The Royal Australian and New Zealand College of Radiologists®

Ethical Principles for Artificial Intelligence in Medicine

The nine ethical principles outlined below guide the development of professional and practice standards regarding the research and deployment of machine learning (ML) systems and artificial intelligence (AI) tools in medicine.[1] These tools should at all times reflect the needs of patients, their care and their safety, and they should respect the clinical teams that care for them. Within this document, the term "AI tools" includes all variations of simple machine learning and complex deep learning acting as AI in clinical decision support. A full list of definitions is included in Appendix One.

These principles are intended to guide all stakeholders involved in research or deployment of AI tools including developers, health service executives and clinicians. They are also designed to complement existing medical ethical frameworks (see appendices), which do not address the issues likely to emerge from use of AI in medicine.

In order to bridge this gap, the Royal Australian and New Zealand College of Radiologists (RANZCR) has developed nine ethical principles specifically to guide the following:

- development of standards of practice for research in AI tools
- regulation of market access for AI tools
- development of standards of practice for deployment of AI tools in medicine
- upskilling of medical practitioners in AI tools, and
- ethical use of AI tools in medicine.

All stakeholders should take heed of all the ethical principles for AI in medicine, noting that some will have greater applicability to them.

Principle One: Safety

Although AI tools have enormous potential, a range of new risks will emerge from AI tools or through their implementation.

The first and foremost consideration in the development, deployment or utilisation of AI tools must be patient safety and quality of care, with the evidence base to support this.

Principle Two: Privacy and Protection of Data

Healthcare data is amongst the most sensitive data which can be held about an individual. Patient data must not be transferred from the clinical environment [2] at which care is provided without the patient's consent, approval from an ethics board or where otherwise required or permitted by law. Where data is transferred or otherwise used for AI research, it must be de-identified such that the patient's identity cannot be reconstructed.

Every effort must be made to store a patient's data securely and in line with relevant laws and best practice.

[1] RANZCR has adopted definitions of AI and ML (see Appendix 1)
[2] A clinical environment is any area relating to patient treatment or diagnosis and may include physical or secure virtual environments.

Principle Three: Avoidance of Bias

AI tools are limited by their algorithmic design and the data they have access to making them prone to bias. As a general rule, AI tools trained on greater volumes and varieties of data should be less biased. Moreover, bias in algorithmic design should be minimised by giving conscious consideration to avoiding bias and involving a range of perspectives and skill sets in the design process.

The data on which AI tools are based should be representative of the target patient population on which the system or tool is being used. The characteristics of the training data set and the environment in which it was tested must be clearly stated when marketing an AI tool to provide transparency and facilitate implementation in appropriate clinical settings. Particular care must be taken when applying an AI tool to a population, demographic or ethnic group for which it has not been proven effective.

To minimise risk of bias, the process, training data set and outcome measures used during development must be transparently stated.

Principle Four: Transparency and Explainability

AI tools can produce results which are difficult to interpret or replicate. When used in medicine, the medical practitioner must be capable of interpreting the basis on which a result was reached, weighing up the potential for bias and exercising clinical judgement regarding findings.

AI tools should ideally employ explainable AI (XAI) techniques to justify the underlying basis for decision-making in a way that is understandable to humans.

When designing or implementing an AI tool, consideration must be given to how a result that can impact patient care be best understood and explained by a medical practitioner.

Principle Five: Application of Human Values

The development of AI tools for medicine should ultimately benefit the patient and society. ML and AI are programmed to operate in line with a specific world view, however the use of AI tools should function without unfair discrimination and not exacerbate existing disparities in health outcomes. Any shortcomings or risks in AI tools should be considered and weighed against the benefits of enhanced decision making for specific patient groups.

The medical practitioner must apply humanitarian values (from their training and the ethical framework in which they operate) to any circumstances in which AI tools are used in medicine, but must also consider the personal values and preferences of their patient in this situation. Entities developing AI tools must demonstrate an understanding of ethical principles and human values.

Principle Six: Decision-Making on Diagnosis and Treatment

Fundamental to quality healthcare is the relationship between the medical practitioner and the patient. The medical practitioner is the trusted advisor on complex medical conditions, test results, procedures and treatments who then communicates findings to the patient clearly and sensitively, answers questions and agrees on the next treatment steps.

While AI tools can enhance decision-making capability, final decisions about care are made after a discussion between the medical practitioner and the patient taking into account the patient's presentation, history, options and preferences.

Principle Seven: Teamwork

ML and AI in research and medicine will need new skillsets and teams. It is imperative that all team members know each other's strengths, capabilities and integral role in the team.

In order to deliver the best care for patients, each team member must understand the role and contribution of their colleagues and leverage them through collaboration.

Principle Eight: Responsibility for Decisions Made

Responsibility for decisions made about patient care rests principally with the medical practitioner. Medical practitioners need to be aware of the limitations of AI tools, and must exercise solid clinical

judgement at all times. However, given the multiple potential applications of AI tools in the patient journey, there may be instances where responsibility is shared between:

- The medical practitioner caring for the patient;
- The hospital or practice management who took the decision to deploy the systems or tools; and
- The manufacturer that developed the ML system or AI tool.

Although the prime responsibility regarding patient care remains with the medical practitioner, when using AI tools, the responsibility is also shared by the managers of the healthcare environment and the manufacturers and developers of AI tools. This potential for shared responsibility when using AI tools must be identified, recognised by the relevant party and recorded upfront when researching or implementing AI tools.

