

Safe and Responsible Artificial Intelligence in Health Care - Legislation and Regulation Review

Submission from The Social Policy Group

The Social Policy Group (SPG) is a national, non-government, not-for-profit body with specialist expertise in social policy and program design with a focus on population diversity, social and community cohesion, gender equality, health and inclusion. The Social Policy Group is funded as a Health Peak and Advisory Body.

SPG is pleased to submit its recommendations to the Department of Health and Aged Care in response to the public consultation on Safe and Responsible Artificial Intelligence in Healthcare. This submission aims to contribute to the ongoing review of legislation and regulations, ensuring that AI technologies are safely and responsibly integrated into Australia's healthcare system.

This submission should be considered alongside submissions by SPG to the following Commonwealth government consultation processes:

- Therapeutic Goods Administration consultation - Clarifying and strengthening the regulation of Artificial Intelligence (AI) Submission (expected submission 20 Oct 2024)
- Department of Industry Science and Resources - Proposals paper for introducing mandatory guardrails for AI in high-risk settings (submitted 4 Oct 2024)

Policy background

The integration of artificial intelligence (AI) technologies is driving transformative change in healthcare, offering unprecedented opportunities for innovation and improved patient outcomes. However, as AI adoption accelerates, so do the risks—raising serious concerns about the unintended consequences that could compromise patient safety. In a sector where decisions often carry life-or-death consequences and affect individuals at their most vulnerable, the stakes are high. Managed poorly, AI has the potential to cause significant harm to both individuals and society. Managed well, AI can help illuminate and

reduce disparities, protect lives, uplift human autonomy, enhance the quality of work for healthcare providers, and safeguard society from public health crises.

Embedding genuine and effective responsible AI principles throughout the entire technology lifecycle—from development to decommissioning—is therefore essential to safeguard public trust and ensure the continued strength of Australia's healthcare system.

Since 2019, global forums of AI ethics and risk discourse have witnessed an explosive proliferation of published principles, frameworks and guidelines, ultimately resulting in a crowded field of “whitepaper noise”. In the past two years, this abundance has translated into a range of national and international efforts to action and legislate the AI safety agenda.

In healthcare, SPG directs particular attention to *Collective action for responsible AI in health*, developed by the Organisation for Economic Co-operation and Development (OECD), which emphasises the transformative potential of AI in healthcare, aiming to enhance resilience, sustainability, and equity within health systems. The report outlines key areas for policymakers to focus on, such as building trust, enhancing capacity, fostering evaluation mechanisms, and promoting collaboration. The core message is that successful AI integration in healthcare depends on human-centred governance rather than just technical solutions. The OECD encourages cooperative learning and collective action to ensure responsible AI deployment that respects individual rights and improves health outcomes for all.

In Australian civil society, several broad framing documents for AI healthcare can be identified, including:

- Royal Australian and New Zealand College of Radiologists (RANZCR) Ethical Principles for AI in Medicine (2023) provides high-level guidance on the ethical deployment of AI in healthcare, particularly in fields like radiology and radiation oncology.
- Australian Alliance for Artificial Intelligence in Healthcare (AAAiH) Roadmap for AI in Healthcare for Australia (2023), a synthesis of extensive community consultation and a national survey of 180 stakeholder organisations, proposes a government-agency-led concept-level strategic plan emphasising patient safety, privacy, governance, and workforce development.

- How should artificial intelligence be used in Australian health care?
Recommendations from a citizens' jury (2024) provide recommendations from a community panel of 30 diverse Australians.

The Australian Government is currently exploring a risk-based, whole-of-economy approach to AI governance. This approach aligns with strategies adopted by other countries like the United Kingdom, Canada, and the European Union, which also focus on regulating AI based on the level of risk posed by different AI applications and sectors. Noting, no country has established a definitive solution for AI in healthcare. These frameworks typically aim to balance innovation with the need to mitigate potential harms associated with AI technologies.

The Australian Government's high-level values have been represented in Australia's AI Ethics Principles, which echo global consensus values of fairness, transparency, privacy, accountability, and human-centred values. Subsequent government activities have been understood to build on these principles to align policy with Australia's commitment to responsible AI use.

