


Perspective

A governance model for the application of AI in health care

Sandeep Reddy ¹, Sonia Allan,² Simon Coghlan,³ and Paul Cooper¹

¹School of Medicine, Geelong, Deakin University, Australia, ²Deakin Law School, Melbourne, Deakin University, Australia, and

³School of Computing and Information Systems, University of Melbourne, Melbourne, Australia

Corresponding Author: Sandeep Reddy, MBBS, PhD, School of Medicine, Deakin University, 75 Pigdons Road, Waurn Ponds VIC 3216, Australia; sandeep.reddy@deakin.edu.au

Received 29 July 2019; Revised 24 September 2019; Editorial Decision 7 October 2019; Accepted 10 October 2019

ABSTRACT

As the efficacy of artificial intelligence (AI) in improving aspects of healthcare delivery is increasingly becoming evident, it becomes likely that AI will be incorporated in routine clinical care in the near future. This promise has led to growing focus and investment in AI medical applications both from governmental organizations and technological companies. However, concern has been expressed about the ethical and regulatory aspects of the application of AI in health care. These concerns include the possibility of biases, lack of transparency with certain AI algorithms, privacy concerns with the data used for training AI models, and safety and liability issues with AI application in clinical environments. While there has been extensive discussion about the ethics of AI in health care, there has been little dialogue or recommendations as to how to practically address these concerns in health care. In this article, we propose a governance model that aims to not only address the ethical and regulatory issues that arise out of the application of AI in health care, but also stimulate further discussion about governance of AI in health care.

Key words: artificial intelligence, healthcare, ethics, regulation, governance framework

INTRODUCTION

Interest in AI has gone through cyclical phases of expectation and disappointment since the late 1950s because of poor-performing algorithms and computing infrastructure.¹ However, the emergence of appropriate computing infrastructure, big data, and deep learning algorithms has reinvigorated interest in artificial intelligence (AI) technology and accelerated its adoption in various sectors.² While recent approaches to AI, such as machine learning, have only been relatively recently applied to health care, the future looks promising because of the likelihood of improved healthcare outcomes.^{3,4} With deep learning algorithms (eg, deep neural networks) meeting, and in some cases surpassing, the performance of clinicians, the promise is already apparent.¹ AI is positioned to have a major role in a range of healthcare delivery areas, including diagnostics, prognosis, and patient management.² However, substantial challenges, not least ethical and regulatory concerns,⁵ could present a barrier to the entry

and use of AI in health care. A single major mishap with a clinical AI system could undermine public and health professional confidence. Therefore, addressing those concerns is a priority.^{5,6} In this article, we elaborate these concerns and propose a governance model to mitigate these risks.

ETHICAL CONCERNS

The successful implementation of AI in healthcare delivery faces ethical challenges.⁷ Three key challenges are potential biases in AI models, protection of patient privacy, and gaining the trust of clinicians and the general public in the use of AI in health care.³ In addition, the ethical integrity and public role of the health professions relies on maintaining broad public trust. The success of AI in health care, and the integrity and reputation of health professions that use AI, depends on meeting these ethical challenges. We outline the previous

3 key ethical challenges in this section and discuss the above ethical principles in the section below on governance.

AI bias

The training of AI models requires large-scale input of health-related or other data.⁴ The computer science adage, “garbage in, garbage out,”⁸ can be restated in the context of AI model training as “biases in, biases out.” Such biases can arise when data used for training are not representative of the target population and when inadequate or incomplete data are used for training the AI models.⁸ Unrepresentative data can occur due to, for example, societal discrimination (eg, poor access to health care) and relatively small samples (eg, minority groups). Unrepresentative data can entrench or exacerbate health disparities. Some AI models deployed in non-healthcare domains have demonstrated biases, such as overestimating risks of criminal recidivism among members of certain racial group.⁹ In health care, biased algorithms may lead to underestimation or overestimation of risks in certain patient populations. Of course, the notion of bias is complex, and humans too have biases. But it may be possible, and hence ethically necessary, to design AI systems that help offset human biases and so lead to fairer (if still imperfect) outcomes.¹⁰ Reducing AI bias is thus necessary for promoting both better and more equitable health outcomes.

