



RACP
Specialists. Together
EDUCATE ADVOCATE INNOVATE

**Submission to the Royal Australian and
New Zealand College of Radiologists
(RANZCR)**
Ethical Principles for Artificial Intelligence in Medicine

May 2019

About The Royal Australasian College of Physicians (RACP)

The RACP trains, educates and advocates on behalf of over 17,000 physicians and 8,000 trainee physicians, across Australia and New Zealand. The College represents a broad range of medical specialties including general medicine, paediatrics and child health, cardiology, respiratory medicine, neurology, oncology, public health medicine, occupational and environmental medicine, palliative medicine, sexual health medicine, rehabilitation medicine, geriatric medicine, and addiction medicine. Beyond the drive for medical excellence, the RACP is committed to developing health and social policies which bring vital improvements to the wellbeing of patients.

The Royal Australasian College of Physicians (RACP) welcomes the opportunity to provide feedback on the The Royal Australian and New Zealand College of Radiologists' (RANZCR) set of ethical principles governing the adoption of artificial intelligence (AI) in medicine, specifically with regards to clinical radiology and radiation oncology

We understand these principles have been developed by RANZCR's AI Working Group to complement existing medical ethical frameworks by addressing the emerging use of machine learning (ML) and AI in medicine and seek to inform:

- the development of standards of practice for research in AI tools
- the development of standards of practice for deployment of AI tools in medicine
- the upskilling of radiologists and radiation oncologists in ML and AI, and
- the ethical use of ML and AI in medicine.

As outlined in your invitation for feedback, you are seeking input from stakeholders on the specific wording of the eight principles; whether anything important is missing and how this ethical framework could be implemented across medicine.

To produce this response, the RACP has consulted with relevant committees including its Ethics Committee, Consumer Advisory Group, e-Health Reference Group members in addition to RACP-affiliated Specialty Societies.

Feedback on the set of principles

Principle One: Safety

This principle states that “the first and foremost consideration in the development, deployment or utilisation of ML systems or AI tools ought to be patient safety and quality of care, with the evidence base to support this”. We would recommend strengthening the wording of this statement as follows to emphasise patient safety: “The first and foremost consideration in the development, deployment or utilisation of ML systems or AI tools **must** be to enhance patient safety and quality of care, with the evidence base to support this”.

Principle Two: Avoidance of Bias

This principle outlines that “the data on which ML systems and AI tools is based should be representative of the target patient population on which the system or tool is being used” and that “particular care must be taken when applying an AI tool trained on a general population to indigenous or minority groups”. We recommend replacing this latter statement with the following wording to avoid ambiguity around who is responsible for mitigating bias when using AI tools: *‘It will be the responsibility of the local clinical governance committee to determine whether the AI tool can be applied to their local populations and the limitations of this application, particularly around minority groups’*. E.g. older adults where evidence is often limited. In addition, we recommend adding in that these reservations regarding potential bias also apply to older Australians particularly those with cognitive impairment as well children, and children with disabilities who are frequently under-represented in research. We recommend that specific attention be paid to identifying and avoiding bias in relation to the needs of First Nations peoples - Aboriginal and Torres Strait Islander people and Māori. As this principle states that “.. bias in algorithmic design should be minimised by involving a range of perspectives and skill sets in the design process.”, then this should involve a specific focus on mitigating bias for Aboriginal and Torres Strait Islander people and Māori.

Whilst we agree that avoidance of bias is key in the use of AI tools, it is not clear how feasible it may be to use data representative of the target population when it applies to Indigenous peoples or minority groups in particular. There is a concern that excising a minority group into a smaller data set may make them more identifiable rather than less. The Maianayri Wingara Indigenous Data Sovereignty Collective has endorsed key principles in relation to Indigenous data sovereignty and Indigenous data governance. These include protocols for assuring anonymity and maintaining confidentiality. Te Mana Raraunga – Māori Data Sovereignty Network have a charter that includes principles for Māori data access and control¹.