Over the past year, the Department of Industry Science and Resources is marshalling efforts to harmonise industry-oriented practices around globally emerging standards, including ISO 42001:2023 and the National Institute of Standards and Technology (NIST) Risk Management Framework (RMF), which are being relied upon to provide a scaffold for management processes, conformity assessment and audit.

Responding to AI in health

To engage meaningfully in discussions about safe and responsible AI in healthcare, we must first centre our attention on the core goals of Australia's health system. These goals are what make the sector unique and guide what should be preserved, enhanced, and protected as AI technologies are integrated. Healthcare values must remain at the heart of decisions on how technology is applied and governed. If AI holds the potential to amplify outcomes, it must be used to amplify what aligns with the principles of quality care, equity, and public well-being.

It is easy to overestimate our understanding of familiar concepts. The temptation to assume that healthcare values are fixed, universally agreed upon, or inherently embedded in future practices must be actively resisted. A thoughtful, deliberate examination of these

values is essential to ensure they are upheld and adapted to evolving healthcare challenges. Even the reasoning behind why improving public trust in AI in healthcare is considered essential is easily dismissed as self-evident. However, the notion that "trust is inherently good" is an oversimplification that warrants further unpacking.

AI technologies, though innovative and promising, can easily capture attention as the latest shiny new thing, creating the risk of focusing prematurely on domain-specific AI safety challenges. Issues like performance, data quality, bias, privacy, explainability, cybersecurity, and human-centric design are crucial to understanding and managing AI risks. However, these technical concerns must not overshadow more fundamental, first-order questions: What does the healthcare sector prioritise, and how can these priorities guide AI's integration? Addressing AI safety challenges effectively requires aligning them with the core values and goals of healthcare. The starting point must be anchored in healthcare principles, ensuring that technology serves to enhance what the sector values most. It must begin with healthcare values in the loop.

While SPG broadly supports a whole-of-economy regulatory approach to ensure the safe and responsible use of AI in healthcare, our feedback to the Department of Industry, Science, and Resources highlights critical limitations in the current framework. Specifically, an organisation-centric, engineering-standards-aligned risk management process needs to fully address the unique needs and protections required to serve the Australian public fairly. In healthcare, AI governance must be informed by sector-specific values, ensuring that core principles—such as patient safety, equity, and accountability—are embedded throughout its application.

This governance is most effectively led by the healthcare sector itself, which possesses the necessary expertise, a deep commitment to patient-centred care, and a profound understanding of the human consequences associated with healthcare decisions. With healthcare professionals ultimately responsible for patient care outcomes, it is essential that AI technologies align not just with technical standards but with the ethical and operational values that underpin the sector.

SPG's feedback on the legislation and regulation review of AI in healthcare presents a series of recommendations designed to reflect the unique priorities of the healthcare sector while addressing the specific challenges posed by AI technologies. Central to these recommendations is the need to establish and maintain an ecosystem that fosters public trust in AI within healthcare. Given that healthcare in Australia is primarily funded by the public, there is a fundamental expectation that health systems and integrated

technologies must be accountable, foremost, to serving public interests. Trust in healthcare hinges on whether Australians believe their healthcare system will protect and care for them, acting in their best interests, particularly for individuals and groups in their most vulnerable moments. Furthermore, trust is a fundamental element of the clinician-patient relationship, and if patients perceive that health decisions are being delegated to an algorithm, it may erode their confidence in the healthcare system. Accordingly, fulfilling these expectations is key to fostering public trust in AI within healthcare.

By framing our recommendations to the government around these core themes, we ensure that the discourse on AI governance extends beyond the immediate technical challenges. It also focuses on sustaining the long-term goals of Australia's healthcare system—goals that prioritise patient welfare, equitable access, and ethical standards as foundational to all AI-driven innovations.

Recommendations:

1. Align with OECD approach to responsible AI in health

As globally recognised consensus on AI in health, **Australia should consider harmonising its AI activities with the OECD approach for responsible AI in healthcare**, as represented by the Collective Action for Responsible AI in Health report. The OECD's guidance emphasises moving beyond technical compliance to focus on broader societal impacts, healthcare values, equitable access, and ethical principles. By aligning with these principles, Australia can ensure that AI supports, rather than undermines, the core mission of healthcare—delivering patient-centred care, safeguarding privacy, while promoting equity.