Privacy

Healthcare data are some of the most sensitive information one can hold about a person.^{8,10} Respecting a person’s privacy is a vital ethical principle in health care because privacy is bound up with patient autonomy or self-rule, personal identity, and well-being.^{5,10} For these reasons, it is ethically essential to respect patient confidentiality and ensure adequate processes for obtaining genuine informed consent from patients both for health interventions and for the usage of their personal health data. AI systems should be protected from privacy breaches to prevent psychological and reputational harm to patients, and patients must provide explicit consent for their data to be used for any specific use.¹¹ The system should be protected from breaches to prevent psychological and reputational harm to patients. It is an expectation that patients must provide explicit consent for their data if their data are shared. However, recent episodes like Cambridge Analytica using personal data collected by Facebook for political advertising¹¹ and the Royal Free London NHS Foundation trust sharing patient data for the development of a clinical application without explicit patient consent¹² present concerns about privacy breaches. Also, increasingly there is concern that anonymized data can be reidentified with few spatiotemporal datapoints. Any such reidentification can breach the trust of patients. Further, method of data collection for AI model training can raise concerns. As mentioned previously, current AI models, particularly deep learning models, require large datasets for high-quality performance.² Apart from the requirement for swathes of potentially sensitive patient information, a potential exists for patient data to be collected without patients being aware of its final usage. For example, AI devices used to support older adults in their homes may collect and transmit data without their knowledge, and health services may supply patient data to AI developers without the informed consent of patients. In some countries, lax rules may permit forms of data collection that promote breaches of privacy.⁸

Patient and clinician trust

Effective health care is predicated on the maintenance of substantial trust between the public and health professions and systems.^{8,10,11}

Professional bodies around the world rightly insist that clinicians have an ethical duty to safeguard and promote patient trust. Trust in clinicians encompasses trust in the clinical tools they choose to use, and in the selection of those tools, including AI-based tools. Because of the nature of AI algorithms, especially deep learning algorithms, a lack of transparency in decision making can result from the use of such tools that may threaten patient trust. The nature of AI algorithms, especially deep learning algorithms, can facilitate a lack of transparency in decision making.³ Deep learning algorithms continuously fine-tune their parameters and evolve rules. This can lead to opaque decision-making processes, hidden even to developers—a situation known as the black-box issue.⁸ This black-box situation can present challenges in validating the outputs of the AI models, guaranteeing safety in unusual input situations, and identifying biases in the data.³ In health care, where transparency in clinical decision making and disclosure to patients of relevant information is paramount, the lack of algorithmic transparency presents particularly acute concerns. The black-box situation also makes it harder to determine if an adversarial attack¹⁰ has taken place (ie, some malicious manipulation of an AI model’s outcome through feeding special cases into it).

Clinicians who cannot understand the inner workings of the model will be unable to explain the medical treatment process to their patients.⁸ Equally, as AI’s predictive and diagnostic ability improves, clinicians may become ever more reliant on AI models; at the limit, decision making itself could become automated. Overreliance on AI models may reduce or eliminate the contact and conversation between healthcare professionals and patients,⁸ which underpins good patient care and respect for patient autonomy. In sum, reduced transparency in decision making, plus the other concerns we have identified in AI models, could engender among the healthcare professionals and the wider public a lack of trust—trust that is so vital to effective health care.

REGULATORY CONCERNS

AI software or devices augmented by AI software have an ability to autolearn from real-world use and can thereby improve in performance over time.¹³ This distinguishes AI software from other software used in health care and presents novel regulatory challenges. It is an objective of regulatory authorities, health services, and clinicians that safe and quality health care be delivered to patients. Algorithms that are unexplainable in their decision making, change continuously with use, and autoupdate, perhaps with features that go beyond the initial approved clinical trials, may require special policies and guidelines.^{8,13} Concerns also emerge about the safety and efficacy of AI medical software that does not necessarily align with current models of care delivery.¹⁴ Regulatory standards to assess AI algorithmic safety and impact are yet to be formalized in many countries.^{1,15} This can both present barriers to entry of AI in health care and enable unsafe practices in which AI is already being used in health care.