In addition, there could also be a risk that overall data is degraded by taking groups out which could result in a reduction in quality of care. At a practical level, there also needs to be greater consideration of the

¹ <https://www.maiamnayriwingara.org/key-principles>
<https://www.temanararaunga.maori.nz/>

importance of taking particular care when applying an AI tool trained on a general population to Indigenous peoples or minority groups can be enforced when the patient identifies as being from these groups.

These concerns need to be appropriately balanced against the opportunity that ML and AI can offer us in tackling disparities and developing targeted and effective approaches dependent on a person's cultural, geographic and genetic profile.

Principle Three: Transparency and Explainability

This principle states that “ML systems and AI tools can produce results which are difficult to interpret or replicate. When used in medicine, the doctor must be capable of interpreting how a decision was made and weighing up the potential for bias. This may require upskilling for medical practitioners.” This principle may be further clarified by outlining that explainability would be enhanced by ensuring the original data used in ML systems and AI tools is robust, therefore making decisions based on this data more reliable.

In addition, this principle could be improved by encouraging the use of information for health consumers in plain everyday language and making it available in different languages along the lines of Consumer Medicines Information (CMI) leaflets provided for over the counter and prescription medications.

We would also recommend rewording this sentence “When designing a ML system or AI tool, consideration must be given to how the decision made can be understood and explained by a discerning medical practitioner” as follows: ‘When designing *and implementing* a ML system or AI tool, *it must be clear* how the decision made can be understood and explained by a discerning medical practitioner”.

One of our members made the following observation on this principle and suggested that this principle be renamed “Transparency and Explicability”: “The main thrust of modern AI currently focuses on gradient descent with back-propagation. This is inherently inscrutable in that all you get at the end is a matrix of coefficients. Even worse, this matrix is not robust in the face of small changes in input data, which can even be modified to deliberately ‘confuse’ it. This limitation should, I believe, be clearly acknowledged, especially in the specific circumstance where a doctor disagrees with the machine, and the outcome is adverse, whether he/she chose to go with the “machine advice” or not.”

Principle Four: Privacy and Protection of Data

This principle rightly acknowledges that healthcare data are amongst the most sensitive data that can be held on an individual and that it must be stored securely and in line with relevant laws and best practice. It also outlines that “where data are transferred or otherwise used for AI research, they must be de-identified such that the patient's identity cannot be reconstructed.” One of our consumer representatives questioned whether such a level of de-identification is possible and if so, whether the resulting data would have research value. This consumer representative's recommendation is that the issue of focus should be liability and that the onus to prove the data is stored safely and is not misused should rest with the researcher or clinician using the data. One of our members also advised that work from [Professor Latanya Sweeney](#) from Harvard University has demonstrated that it is often possible to ‘re-identify’ patients from their data.

Another consumer representative questioned whether the statement “Patient data must not be transferred from the clinical environment at which care is provided without the patient's consent or approval from an ethics board” was accurate as ethics boards may require patient consent.

The issue of Indigenous data sovereignty must underpin all decisions involving data. Aboriginal and Torres Strait Islander people and Māori have the right to govern the creation, collection, ownership and application of their data². The collection of information needs to be respectful of communities' self-determination, including protocols for assuring anonymity and maintaining confidentiality. Data should be appropriately stored with confidentiality and access protocols in place.

Secondary use of data pertaining to Aboriginal and Torres Strait Islander peoples should require approval from Aboriginal Human Research Ethics Committees (HRECs) as these HRECs have specific expertise on these matters. Data custodians should be guided by these specific data governance agreements pertaining to the secondary use of data involving Aboriginal and Torres Strait Islander peoples. Use of data for research should also be consistent with frameworks for Māori research ethics (i.e. Te Ara Tika).

² <https://www.maiamnayriwingara.org/key-principles>
<https://www.temanararaunga.maori.nz/tutohinga>

We would also recommend that this principle specifies that audit is necessary to protect patients against systemic errors and that a balance needs to be achieved: if audit and quality improvement are impeded by excessively onerous requirements and failure of trust, then there is a risk that patients could be seriously harmed.

Principle Five: Decision Making on Diagnosis and Treatment

We agree that AI tools and ML systems are tools to be used as an aid to a medical professional and that the human element in health care must be preserved. We recommend rewording this statement slightly to re-emphasise that treatment decisions are recommended by the doctor in consultation with the patient in line with good practice and evidence-based patient-centred care.