The OECD's focus on collective action and cooperative learning is crucial for fostering trust and accountability, particularly in publicly funded healthcare systems like Australia's. A strong emphasis on ethical values and public interest is essential to promote trust in AI technologies and ensure they genuinely serve societal needs. Furthermore, a sector-specific governance framework, informed by OECD-aligned best practices, will provide the necessary focus on healthcare-specific issues, such as individual vulnerability, explainability and equity. By building AI governance that aligns with these ethical obligations, the government can best integrate technological innovation in a manner that is not only effective but also fundamentally democratic and ethical by global standards, serving the well-being of the Australian public.

2. Establish health sector-specific AI regulation

The healthcare sector has unique needs that demand tailored regulatory approaches. While a whole-of-economy AI regulatory approach may provide a broad framework, healthcare, given its distinct characteristics, requires specific governance that addresses the particular risks and challenges inherent in public health and patient care. AI implementation in healthcare impacts human life, involving decisions that can have irreversible outcomes. Therefore, governance frameworks must be equipped to respond to the complexities of human health, integrating healthcare values and ethical principles into each stage of AI deployment.

Legislative changes and capacity building within healthcare institutions are essential to accommodate AI's transformative impact. Current regulatory frameworks, including those under the TGA, must be expanded to account for the complexities of AI integration, particularly self-learning and adaptive AI systems that pose new regulatory challenges not envisioned under earlier legislative constructs. This entails reevaluating the resources, authority, and expertise required to ensure AI technologies are appropriately governed across their entire lifecycle—from design through to deployment and eventual decommissioning.

A health sector-specific wrap-around should accompany broader, whole-of-economy legislative changes. This tailored approach should recognise healthcare's unique role in protecting individual lives and societal well-being. While supporting the Department of Industry's broad efforts to harmonise AI governance, specific healthcare sector oversight is required to ensure that all regulatory measures uphold the high standards of public safety, trust, and equity the Australian people expect.

3. Maintain healthcare values at the centre

As Australia moves toward integrating AI technologies in healthcare, maintaining and enhancing core healthcare values—such as equity, patient safety, autonomy, and privacy—remains paramount. **Healthcare values must remain at the heart of decisions on how technology is applied and governed.** If AI holds the potential to amplify outcomes, it must be used to amplify what aligns with the principles of quality care, equity, and public well-being.

The healthcare sector brings an established ethical foundation to AI governance in the form of medical bioethics, including the four principles of autonomy, beneficence, non-maleficence, and justice. These must be explicitly reinforced as AI technologies become more integral to healthcare delivery. AI introduces ethical dimensions unique to the technology that fundamentally substitutes human decision-making with machine decision-making processes. This requires that ethical frameworks address the need for human-centricity and ensure human oversight remains embedded in decision-making processes (human-in-the-loop oversight). These considerations will mean that AI remains an extension of human care, never unacceptably replacing or diminishing the role of human judgment.

The current governance approaches, including those promoted by the Department of Industry, often adopt a values-agnostic perspective, focusing primarily on technical compliance and an organisation-oriented rather than a public-oriented accountability lens. Such an approach risks neglecting the foundational healthcare values that have traditionally guided patient care. Unlike many other sectors where commercial and efficiency-driven goals predominate, healthcare's mission is centred around promoting well-being and protecting the vulnerable, which requires ethical, human-centred decision-making.

To this end, **Australia's AI governance should include prohibited use cases** where specific applications of AI are considered inconsistent with the ethical obligations of healthcare. Absolute boundary setting on grounds of public safety has established precedent across other technology classes, and mirrors similar global regulatory approaches such as the European Union's AI Act. The emphasis should be on ensuring that technology enhances healthcare's mission rather than undermining patient trust or creating unintended harms that could disproportionately impact vulnerable groups.