Issues of liability are also of concern: for example, there is the question of who is responsible when errors result from the use of AI software or AI-augmented devices in the clinical context. Current medicolegal guidelines, across the world, are also unclear regarding where the lines of responsibility begin or end when AI agents guide clinical care.⁷ The lack of explainability affecting some algorithms, and the fact that treatment strategies are generally less effective in routine clinical practice than in the preliminary assessment, adds to regulatory complexity. A further concern may arise when clinicians dismiss appropriate AI-recommended treatment strategies because

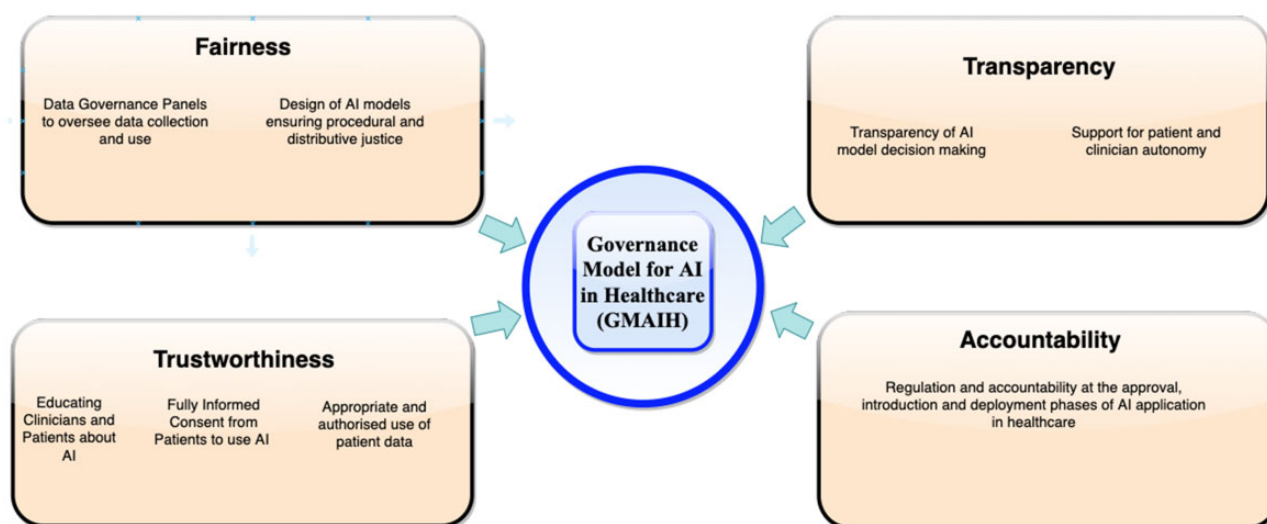


Figure 1. Outline of the Governance Model for Artificial Intelligence (AI) in Health Care.

of lack of trust in the AI agent.¹⁶ What the implications will be for medical malpractice in the context of dominant AI-driven diagnostics is yet to be seen.¹⁷

IMPLICATIONS FOR HEALTH CARE

Given the access many countries have to infrastructure that can run AI software, the speed of investment in AI, the fast pace at which AI-based applications can be developed, and the countless opportunities AI presents for health care, it is becoming increasingly evident that it is not a question of “if” but “when” AI will become part of routine clinical care.^{1,2,7,13,17} Clinical use of AI models is certain to transform current models of healthcare delivery; indeed, their reach will extend beyond clinical settings.¹⁸ AI has an ability to overcome limitations with traditional rules-based clinical decision support systems and to enable better diagnostic and decision support.¹⁹ Opportunities to automate triage and screen and administer treatment are also becoming a reality. AI embedded in smart devices, supported by the Internet of things and fast Wi-Fi, could bring AI-enabled health services into the homes of patients, thus democratizing health care.^{1,8} However, some concerns must be emphasized. In the absence of appropriate regulatory and accreditation systems, rapid progress in development and deployment of AI models could lead to unsafe and morally flawed practices in health care. So far, relatively little attention has been paid to this aspect. Consequently, it is imperative to explore governance models for the use of AI in health delivery.