In addition, this principle needs to acknowledge that to justify their use, there must be robust evidence that the use of ML/AI enhances relevant clinical and patient centred outcomes at an acceptable cost to the community, as would be the case for any other new intervention or treatment.

One of our members observed that “An even more important principle here—one not apparently acknowledged by the document—is that modern ‘AI’ does not and currently cannot accommodate causal hypotheses and counterfactuals. As the heart of modern medicine should be scientific testing of causal hypotheses and exploration of alternatives, we need to emphasise this deficiency, and point out that for this reason, ‘the data’ cannot drive decision making.

Principle Six: Liability for Decisions Made

One of our consumer representatives raised concerns about this principle potentially making it more difficult for the consumer to know who is responsible if anything goes wrong as a result of the use of AI tools or ML systems for their care. Information about potential risks and who is responsible for managing these risks must be communicated clearly to consumers prior to the use of AI tools and ML systems in their care.

Another consumer representative remarked that this principle could be contentious amongst health technology researchers, regulators, community members and medico-legal experts who may have differing views about liability. Issues of liability also depend on the maturity and advancement of the ML/AI technology used, the extent the technology is activated, and what the diagnostic or treatment procedure is.

We recommend strengthening this principle by adding that the burden of responsibility is shared with quality assurance, compliance/accreditation requirements and quality improvement programs and that it is important that such programs are established in advance and not on an ad hoc basis.

One of our members observed that there are a number of issues not currently addressed by this principle and suggested RANZCR gives consideration to the following:

- “The mere presence of the software may impair decision making in several ways. Apart from bad advice, there is also the issue of distraction and information overload. The “ML” must be viewed as a component in a complex system that can and likely will produce revenge effects.
- As shown recently by Boeing, software can kill. In the medical context however, “hold harmless” clauses may deflect liability away from the software designer, and onto the institution or doctor. This is so serious a threat that I believe such clauses should specifically be forbidden, if software is to be used at all.
- Similarly, confidentiality agreements and hiding of code may prevent adequate examination of software to find the latent faults. A reasoned case should be made for confidentiality agreements not being binding in the case of suspected harm due to software, and an even more clear case can be made for mandating open-source code where patient harm is possible as a consequence of the behaviour of software.”

Principle Seven: Application of Human Values

The importance of patient-centred care is currently missing from this principle and needs to be added. We would recommend that this principle also acknowledges that doctors and patients will have different interpretations of ethics and that these interpretations may also differ between professionals.

In addition, this principles should acknowledge that AI tools and ML systems should only be used to recommend a series of treatment options rather than to prescribe an outcome which should always be decided in the best interest of the patient in a manner where informed choices can be made through a mutual exchange of information via tools such as advance health care directives and facilitating mutual health literacy approaches between physicians and patients. For example, if a patient does not want resuscitation, this should be taken into account by the AI tools and ML systems despite the fact that it would reduce options.

We would recommend adding in explicit references to patients' goals of care and to the [National Safety and Quality Health Service Standards](#) in this principle.

Principle Eight: Governance

We recommend this principle specifies that the accountable governance committees tasked with overseeing the implementation of AI tools and ML systems and to ensure compliance with ethical principles and standards should be required to have trained consumer representatives and a practicing clinician as members.

How this ethical framework could be implemented across medicine?

Educating future trainees and existing specialists in the art of interpreting results from AI, and the ethics around AI in medicine will become necessary. This could be achieved by including the development of workshops or additions to existing lecture series in order to incorporate this facet of medicine into training. With regards to the implementation of this ethical framework across medicine, one of our consumer representatives suggested the use of 'ethics' promotional tools and a focus on improving health literacy.