4. Focus on public accountability as the solution for public trust

Public accountability is fundamental to ensuring AI technologies serve the healthcare sector effectively, ethically and in a manner that sustains public trust. **As healthcare in Australia is predominantly funded by the public, there is a principal need for AI governance that guarantees transparency, fairness, and accountability to the public.** The integration of AI should be oriented towards enhancing public trust, recognising that Australian citizens will be both the funders of healthcare technology and the ones most directly impacted by AI integration, for better or worse.

In a system where public funding plays a central role, the expectations around public accountability are heightened. Australians expect that the healthcare services they finance will uphold the highest standards of safety, equity, and ethical care. This expectation extends to the integration of AI technologies, which must be demonstrably aligned with the values and interests of the public. Publicly funded healthcare implies a social contract where the government and healthcare providers are accountable to citizens for the services rendered, and AI must be no exception. Ensuring that AI systems meet public expectations for quality and ethical behaviour is key to maintaining the legitimacy and trustworthiness of the healthcare system.

Global ISO standards architecture defines trustworthiness as "the ability to meet stakeholder expectations in a verifiable way" and that "depending on the context or sector, different characteristics apply and need verification to ensure stakeholders expectations are met." [ISO/IEC TR 24028:2020]. For AI technologies in healthcare, this definition of trust underscores the need for AI to reliably align with the ethical and operational expectations of the healthcare system and its stakeholders. Trust must be sector-specific, meaning that AI in healthcare must meet standards that are consistent with the unique requirements and values of the health sector. This includes prioritising patient safety and equity with transparency and accountability and ensuring that AI technologies are continuously monitored to meet these expectations over time to achieve necessary "verification".

Public accountability demands that AI systems in healthcare not only comply with technical standards but also meet the ethical expectations of the public they serve. Transparency in AI processes and decision-making is crucial, particularly when AI technologies are involved in critical, potentially life-altering decisions. The public must be assured that AI systems are being used responsibly, with mechanisms in place to address any adverse impacts promptly and equitably. This is especially important in healthcare, where decisions made by AI systems can directly affect patient outcomes and where the potential for harm must be minimised through robust accountability structures.

To further strengthen public accountability, it is essential to establish meaningful public investigation and recourse mechanisms. These mechanisms would provide the public with clear avenues to raise concerns, challenge decisions, initiate investigations, and hold parties accountable for failures in AI governance, especially with regards to use of personal data and contest of automated outcomes. Overreliance on existing legislative instruments, such as privacy, administrative law, online safety, corporations, intellectual property, competition and consumer protection, and anti-discrimination laws, risks

creating gaps in meaningful oversight. When developing new policy and legislation, deferring responsibility to adjacent existing frameworks can fail to recognise that they are not always fit for purpose in capturing the specific nuances of AI technologies or in delivering enforceable protections that can uphold stakeholder's rights effectively.

Such an overreliance on existing legislation may evade addressing the "hard stuff"—the complexities and emergent risks of AI that are beyond the reach of conventional legal frameworks, particularly when dealing with large, foreign-owned corporate entities. A public accountability framework must move beyond mere legislative compliance and actively engage and address the unique challenges that AI technologies present. The establishment of independent, transparent avenues for public recourse is crucial for avoiding the pitfalls of relying solely on existing instruments, which may not adequately capture or enforce protections in a way that meets public expectations for justice and restitution.

5. Develop independent evaluation of AI systems

A significant element of maintaining public trust is ensuring independent evaluations of AI technologies. The Department of Industry model explores reliance on conformity assessments, which lacks the objectivity required for maintaining transparency at a level expected by the public. Establishing independent oversight, similar to how Certified Public Accountants (CPAs) provide external evaluations in financial sectors, is necessary to build public confidence. Such oversight would involve independent experts conducting evaluations to verify that AI systems are safe, ethical, unbiased, and aligned with healthcare values. Accreditation and certification models for AI evaluation, as applied to other areas of healthcare safety and quality, could serve as effective tools for achieving trustworthy and reliable AI governance.

