Australian Academy *of* Health and Medical Sciences

ABN 55 167 124 067

Australian
Academy *of* Health and
Medical Sciences

Gabba Towers, 411 Vulture Street, Woolloongabba, QLD 4102
PO Box 6114, Woolloongabba, QLD 4102
tel.: +61 (0) 7 3102 7220 · email: info@aahms.org

RESPONSE TO THE DEPARTMENT OF INDUSTRY, INNOVATION AND SCIENCE CONSULTATION ON THE *ARTIFICIAL INTELLIGENCE: AUSTRALIA'S ETHICS FRAMEWORK*, MAY 2019

Key messages

- The draft provides a comprehensive ethical framework and set of principles, which lay the groundwork for the ethical implementation of AI in Australia. Application of AI-driven technologies in health is one area that demands especially careful consideration. Inappropriate use of these technologies can put patients at risk, jeopardise privacy and deteriorate public trust.
- A fundamental principle of AI in health, as it currently stands, is that it supports augmented decision making – enhancing a function currently undertaken by humans, but those humans remain critical to the development, implementation, monitoring and oversight of the AI.
- The framework would benefit from the inclusion of greater detail on the challenges around patient (or more broadly, user) consent and data sovereignty across all populations. There are unique opportunities and challenges associated with collecting, storing, using and sharing health data, which often demand careful consideration.
- A mechanism for consumer reporting of issues (by patients and clinicians) and empowering authorities to act in a timely manner in the event of suspected breaches will be important.
- We fully support the paper's conclusion that investing in avenues for public feedback and dialogue on AI will be key to developing and implementing valuable technologies.
- The Academy is progressing debate about the use of AI-driven technologies in health and would be delighted to share the developments and outcomes with the consultation team as they emerge.

Introduction

The Australian Academy of Health and Medical Sciences' (AAHMS) is Australia's Learned Academy for health and medical sciences. Our mission is to advance health and medical research in Australia and its translation into benefits for all, by fostering leadership within our sector, providing expert advice to decision makers, and engaging patients and the public. The Academy is the impartial, authoritative, cross-sector voice of health and medical science in Australia. We are an independent, interdisciplinary body of 357 Fellows – elected by their peers for their distinguished achievements and exceptional contributions to health and medical science in Australia.

We welcome the opportunity to comment on the discussion paper on Australia's Artificial Intelligence (AI) Ethics Framework. Our Fellows represent the spectrum of health and medical sciences in Australia, many of them are therefore working in areas of healthcare that are close to implementing AI-driven technologies, or indeed are already doing so. Consequently, our response focuses on the context of health and has been informed by Fellows and Associate Members from a range of medical specialties, who provided insights on ethical challenges faced by the health sector in relation to the core principles outlined in the discussion paper.

AI in health requires careful consideration

AI is already beginning to transform the health sector and with increasing levels of sophistication, AI-driven technologies bring the potential to improve the prediction, prevention, diagnosis and management of diseases – potentially offering precise, efficient and cost-effective solutions. However, as the paper makes clear, challenges remain – in health and other areas – around the safe, ethical and equitable implementation of AI.

The discussion paper presents a comprehensive ethical framework and principles, which lay the groundwork for the ethical implementation of AI in Australia. As noted in the consultation paper, 'AI is a broad family of technologies which requires careful, specialised approaches'.[1] **Health is one area that demands especially careful consideration regarding the ethical and equitable use of AI.** Inappropriate or poorly thought-through use of these technologies can put patients at risk, jeopardise privacy and deteriorate public trust.

The discussion paper specifies that the ethics framework is mainly targeted at business, government and academia, and this breadth is welcome. It is important to note that areas like health may bring extra complexity, since they sit across both the private and public sectors, and through research and training, in fact link to all three. The health sector would benefit from a unified approach to addressing the ethical challenges of AI.

The Academy recognises the pressing need to ensure ethical use and implementation of AI before serious issues arise. We are progressing debate in this domain to inform such specialised approaches. For example, we are convening a roundtable meeting in July 2019 with leaders from academia, industry, health and government to discuss the opportunities, challenges, benefits and risks of AI-driven technologies in health. We would be delighted to engage the Department and Data61 with this work and share its outcomes as they become available. Looking more broadly, we would also highlight the work of our colleagues at the Australian Council of Learned Academies (ACOLA), whose project on the effective and ethical development of AI will provide a broader perspective on this topic.

We note the following points in relation to the principles of the draft framework in the context of health.

Principle 1 – Generates net-benefits

This is a valuable first principle. In health it is important to keep the individual's best interest at the centre of care. This first principle is important for maintaining focus on patients receiving the best possible care and treatment, be it through AI-enabled methods or otherwise. However, it is important to recognise that unexplained variability between individual patients will usually remain, such that benefits can generally only be known at a 'macro' level (at least for the time being while personalised medicine is still evolving). Understanding how to assess net benefits will require careful application of scientific methods that ensure valid causal inference for the effects of interventions.