GOVERNANCE MODEL

To address the aforementioned ethical, regulatory and safety and quality concerns, we propose a governance model for AI application in health care. The model we present is termed Governance Model for AI in Healthcare (GMAIH). The 4 main components of the proposed governance model are fairness, transparency, trustworthiness and accountability (Figure 1).

Fairness

Data in the health context may include (but not be limited to) medical images, text from patient records about medical conditions,

diagnosis and treatment, and reimbursement codes.¹ As discussed, inappropriate and poorly representative training datasets for AI models can lead to biases, inaccurate predictions, medical errors, and even large scale discrimination.^{3,5,8} Therefore, we recommend a data governance panel constituted by the AI developers that includes patient and target group representatives, clinical experts, and people with relevant AI, ethical, and legal expertise. The panel would review datasets used for training AI to ensure the data is representative and sufficient to inform requisite model outcomes. This initiative is akin to co-design of research and service provision through the involvement of patient and public representatives and healthcare professionals.^{20,21} The panel would work to achieve a clearly articulated data collection and utilization strategy that will guide documentation, workflow, a review of influencing factors and monitoring standards. The panel’s remit would also be to review algorithms—noting that data and algorithms go together in developing AI models.¹

Normative standards for the application of AI in health care should be developed by governmental bodies and healthcare institutions as part of governance. These standards should inform how AI models will be designed and deployed in the healthcare context and should conform to the requirements of one of the classic biomedical ethical principles, namely justice.²² The principle of justice includes fairness in access to health care. Accordingly, AI applications should not lead to, or exacerbate, discrimination, disparity, or health inequities. The design should ensure procedural (fair process) and distributive justice (fair allocation of resources) is abided by, with a view to protect against adversarial attack or the introduction of biases or errors through self-learning or malicious intent.

Transparency

While the performance of deep learning models in medical imaging analysis and clinical risk prediction has been exceptionally promising, the models are also hard to interpret and explain.² This poses a particular problem in medicine, where transparency and explainability of clinical decision is paramount.^{3,8} In fact, this issue has been cited as the single most significant difficulty for acceptance, regulation, and deployment of AI in health care.⁸ Limited transparency can reduce trustworthiness of AI models in health care. Limited

transparency also impairs validation of the clinical recommendations of the model and identification of any errors or biases.³ Earlier AI models used in medicine were logical and symbolic based.²³ While they lacked the accuracy and predictive powers of current algorithmic models, those earlier models offered a trace of their decision steps. In contrast, there are limits to the explainability of current models such as deep learning AI.

To address this issue at a general level, a field termed explainable AI (XAI) has emerged.²⁴ The intention of XAI is to enable a set of techniques that allow explainability while maintaining high performance. While it is beyond the scope of this article to discuss individual techniques, we will mention that the focus of XAI techniques in medicine relates to the functional understanding of the model as opposed to low-level algorithmic understanding of the model.²³ That understanding can be targeted at a global level (understanding of the whole logic of a model) or local level (explaining the reasoning for a specific decision or prediction).²³ Whereas these measures relate to addressing the explainability drawbacks of deep learning models, there have also been suggestions for using AI algorithms that are explainable in the context of medicine.²⁵ Although these explainable algorithms have less accuracy and predictive performance, they lend themselves to greater interpretability, which is crucial in medicine. It is also important that AI agents designed to have human appearance in voice or visual look do not deceive humans (ie, they should introduce themselves as AI agents).

Sufficient transparency and explainability is demanded by the classic ethical principle of respect for autonomy.²² Autonomy can be understood as self-rule, which in the health context implies the freedom and ability of patients to make decisions in accordance with their preferences and values. AI agents must therefore support rather than diminish the provision of a level of transparent understanding sufficient to meet patients' individual requirements for decision making. They must also allow patients the freedom to make health-related decisions without coercion or undue pressure.