An independent evaluation body should be established to oversee AI applications in healthcare, ensuring these technologies are continually reviewed for compliance with safety and ethical standards. **This would not only protect the public interest but also help uphold public trust by providing a transparent, third-party perspective on AI performance and risk management.** Such independent evaluations would also be critical in identifying emergent risks that may arise as AI systems adapt and evolve, thereby providing a proactive approach to safeguarding public safety and ensuring accountability.

Furthermore, the independence of evaluations is necessary to avoid conflicts of interest that could arise when organisations self-assess the safety and efficacy of their AI systems.

Just as financial audits seek to provide an unbiased assessment of financial integrity, independent evaluations of AI technologies would offer an objective view of their safety, fairness, and alignment with public values. Independent oversight would be instrumental in ensuring that AI technologies deployed in healthcare continue to meet the high ethical standards expected by the public, ultimately contributing to a trustworthy and accountable healthcare system.

6. Understand AI in healthcare as an augmented human service

AI in healthcare should be seen as an extension of human services rather than just a tool. Its broad purpose is to assist and enhance human labor by taking on tasks with cognitive dimensions - whether analyzing data, making predictions, or automating routine tasks. Given that healthcare is inherently a service industry centered on interactions between people, AI systems must be treated as augmentations to these services, not as standalone products. In identifying AI as an enhanced human service, the governance of AI in healthcare must also align with the standards and expectations traditionally associated with the health workforce. Expectations of competence and accountability that apply to healthcare providers should also extend to AI systems delivering, ensuring they meet rigorous standards and operate under appropriate oversight to maintain the quality of care.

Scope of practice and responsibility must match proven performance. There is a need to reflect on the extent to which, akin to the service workforce, that AI augmented services provision will be expected to achieving and maintaining relevant certifications, ongoing performance evaluation and performance upkeep, adhering to evidence-based practice, participating in regular competency assessments, and commitment to lifecycle upkeep activities to provide high-quality patient care and meet evolving healthcare practices.

7. Recognise AI as a behavioural and decision-affecting technology

AI technologies are inherently behavioural and decision-affecting by design, meaning they do not merely process data but actively influence the decisions made by healthcare services, care providers and the public. In the healthcare context, these decisions can be life-changing, involving choices about triaging, diagnosis, treatment plans, and dictate resource allocation. The influence of AI on decision-making necessitates scrutiny to ensure that its deployment aligns with ethical healthcare practices and protects public well-being.

The impact of AI on human decisions extends beyond the immediate context of care delivery to potentially shape long-term cultural shifts within healthcare. As AI systems become more integrated, there is a risk that healthcare providers may develop an overreliance on these tools, which could erode the clinical decision-making skills of practitioners over time. Without active vigilance, entrenchment of automation bias and skills atrophy carries the potential for undermining human-in-the-loop oversight if unrecognised or unmitigated. This shift could lead to a diminished capacity for independent judgment, ultimately affecting the quality of care delivered to patients. It is, therefore, crucial that AI systems are designed and implemented in ways that support, rather than replace, human decision-making, ensuring that clinicians retain control and responsibility for patient outcomes.

Moreover, AI's ability to influence behaviour means that it can inadvertently perpetuate biases or introduce new forms of bias if not carefully managed. AI systems learn from existing data, which may contain inherent biases that, if not corrected, can lead to inequitable outcomes. This is particularly concerning in healthcare, where biased decision-making can result in disparities in care quality across different patient populations. Ensuring that AI systems are transparent, explainable, and subject to regular ethical evaluations is essential to mitigate these risks and uphold the values of equity and fairness in healthcare.

The long-range implications of AI integration also include cultural changes in how healthcare is delivered and perceived. AI tools have the potential to reshape patient-provider relationships, shifting the dynamic from one of direct human interaction to a more technologically mediated experience. While AI can enhance efficiency and provide valuable support in decision-making, it is essential to maintain the human element that is central to effective healthcare. Patients must feel that they are being treated as individuals, with their unique needs and preferences considered, rather than being subjected to impersonal, algorithm-driven decisions. Balancing the advantages of AI with the need for empathetic, patient-centred care will be crucial in maintaining public trust and ensuring that AI serves as a beneficial tool in healthcare rather than a detriment to human connection.