Additional feedback and observations

In addition to specific feedback on the principles outlined above, the following additional comments and observations have been provided by RACP-affiliated specialty societies and individual members for the RANZCR's consideration:

- The Medical Oncology Group of Australia's (MOGA) Ethics sub-committee
 - The evidence required to adopt AI in Medicine as standard practice should be detailed or referenced in the set of principles.
 - This ethical framework would benefit from outlining how their practice will be introduced if these ethical principles are adopted/implemented including what training will be provided, how the practice will be monitored and what feedback loops will assist in constant learning and quality improvement.
 - The key issue is that introduction of AI, as for any new standard of care in medicine, should be subject to ethical evaluation including clinical trials in the human population as per the current NHMRC standards. Evidence supporting the 'Ethical Principles' should be critically evaluated, in reference to the target populations and the intended applications.
 - As with any other diagnostic test a patient should be informed about its risks and benefits (including its sensitivity and specificity) and patient confidentiality should be maintained. The governance structure of a hospital should ensure that these issues are addressed, and AI is a tool of the doctor and therefore should not be seen as a threat to the doctor/patient relationship.
- The Australian Rheumatology Association's (ARA) via its Australian Musculoskeletal Imaging Special Interest Group (AMISIG)
 - This is a great first step towards a very important issue.
 - The key factors in the use of MLS and AI go beyond the images created, and centre on the decision tools and computer-derived algorithms that are increasingly being implemented.
 - A reliable mechanism for feedback from patients and doctors who use the AI tools is essential in the early roll-out of these tools. This aspect should be emphasised in the set of principles, in particular Principles One and Six.
 - The ARA agrees that doctors who are ultimately responsible for the final decision (whether right or wrong) will require specialised training and upskilling, to implement ML systems and AI tools.
 - The ARA does not fully agree with the statement that MLS and AI tools "trained on greater volumes and varieties of data should be less biased." Algorithm bias should be known and measurable, rather than 'minimised,' so as to inform users of potential limitations and

precautions with the findings. For example, in Principle Two, care needs to be taken with the application of MLS and AI tools to populations across different ethnic groups, such as using an app created using Asian population information as the reference point for the diverse Australian population.

- In relation to the ethical use of MLS and AI, the ARA has formed the view that the RANZCR may need to explore specific policies on the ethical use of MLS and AI. The ethical framework of the Australian Medical Council's Good Medical Practice – *Professionalism Ethics and Law* is insufficient to cover the broader issues in the ethical use of MLS and AI in practice.
 - This draft is a sound beginning, and the ARA believes that as rheumatologists who use imaging, they are able to offer assistance and further involve themselves from an ultrasound perspective to help with imaging interpretation and other clinical and blood parameters for developing valid and reliable computer-aided decision-making systems. It is anticipated that these systems will in the future serve to enhance the doctor's decisions with diagnosis, prognosis and provide supports for treatment decisions.
- The Australian and New Zealand Society for Geriatric Medicine (ANZGSM) via its Clinical Issues Committee and other members:
 - Under Principle 3: ANZGSM agrees that the issue of consent needs to be clear. Given the concern around the decision-making capacity of older patients, has the issue of surrogate consent been addressed?
 - Regarding Principle 6, using analogies from other areas such as autonomous vehicles and Facebook-based 'research', as has been reported in the general media, responsibility/liability will be a significant hurdle to overcome. For example, social media platforms avoid responsibility and see themselves as just facilitators or conduits of information with no responsibility for how their technology is used.
 - ANZGSM agrees with MOGA's comment regarding informed consent including risks and benefits and comparison with alternative treatments offered, including doing nothing and confidentiality. In a broader sense, a focus on including the education of future trainees and existing specialists in the art of interpreting results from AI, and the ethics around AI in medicine will become necessary. This could be achieved by including the development of workshops or additions to existing lecture series in order to incorporate this facet of medicine into training.
 - Individual member:
 - The principles outlined are probably suitable if the item being reviewed is routine with no variance. However, it needs to be checked by a human specialist if there is any variation no matter how minor. This is a similar concept to aviation where if the human cannot override the flight systems when there is a problem or variance from that programmed disaster is likely, cf 737-8 Max. Also, the clinical detail provided by those requesting the imaging needs to be complete, because if not algorithm can go in wrong direction e.g. request for CT Brain misses possibility of head trauma so the appropriate procedure is not performed and the base fracture of the skull is missed as has happened in a real case.