Based on these considerations, we propose through our governance model an emphasis on ongoing or continual explainability. Where deep learning or other AI models (which have explainability issues) are deemed to be necessary, under this governance model interpretable frameworks are expected to be utilized to enhance the decision-making process. Lately, several medical studies have showcased how this is possible with the use of explainable tools, ranging from visual to direct measurement tools.^{26–28}

Trustworthiness

It is important for clinicians to understand the causality of medical conditions, and in the case of AI, the methods and models employed to support the clinician decision-making process.²³ In addition to the explainability issues discussed in the previous section, the potential autonomous functioning of AI applications and potential vulnerability of these applications to being accidentally or maliciously tampered with to yield unsafe results may present major hindrances for clinicians in accepting AI in their clinical practice.^{1,19} Also, recent episodes of hospitals sharing patient data with AI developers without the patients' informed consent has added to the problem of trusting AI developers and AI itself.^{12,29} This has been further compounded by the ability of AI agents to collect and learn from data in real-world settings¹³ and certain AI applications overpromising and underdelivering on clinical outcomes in the recent period.³⁰ To address these issues, we propose through our governance model a multipronged approach that includes technical education, health literacy, full informed consent, and clinical audits.

Admittedly, understanding the full spectrum of AI, including its relevant mathematics and programming, takes time. Nevertheless, there have been recommendations and initiatives to educate healthcare professionals about the basics of AI (ie, techniques, application, and impact).^{31,32} We believe these initiatives are a vital element in building trust for AI among healthcare professionals. By understanding how AI works, and what advantages and limitations it has in healthcare delivery, clinicians will very likely be more accepting of AI. Crucially, this approach would enable clinicians to be partners in the control of the technology, rather than merely being passive recipients of the AI outputs.

In addition, education should extend to the patient community and public. We recommend an education approach that adopts principles of health literacy, to ensure patients receive the information they need to make informed and autonomous health choices.³³ To enable such education (of both health professionals and the patient community), we recommend partnerships between academic institutions and health services, thereby ensuring complementary use of skills in AI technology, pedagogy, healthcare policy, and clinical practice. The base education content can be repurposed to suit different audiences and adapted as AI technology and its application evolve.

We also recommend that institutional policies and guidelines be reworked to ensure patients are aware that the treating clinician is drawing support from AI applications, what the limitations of the applications are, and that the patients are in a position, where relevant, to refuse treatment involving AI.³⁴ Where patient data may be shared with AI developers, there must be a process to seek fully informed consent from patients and if it is unrealistic to seek approval, data must be anonymized to that extent individual patient details cannot be recognized by the developers.³⁵ The permissions to provide data should be rescindable. Also, differential privacy, a technological solution, which minimizes the risks of analyzing confidential and sensitive data should be considered.³⁶ Through this approach, a high standard of data anonymization is achieved by shrinking the risks associated with reidentification, thus upholding privacy of patients.

Further, we recommend, where possible, the use of public datasets to develop AI software to minimize privacy breaches. There should be clear clinical objectives associated with AI applications and the veracity of the claims made by AI developers should be tested. Professional medical bodies have a role in issuing clinical guidelines regarding where AI applications can be utilized in the diagnosis and treatment process (see also the following section). Such guidelines would increase not only the confidence of physicians using AI, but also their trust in AI applications. It would also respect the autonomy of patients.

Accountability

Accountability, the fourth and final component of our governance proposal, commences with the development of the AI model and extends to the point the model is applied in clinical care and finally retired. This spectrum involves a number of players including software developers, government agencies, health services, medical professional bodies, and patient interest groups, among others. Therefore, we consider the accountability component as the most challenging of the governance components to implement. So how do we frame accountability for such a diverse range of players and situations? We recommend identifying appropriate stages for which monitoring and evaluation is critical to ensure the safety and quality of AI-enabled services. These stages include approval, introduction, and deployment.