8. Ensure risk capture for AI for use in healthcare

All AI technologies intended for use in healthcare should undergo comprehensive risk and impact assessment. The current model proposed by the Department of Industry

lacks the depth required to adequately identify potential risks, relying too heavily on a simplistic distinction between "high-risk" and "low-risk" systems. Such categorisation without a thorough evaluation of emergent or unintended harms can leave significant gaps in patient safety. AI technologies, particularly those driven by complex algorithms and machine learning, evolve over time, introducing new risks that may not be evident at the point of deployment. As such, continuous monitoring and oversight are crucial, even for systems initially deemed low-risk.

Given the adaptive nature of AI, healthcare-specific challenges require a dynamic, lifecycle-based risk assessment. AI systems have the potential to change functionality, scope, or impact as they learn and adapt, making static risk categorisations inadequate. To ensure ongoing safety, AI systems must be regularly re-evaluated and adjusted to mitigate risks as they interact with real-world data and environments. This approach ensures that AI technologies remain aligned with patient safety, quality of care, and ethical standards throughout their lifecycle.

Moreover, a risk assessment should extend beyond technical performance to consider the broader ethical, social, and behavioural implications of AI use in healthcare. Evaluating AI technologies must account for their impact on safety, equity, and quality of care, especially considering the unique vulnerabilities of patients. Stakeholder engagement, including input from healthcare professionals and patients, is crucial to ensure that those affected by AI tools have a voice in assessing their safety and impact.

To adequately address these risks, it is also essential to establish independent auditing and evaluation mechanisms. Current internal compliance assessments may lack the necessary objectivity for thorough oversight. Independent reviews would provide an unbiased evaluation layer, ensuring that AI systems deployed in healthcare meet rigorous standards for safety, equity, and accountability.

9. Secure Diversity, Inclusion, and Equity in AI Governance for Healthcare

AI governance in healthcare must be explicitly designed to maintain, protect and promote diversity, inclusion, and equity. The core purpose of healthcare is to protect vulnerable individuals and populations, ensuring access to high-quality care regardless of one's background or circumstances. In a country as diverse as Australia, this requires that AI technologies be developed and deployed in a manner that is inclusive of all cultural,

ethnic, and social groups, with particular attention to the needs of vulnerable and underrepresented communities.

The omission of diversity and inclusion considerations from the current proposed AI guardrails, as exemplified by the removal of Voluntary Guardrail 10, represents a significant oversight. Healthcare systems are tasked with not only treating illness but also safeguarding the well-being of individuals in their most vulnerable states. This inherently requires a human-centred approach that respects and adapts to the diverse cultural contexts of those it serves. AI technologies must be subject to governance frameworks that prioritise these values and ensure that marginalised communities are not disproportionately disadvantaged by the implementation of new technologies.

To sustain cultural inclusion measures necessary for Australia's diverse population, it is essential that legislative and regulatory frameworks incorporate explicit provisions for diversity and equity. This includes involving stakeholders from diverse backgrounds throughout the AI lifecycle—from design to development to deployment and evaluation. By actively including input from a wide range of voices, healthcare can ensure that AI technologies not only avoid perpetuating existing biases but also actively contribute to reducing health disparities.

A commitment to diversity and equity is not merely a matter of ethical preference but a foundational requirement for ensuring that AI serves all Australians effectively. AI technologies that fail to consider the varied needs of the population risk exacerbating existing inequities in healthcare outcomes. The omission of these considerations from regulatory guardrails leaves vulnerable groups at risk of exclusion or harm, undermining the public trust that is essential for the successful integration of AI into healthcare. Therefore, any legislative course for AI in healthcare must include concrete requirements for diversity, inclusion, and equity, ensuring that these principles are upheld across all levels of AI governance.

10. Address Australia as a deployer marketplace

Recognising Australia's role as primarily a deployer marketplace for AI technologies developed overseas brings unique considerations for healthcare, particularly regarding public trust and data sovereignty. As a country that predominantly adopts and implements AI technologies developed elsewhere, Australia faces distinct challenges in ensuring that these technologies align with national values and meet public expectations for safety, accountability, and ethical standards. The reliance on externally developed AI

means that Australia must be vigilant in assessing these tools to ensure they are suitable for the local healthcare context and uphold the standards expected by Australian citizens.