Principle 2 – Do no harm

This is an important principle and clearly applies to the domain of health. We would highlight the role of education and training here – not only the need for adequate training in using and understanding an AI-driven technology, but **the importance of retaining relevant underpinning education and training**. For example, in health, as AI systems are adopted it is important that clinicians continue to be trained as diagnosticians and expert clinical decision makers who are equipped to maximise the value of the new technology and capable of intervening when system performance is compromised or where system error occurs. This should remain as a fundamental safeguard for patients. More broadly, it will be important to

[1] Consultation paper, page 10 (Dawson D and Schleiger E, Horton J, McLaughlin J, Robinson C, Quezada G, Scowcroft J, and Hajkowicz S (2019) Artificial Intelligence: Australia's Ethics Framework. Data61 CSIRO, Australia).

strengthen our capacity to monitor and evaluate the impact of AI technologies in the health system, including assessing the quality of data and the strength of evidence for benefits in a wide range of potential health-related applications. This will require investment in the data sciences including epidemiology and biostatistics.

Principle 3 – Regulatory and legal compliance

Ethical norms vary between countries and regions. This has potential implications for development and deployment of AI systems in health, particularly systems that influence the distribution of finite health resources. This is well exemplified by the MIT autonomous vehicle study (outside of health, but outlined in the discussion paper), which showed that while there can be broad agreement in moral decision making, there is significant variation in moral preferences that could impact on the local appropriateness of decisions informed by AI. **This underscores the importance of human oversight of AI deployment.**

As an extension of this point, in health the creation of common standards, e.g. in relation to disease classification and record-keeping, can enhance interoperability. For instance, where data is being mined from medical records, it is helpful if those data – whether a diagnosis, treatment or otherwise – are consistently recorded. Other issues include disagreement in the definitions of the phenotype and what constitutes the reference disease label or gold standard classification. Examples like this may result in sub-optimal system performance.

Principle 4 – Privacy protection

The healthcare sector faces a particular set of challenges in regard to privacy protection and the safe use of data, especially given the sensitive nature of the data involved. For instance, medical imaging and genomics are some of the first areas to develop and apply AI-driven technologies, in specialties such as oncology, ophthalmology and dermatology. Resulting ethical challenges include the identification of potentially sensitive personal information through retinal (eye) imaging, skin images and genomic information – and not always in the most obvious ways. For example, some AI systems can classify the gender and the age of an individual on the basis of an eye's retinal image alone. The potential ramifications of identifying features of an image become more significant when these images may also contain information about other aspects of health, such as cardiovascular and central nervous system disease risk. As highlighted in the discussion paper, the re-identification of de-identified data also becomes more likely when multiple data sources are linked, as is often the case for medical data. Linked data can bring significant benefits, such as identifying causes and risks of disease, improve care and reducing costs – careful thought is clearly needed on this complex area.

Patient consent

Consent for the use of data, including images, is another important issue raised in the paper. The potential for data sharing with third parties and potential uses of these data need to be considered in relation to the consent process, including how to achieve genuinely informed consent. How consent is traced when data is passed on is important and potentially complex. The opportunity to revoke consent is important, but the practicalities of implementing this become complex when data, such as an image, has been used to train an AI system. Data sovereignty across all population groups is an important consideration if data is to be shared across geographical and legal boundaries. **The framework would benefit from the inclusion of greater detail on these challenges – on a general level, which could in turn be carried through to more specific areas including health.**

The consultation paper mentions the EU General Data Protection Regulation (GDPR) on several occasions and it is worth noting that this legislation includes clauses that relate specifically to the use of patient and public data for research purposes, including in relation to consent**. We mention this as it highlights the unique opportunities and challenges associated with collecting, storing, using and sharing health data,**

which often demand careful consideration. Informed consent must be gained in a regulatory context that enables the safe and secure use of data for legitimate purposes, while also protecting the rights and interests of individuals.

Principle 5 – Fairness

The discussion paper identifies issues around the potential for bias in AI algorithms and rightly addresses fairness as an important principle to guide the ethical use of AI. In the health sector, the bias of AI can lead to serious discrimination of vulnerable populations and minorities. The paper outlines a selection of case studies to demonstrate this point and additional examples from health could also be added. For example, AI-driven technologies used in medical image analysis in ophthalmology face the issue that many image sets that have been used to train AI systems have been from relatively homogeneous populations. This poses a risk of limited generalisability to other populations and consequently the potential for bias. There is an increasing need for the diversification of training and validation sets. This is an embodiment of the 'fairness' principle of the proposed ethics framework.

