop an AI tool. Unlike most healthcare AI tools trained using proprietary datasets, some generative AI systems use unlicensed content that raise concerns about intellectual property infringements.[13]. In addition to generic and technology-specific solutions, **problem-** or **task-specific** solutions may be appropriate. This approach is concerned with what *problem* the technology is meant to solve or what *task* the technology is designed to perform. Any task involving clinical work, such as establishing diagnosis and identifying appropriate therapy, should be considered higher risk (and require more robust regulation) compared to tasks to assist in grocery shopping. Therapeutic Goods Administration's SaMD regulation[14] uses a risk model that classifies levels of risk depending on the intended purpose or task (e.g. tasks intended to specify a treatment that may lead to death are considered high risk). It is possible that different tasks may involve the same underlying predictive algorithm, but have different levels of risks or harms.

Governance of AI should consider specific rules for **transparency** about the use of AI. Concern about transparency (and the lack thereof) is a common finding across our consumer-facing studies.[15] Transparency is prominent in the EU approach,[16] as well as in local guidelines including the NSW Artificial Intelligence Assurance Framework[17] and the Automated Decision-making Better Practice Guide.[18] Transparency requirements should cover any ADM/AI-related recommendation or decision with direct implications to clients or consumers. Preliminary findings of our scoping review[19] of literature on public views about healthcare AI ethics and governance show that there is interest in transparency with regards to:

- How accurate or effective the model is at performing tasks
- To what extent a recommendation by an AI system is overseen by a person

---

[11] "Professional perspectives on the impact of healthcare artificial intelligence on clinical roles and skills." *International Journal of Medical Informatics* (2023) https://doi.org/10.1016/j.ijmedinf.2022.104903.

[12] See https://www.fda.gov/regulatory-information/search-fda-guidance-documents/marketing-submission-recommendations-predetermined-change-control-plan-artificial

[13] "Generative Ai Has an Intellectual Property Problem." *Harvard Business Review* (2023). https://hbr.org/2023/04/generative-ai-has-an-intellectual-property-problem.

[14] See https://www.tga.gov.au/sites/default/files/how-tga-regulates-software-based-medical-devices.pdf

[15] "Australian women's judgements about using artificial intelligence to read mammograms in breast cancer screening". *Digital Health* (in production).

[16] See https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206

[17] See https://www.digital.nsw.gov.au/policy/artificial-intelligence/nsw-artificial-intelligence-assurance-framework

[18] See https://www.ombudsman.gov.au/__data/assets/pdf_file/0029/288236/OMB1188-Automated-Decision-Making-Report_Final-A1898885.pdf

[19] "Public views on ethical issues in healthcare artificial intelligence: protocol for a scoping review". *Systematic Reviews* (2022). https://doi.org/10.1186/s13643-022-02012-4

- Which data sources were used to train the AI system
- If and how the AI systems collect and store user data

Transparency should be meaningful and understandable for clients/consumers at point of use. For example, the transparency solution offered by the City of Amsterdam's AI register (page 30 of the discussion paper) does provide the possibility of transparency. However it requires knowledge of its availability, and effort and skills to access the information. To ensure that all service users, regardless of literacy levels, can access advice or explanation about AI use, information should be provided at the point of service as well. Our colleagues at the Sydney Health Literacy Lab provide strategies (e.g., testing for readability) to make complex and technical information understandable to the public.[20]

Complete banning of certain high-risk AI applications should be considered to protect the safety and welfare of Australians. We note that this approach has been taken in some jurisdictions (e.g. the draft EU AI Law). Risk evaluations that could lead to the banning of AI come down to people's values and priorities. Thus, engagement with the public will be important to determine what applications should and should not be prohibited in the Australian context.

Regarding **public trust** in AI. The discussion paper focuses on improving public trust as a central goal, and conceptualises public *distrust* as one of the main barriers to AI adoption. We suggest that this should be reframed. Rather than improving public trust, the goal should be increasing the trustworthiness of AI. Publics are diverse, and hold diverse values. They are active stakeholders in defining what counts as trustworthy behaviour from governments and corporations – whether or not this is acknowledged. Question 11 asks "What initiatives or government action can increase public trust in AI deployment to encourage more people to use AI." This puts the focus on changing the views of the public, rather than on changing AI and its implementation. This strategy risks diluting the responsibilities of manufacturers, developers and regulators to ensure AI is trustworthy—that the technology is proven safe and effective.