Principle Nine: Governance

ML and AI are fast moving areas with the potential to add great value but also to do harm. The implementation of AI tools requires consideration of a broad range of factors including how the ML or AI will be adopted across a hospital or practice and which patient groups will be affected and how it might align with patients' goals of care.

A hospital or practice using or developing AI tools for patient care applications must have accountable governance to oversee implementation and monitoring of performance and use, to ensure the practice is compliant with ethical principles and standards.

Broader Ethical Frameworks

Other ethical frameworks cover the expected approach and behavior of medical practitioners when delivering care to patients and provide general guidance relating to the development and adoption of new technologies in medicine.

Medical practitioners in Australia are expected to practise in accordance with the Medical Board of Australia's Good Medical Practice: A Code of Conduct for Doctors in Australia [i] and the Australian Medical Council's Good Medical Practice.[ii]

Medical practitioners in New Zealand are expected to practise in accordance with the New Zealand Medical Council's Good Medical Practice [iii] and the Code of Ethics set by the New Zealand Medical Association.[iv] Medical Practitioners in New Zealand must also comply with the Code of Health and Disability Services Consumers' Rights. [v]

RANZCR has also developed a more explicit Code of Ethics for clinical radiologists and radiation oncologists.[vi]

Appendix One – Definitions

Technical definitions for artificial intelligence are available from the International Organization for Standardisation (ISO) [vii], general definitions are included below.

Artificial Intelligence

"An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy." [viii]

Explainable Artificial Intelligence (XAI)

"A set of processes and methods that allows human users to comprehend and trust the results and output created by machine learning algorithms". [ix]

Algorithm

"A series of instructions for performing a calculation or solving a problem, especially with a computer. They form the basis for everything a computer can do and are therefore a fundamental aspect of all AI systems." [x]

Bias

"A systematic deviation from the truth." [xi]

Variance

"A random deviation from the truth." [xi]

Expert system

"A computer system that mimics the decision-making ability of a human expert by following pre-programmed rules, such as 'if this occurs, then do that'. These systems fuelled much of the earlier excitement surrounding AI in the 1980s, but have since become less fashionable, particularly with the rise of neural networks." **Error! Bookmark not defined.**

Machine learning

"One particular form of AI, which gives computers the ability to learn from and improve with experience, without being explicitly programmed. When provided with sufficient data, a machine learning algorithm can learn to make predictions or solve problems, such as identifying objects in pictures or winning at particular games, for example." **Error! Bookmark not defined.**

Supervised Machine Learning

"A type of ML for which the algorithm changes based on data with known labels. In clinical radiology to evaluate medial images, supervised ML is a repetitive process to match images to existing labels." [xi]

Unsupervised Machine Learning

"In supervised ML the algorithm is fed an unlabelled dataset (i.e. without answers). In this case the algorithm groups the image findings into clusters based on one or more features it "learns"." [xi]

Deep learning

"A more recent variation of neural networks, which uses many layers of artificial neurons to solve more difficult problems. Its popularity as a technique increased significantly from the mid-2000s onwards, as it is behind much of the wider interest in AI today. It is often used to classify information from images, text or sound." [x]

Neural network

"Also known as an artificial neural network, this is a type of machine learning loosely inspired by the structure of the human brain. A neural network is composed of simple processing nodes, or 'artificial

neurons', which are connected to one another in layers. Each node will receive data from several nodes 'above' it, and give data to several nodes 'below' it. Nodes attach a 'weight' to the data they receive and attribute a value to that data. If the data does not pass a certain threshold, it is not passed on to another node. The weights and thresholds of the nodes are adjusted when the algorithm is trained until similar data input results in consistent outputs." **Error! Bookmark not defined.Error! Bookmark not defined.**

References

[i] Medical Board of Australia (2014). [Internet] [Cited 2022 March 31]. Available from: https://www.medicalboard.gov.au/Codes-Guidelines-Policies/Code-of-conduct.aspx

[ii] Australian Medical Council (2009). [Internet] [Cited 2022 March 31]. Available from: http://www.amc.org.au/about/good-medical-practice

[iii] New Zealand Medical Council (2016). [Internet] [Cited 2022 March 31]. Available from: https://www.mcnz.org.nz/about-us/publications/good-medical-practice/

[iv] New Zealand Medical Association (2014). [Internet] [Cited 2022 March 31]. Available from: https://www.nzma.org.nz/publications/code-of-ethics

[v] Health and Disability Commissioner (1996). [Internet] [Cited 2022 March 31]. Available from: https://www.hdc.org.nz/your-rights/about-the-code/code-of-health-and-disability-services-consumers-rights/

[vi] RANZCR Code of Ethics. [Internet] [Cited 2022 March 31]. Available from: https://www.ranzcr.com/documents/3958-ethics/file

[vii] ISO (International Organization for Standardisation) and IEC (the International Electrotechnical Commission). [internet] [cited 2022 March 31]. Available from: ISO/IEC DIS 22989(en), Information technology — Artificial intelligence — Artificial intelligence concepts and terminology

[viii] The Organisation for Economic Co-operation and Development (OECD). [internet] [cited 2022 March 31]. Available from: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

[ix] IBM Watson Explainable AI. [internet] [cited 2020 March 31]. Available from: https://www.ibm.com/watson/explainable-ai

[x] Select Committee on Artificial Intelligence. AI in the UK: ready, willing and able? [Internet] [cited 28 March 2019]. Available from: https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf

[xi] J. Raymond Geis, Adrian P. Brady, Carol C. Wu, et al, Ethics of Artificial Intelligence in Radiology: Summary of the Joint European and North American Multisociety Statement. *Canadian Association of Radiologists Journal* 2019 Nov;70(4):329-334