Public trust is a cornerstone of effective healthcare delivery, and the integration of AI technologies heightens the need for robust trust-building measures. Given the public funding of healthcare in Australia, citizens expect that any technology integrated into their healthcare system will serve their best interests, with particular attention to privacy, safety, and ethical use. AI technologies must be transparent in their operation, and the public must be confident that these tools are being used responsibly, with clear mechanisms in place for accountability. This includes providing patients with understandable information about how AI is being used in their care, what data is being collected, and how decisions are made.

The issue of data sovereignty is particularly important in the context of Australia's healthcare system. AI systems rely on vast amounts of data, often including sensitive patient information. Ensuring that this data is stored, processed, and managed within Australian jurisdiction is critical for maintaining control over how the data is used and protecting patient privacy. Geographic data residency ensures that Australian laws govern the use of healthcare data, providing an additional layer of security and accountability that is crucial for public trust. Without proper data sovereignty measures, there is a risk that sensitive health data could be subject to foreign regulations or used in ways that do not align with Australian standards for privacy and patient rights.

Balancing the need for access to cutting-edge AI technologies with the requirements for data sovereignty and public trust is essential for the successful integration of AI in Australian healthcare. This involves not only stringent regulatory oversight but also clear communication with the public about how their data is being protected and how AI technologies are enhancing healthcare services. By prioritising data sovereignty and aligning AI deployments with national values, Australia can build a healthcare system that leverages the benefits of AI while safeguarding the rights and interests of its citizens.

11. Build a comprehensive skills base to support healthcare AI Governance

To effectively govern AI integration in healthcare, **it is crucial to establish and maintain AI literacy and proficiency that encompasses both health sector-specific workforce training and public education to empower health consumers.** This approach will ensure

meaningful human oversight and sustain clinical autonomy in a future where AI technologies are increasingly integrated into healthcare decision-making.

For the public, there must be a focus on AI literacy to empower individuals to understand how AI impacts healthcare services, enabling them to make informed choices and engage meaningfully with AI-enhanced healthcare. Public education initiatives should provide insights into how AI influences decision-making, patient care, and data usage, thereby fostering trust and ensuring that individuals can navigate an AI-driven healthcare landscape with confidence.

Within the healthcare workforce, specialised training is essential to equip care providers with the skills to work effectively with AI tools and within AI service ecosystems. Healthcare providers need to be trained to interpret AI-generated insights, understand AI's limitations, and recognise potential biases in machine-driven recommendations. This training will empower providers to critically assess AI outputs and interrogate "black box" processes for adequate explainability, ensuring that AI augments rather than replaces human decision-making. Only by maintaining the human clinician authority and autonomy, can the healthcare workforce leverage AI tools while ensuring that health remains high-value care and human centered.

Human-in-the-loop oversight is a key component of patient safety and upholding ethical standards in healthcare. HITL oversight places humans at the centre of AI-supported decision-making, especially in high-stakes medical scenarios where AI outputs can significantly impact patient outcomes. To achieve this, healthcare workers must be equipped not only with AI proficiency skills but also with the confidence to question, adjust, or override AI recommendations when necessary.

Collaboration among healthcare institutions, educational providers, government bodies, and industry stakeholders is required to develop both public and workforce skills. Investment in continuous professional development will ensure that healthcare professionals keep pace with AI advancements, adapting their knowledge and practices accordingly. By building a comprehensive skills base across both the public and the healthcare sector, Australia can create a healthcare environment where AI genuinely supports clinicians, enhances patient outcomes, and maintains public trust in the evolving healthcare system.

12. Develop sector-specific legislative instruments and agencies

A radical review of health-specific legislative instruments will be necessary to meaningfully action the raft of proposed new AI governance activities. While this recommendation is addressed in further detail in the accompanying submission to TGA, a summary of the issues is provided as follows:

The current scope of TGA legislation is insufficient for managing the complexities of AI technologies, particularly as AI increasingly blurs the lines between professional and consumer-facing applications and between discrete clinical tools and fully integrated healthcare systems. AI tools now operate across multiple contexts, introducing complexities that the existing TGA framework is not equipped to handle. For instance, TGA's regulations were initially designed for specialised hardware and have since been extended to software as a medical device (SaMD). However, these adjustments fail to capture the nuanced risks posed by AI, especially those related to iterative learning, continuous adaptation, and the broader societal impacts beyond individual patients.