Principle 6 – Transparency and explainability

There are concerns among clinicians and patients regarding the lack of transparency regarding the basis for the output of the system – the so called 'black box' problem. Visualisation strategies can assist in gaining trust and aid transparency in the implementation of AI-technologies. Medicine is increasingly complex and there are already many barriers to patients understanding their condition(s) and associated management/treatment; the added complexity of comprehending the role of AI in diagnosis and clinical decision-making is not trivial in this context. Further complexity is added when consumer-facing AI technologies are used by individuals for the pre-assessment of medical symptoms e.g. chatbots. These AI-driven applications can for instance be used by the consumer to identify whether they should consult a health professional. Depending on the symptoms and associated condition, the recommendation made by the AI-driven application could have life-changing consequences. Since the AI's assessment is heavily data driven, recommendations linked to symptoms which are associated with more common conditions would potentially be more accurate than those for rare diseases where less data are available. For transparency it is therefore important to consider whether patients should be made aware of the accuracy of the AI's recommendation, i.e. the confidence level of the assessment. It would also have to be considered whether information on the data sets and sample size used for machine learning should be made available to consumers, particularly for rare conditions with less available data and lower sample sizes. This also relates to the points about contestability in the next section.

To assure explainability, the communication and interaction between patients and health professionals remains vital. In particular, the need for clinician and patient education and engagement is important if AI is to be adopted in health in a meaningful and ethical manner (we expand on this point below). This is addressed by the 'transparency and explainability' principles of the ethics framework and by the 'contestability' principle – as understanding is a key component of the latter.

Principle 7 – Contestability

In health, an example of where this principle might apply is where the clinical classification generated by the AI system differs from the human expert classification, and this will be particularly challenging where an AI system has been demonstrated to perform at a high(er) overall level of accuracy. A mechanism to document and appraise discrepancies between human expert and AI system classifications/decisions will be important in defining system constraints, driving system improvement and identifying the limitations and biases of both human experts and AI systems. The manner in which this will be achieved in practice is a matter for further consideration. This is an extension of the concept of 'contestability' outlined in the proposed ethical framework. As highlighted in the discussion paper, expert human oversight should guide the implementation of AI system outputs. Medicolegal ramifications of the implementation of AI need to be mapped out – for instance in cases in which an adverse clinical outcome arises from reliance on AI output, or when AI output was overruled by a clinician. This also relates to the next principle around accountability.

Principle 8 – Accountability

Prediction, prevention, diagnosis and management are constantly changing as we grow our understanding of health and disease. Medicine is complex, and individuals increasingly suffer from multiple conditions, the management of which can be very challenging. This evolving landscape and increasing complexity means that mistakes can unfortunately occur, the consequences of which can be highly significant for the individual. We agree that monitoring system performance and the timely reporting of adverse outcomes are critical to minimising the potential harms of AI systems and building trust. This may involve a mechanism via existing regulators (e.g. the Australian Health Practitioner Regulation Agency or Therapeutic Goods Administration), professional bodies or new authorities – and thought needs to be given to the extent to which different fields (e.g. health, transportation and so on) should be aligned in this regard.

As highlighted in the discussion paper, accountability for error is also important. AI-driven solutions need to fit the legal, ethical boundaries and standards of Australian society and law. The notion of extending the attribution of accountability to the developers of AI systems and those involved in their implementation is helpful, however understanding how this will work in practice is less clear in particular where technologies have been developed by multinational companies and then implemented in Australia. The paper recognises that overregulation poses the potential risk of stifling innovation or driving smaller companies out of the field if the risk is such that only large technology companies can manage this exposure.

It is also important to note that medicolegal systems can vary significantly in different regions across the globe. Enforcement of regulatory breaches and unethical AI systems is likely to be a major challenge in practice. We welcome the inclusion in the draft framework of a system of risk assessments, system tests and reporting mechanisms as a means to promote ethical and fair AI use. Nevertheless, challenges are likely to arise from the rapid proliferation of many different AI systems in health. **A mechanism for consumer reporting (patient and clinician) and empowering authorities to act in a timely manner in the event of suspected breaches will be important.**

Consultation

We welcome the inclusion of public consultation in the paper and agree that public support will be critical to the successful deployment of AI-driven technologies, including in health. **We fully support the paper's conclusion that investing in avenues for public feedback and dialogue on AI will be key.** The mechanisms used for these purposes must enable meaningful conversations in order to provide valuable insights, and appropriate investment will be critical to success here.

Conclusion

The ethical principles outlined in the discussion paper clearly apply to the use of AI in health and would lay the foundation for an additional framework to address the ethics of AI in this context. Although the use of AI in various medical disciplines may differ, targeted ethical principles will be crucial in guiding the safe, effective and equitable use of AI in health.

We are grateful to the Fellows and Associate Members who contributed to this response.

For further information about this response, please contact Katrin Forslund, Policy and Projects Officer at the Australian Academy of Health and Medical Sciences: katrin.forslund@aahms.org.