In relation to the issue of ensuring trustworthy AI, we note that the 15 recommendations of the Community Jury, considered as a suite, can inform this question. The Jury recommended: an independently-governed charter, a guarantee that benefits outweigh harms and healthcare system performance is not undermined by AI deployment, protection of patient rights and choice, equitable access to AI systems, mitigation of bias including representative and inclusive training datasets, training for healthcare workers, clear information about the strengths and weaknesses of particular AI systems in use, ensuring robust and real-world evaluation, regulatory oversight and ongoing mandatory monitoring and reporting, considering open source software, and implementing a comprehensive and fully-funded public community education program. This set of recommendations, taken together, arguably outline what the Australian community would expect before they would be willing to trust AI in a healthcare context. We would welcome opportunities to provide more information about these recommendations and the jury process as decision-making proceeds.

Risk-based approaches

AI is highly disruptive and has the potential to transform society. Because of this, we should take a precautionary approach and be more rather than less risk-averse.

A risk-based approach should include consideration of the **evidence base** for AI use in all sectors and industries. It is important to apply robust and clear rules of evidence, as well as ensuring that the evidence is gathered from real-world contexts. Current practices in healthcare AI technology development still relies on synthetic datasets, which contain computer-generated rather real-world data. In addition, studies have shown that there is often insufficient validation of AI systems in real-world practice.[21]. A scoping review of AI applications for breast cancer detection, for example, showed evidence gaps that suggest the performance of AI applications may match what is claimed.[22] These gaps include use of imaging data that may not represent the practice setting, the potential for bias in model training, and the lack of tests comparing AI versus human interpretation. A key limitation of a risk-based approach is that emerging technologies make it difficult to assess risk, and that the risk and harms may only be discovered once the technology is already widely deployed. An approach based on findings from research and evidence can clarify the likelihood and severity

[20] See https://www.sydneyhealthliteracylab.org.au/tips-and-tricks

[21] "A Comparison of Deep Learning Performance against Health-Care Professionals in Detecting Diseases from Medical Imaging: A Systematic Review and Meta-Analysis." *The Lancet Digital Health* (2019). https://doi.org/10.1016/S2589-7500(19)30123-2.

[22] "Artificial Intelligence (Ai) for the Early Detection of Breast Cancer: A Scoping Review to Assess Ai's Potential in Breast Screening Practice." *Expert Review of Medical Devices* (2019) https://doi.org/10.1080/17434440.2019.1610387.

of harm, or in some cases the absence of adequate evidence to support claims. Given the rapid advances in AI, regulatory approaches should be dynamic enough to respond to new evidence.

We find the discussion paper lacks emphasis on **the role and involvement of the public** in defining risk, and which application, case or industry should be considered high or low risk. Our community jury recommended a comprehensive and fully funded community education program to ensure that the community is brought along with developments in and the application of AI in health. Similarly, the *Montréal Declaration for a Responsible Development of Artificial Intelligence* (2018)[23] includes "democratic participation" as one of its principles. The democratic participation principle not only upholds the responsibility of authorities to inform the public about AI, but also the opportunity for citizens to deliberate on the social parameters of AI systems, their objectives and the limits of their use.

CONCLUSION

We welcome the government's efforts to identify gaps in the governance of AI and seek feedback on how to move forward. Based on our work on the ethics of healthcare AI, we recommend regulatory approaches that take into account both individual and societal risks, are based on research and evidence, and uphold public participation.

Sincerely,

Dr Yves Saint James Aquino
Emma Frost
Prof Stacy Carter

AUSTRALIAN CENTRE FOR HEALTH ENGAGEMENT, EVIDENCE AND VALUES (ACHEEV), BUILDING 29
UNIVERSITY OF WOLLONGONG, NSW 2522 AUSTRALIA
https://www.uow.edu.au/the-arts-social-sciences-humanities/research/acheev/ | uow.edu.au
CRICOS Provider No: 00102E | TEQSA Provider ID: PRV12062 | ABN: 61 060 567 686

[23] See https://monoskop.org/images/d/d2/Montreal_Declaration_for_a_Responsible_Development_of_Artificial_Intelligence_2018.pdf