The current legislation relies heavily on a single evaluation point, which is not suitable for AI technologies that continuously evolve after deployment. AI systems are inherently dynamic, requiring ongoing assessment and regulation rather than a one-time approval. This is particularly problematic for large-scale software systems, such as state-wide electronic medical records, which are too embedded to be recalled or re-evaluated easily. The absence of provisions for iterative evaluations exposes healthcare systems to the risk of unaddressed safety concerns, emergent harms, and scope creep.

Additionally, current TGA legislation does not adequately account for non-individual impacts of AI, such as societal or systemic risks. The use of AI in healthcare can have far-reaching consequences that extend beyond individual patients, affecting groups or entire communities. The lack of mechanisms to address these broader impacts means that public accountability is not sufficiently met under the current framework. As healthcare in Australia is predominantly funded by the public, this shortfall undermines the expectation that AI in healthcare should serve the collective public interest.

To address these limitations, comprehensive reform is required - not only through significant legislative updates but also by expanding the powers and resources necessary to enforce these new regulations effectively. This may involve reconsidering whether the TGA or another regulatory body is best suited to oversee AI governance in healthcare. Regardless of which entity takes responsibility, the distinct values and priorities of the

healthcare sector demand a specialised regulatory approach that aligns with healthcare ethics and the expectations of the Australian public. While harmonising with a whole-of-economy AI framework is important, a health-specific regulatory framework is crucial to ensure responsible AI deployment, maintain public trust, and uphold the integrity of healthcare systems.

Conclusion

The integration of AI in healthcare presents an unprecedented opportunity to improve the quality, accessibility, and efficiency of care. However, these benefits can only be fully realised if AI technologies are carefully governed to ensure they genuinely serve the interests of patients and the public. To achieve this, a strong emphasis on sector-specific governance, alignment with globally recognised ethical principles, and robust public accountability mechanisms is essential. AI in healthcare must be deployed not as an isolated technology but as part of an ecosystem that is committed to enhancing human-centred care, protecting privacy, and upholding the values that form the bedrock of healthcare.

The recommendations outlined in this document call for the development of a comprehensive AI governance framework tailored to the unique needs of the healthcare sector. By building on existing bioethical foundations, emphasising the importance of public accountability, and ensuring that independent oversight is a core element of AI evaluation, Australia can develop an AI ecosystem that is not only technologically advanced but also safe, trustworthy, and equitable. These measures will help guarantee that AI technologies enhance—rather than undermine—the trust that underpins Australia’s healthcare system.

Ultimately, the success of AI in healthcare depends on our ability to align innovation with ethical obligations and public interest. The path forward must be one that integrates AI in a manner that prioritises patient welfare, promotes transparency, and holds technology developers and implementers accountable. By focusing on these key areas, we can harness the transformative potential of AI to create a healthcare system that is more responsive, resilient, and effective—ensuring that technological progress is always aligned with the overarching goal of improving the health and well-being of all Australians.

References

Anderson, B. and E. Sutherland (2024), "Collective action for responsible AI in health", OECD Artificial Intelligence Papers, No. 10, OECD Publishing, Paris, <https://doi.org/10.1787/f2050177-en>.

ISO/IEC. (2020). *ISO/IEC TR 24028:2020 Information technology – Artificial intelligence – Overview of trustworthiness in artificial intelligence*. International Organization for Standardization.

ISO/IEC. (2023). *ISO/IEC 42001:2023 - Information technology – Artificial intelligence – Management system*. International Organization for Standardization.

National Institute of Standards and Technology. (2023). *AI Risk Management Framework (AI RMF) 1.0*. U.S. Department of Commerce. Retrieved from <https://www.nist.gov/itl/ai-risk-management-framework>