Approval stage

For the approval stage, which covers permission for the marketing and use of AI in healthcare delivery, governmental bodies or regulatory authorities have an important role. In the United States, the Food and Drug Administration (FDA), which regulates medications and medical devices, has introduced steps to approve software for medical use.¹³ The FDA terms such software as software as a medical device (SaMD).³⁷ As part of the SaMD risk categorization and premarket approval, several AI-based SaMD have been approved for use in healthcare delivery.³⁷ In addition to the current process of risk review and premarket approval of AI-based SaMD, the FDA is mulling a “predetermined change control plan” to anticipate changes in the AI algorithm after market introduction.¹³ This means that when the AI software has a significant medication that affects the safety or effectiveness, the developer would have to revert to the FDA for review and approval. We consider the FDA approach both in terms of current and proposed review and approval processes forward thinking and commendable. The FDA adopts a balanced approach toward ensuring the safety and quality of AI-based SaMD, while not creating unnecessary barriers for AI developers to introduce SaMD to the market. The FDA process could be similarly adopted by respective regulatory agencies across the world. In countries that do not have established regulatory processes for evaluation and monitoring of SaMD,³⁸ there is a role for international bodies (eg, the World Health Organization) to guide and support relevant countries to adopt appropriate processes to regulate SaMD.

Introduction stage

The introduction stage involves health services reviewing AI products in the market, assessing them for their suitability in their healthcare delivery and establishing relevant policies and procedures to allow for incorporation of AI software in clinical care. It is often that health information technology products fall short of expectations and is indeed the case with AI models in recent history.³⁰ AI models need to be reviewed for their data protection, transparency, and bias minimization features in addition to safety and quality risks and protections against malicious attack or inadvertent errors.^{8,19} Health services can constitute or use existing panels to review alignment of the AI models with their specific clinical or health service needs. However, the rapid progression in AI technology and varied techniques means that not all panels would have the capacity to make the assessment of AI products on their own. It has been proposed that a benchmarking system that scrutinizes the performance and robustness of AI medical software be available to guide health services.¹ The benchmarking system could be a result of public-private partnerships. The benchmarking platform would allow for comparison of different AI models through a dashboard of performance metrics. These benchmarking platforms can guide individual health services about their choices.

Deployment stage

The deployment stage takes into account liability, monitoring, and reporting factors. If we expect AI models to incorporate ethical principles, it is also pertinent to assess and hold the models accountable in deployment.³⁹ Use of AI in clinical care and the potential liability issues that may emerge are complex and filled with many uncertainties.^{16,39} The use of AI software for clinical practice risks increased liability for clinicians and health services.¹⁶ The issue of who becomes responsible when safety and quality issues arise because of the use of AI medical software necessitates appropriate

legal guidance. Current medical malpractice or negligence laws may not be able to accommodate this scenario, and remain untested in, if not ill-suited to, the context of use of autonomous or semiautonomous medical software.³⁹

Of course, any legislative change to address such issues should not be at the cost of innovation and should not preclude the use of AI models in clinical care. A balanced approach in which the safety of patients, autonomy of clinicians, and clinical decision support derived from AI models is required. We recommend a responsive approach to regulation that allows for ongoing monitoring of safety and risk of the AI models in clinical practice, which should include regular audits and reporting. Audits could test the model's bias, accuracy, predictability, transparency of decision making, and achievement of clinical outcomes. The same measures could be considered for reporting. We also recommend drawing on the TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis) model as guidance for constituting the reporting framework.⁴⁰ The TRIPOD model is a checklist of 22 items considered important for transparent reporting of predictive models including model specification, performance and validation. In addition, the GMAIH model suggests the accountability and reporting process to mirror the strategy recommended by authors Halligan and Donaldson⁴¹ for implementing clinical governance, which covers composition of national standards to be used by health services to assure safety and quality, local clinical governance models, annual appraisal of AI model performance, site visits, learning mechanisms including adverse event reporting, incorporation of patient views, and education and training of clinicians and patients.

Integration

While the preceding discussion focused on the governance model itself, a very important consideration is how the GMAIH model integrates into clinical workflow. Clinical workflow is represented in the routine tasks performed by clinicians and the results generated by it.⁴² These include administrative tasks such as appointment scheduling and billing and clinical tasks such as medical treatment and patient education. To ensure that AI applications yield necessary value to the clinicians and patients, they have to be integrated into clinical workflow. The steps to integrate AI application are outlined in Figure 2. The GMAIH model interplays with the integration at critical steps by ensuring that applications generate appropriate data, there is transparency in decision making, clinicians' and patients' views are considered in the integration, and there is accountability of the applications through inspections and reporting.

To support the integration and governance, we recommend that governance be provided by a clinical governance committee formulated with specific skills and experience to oversee the introduction and deployment of AI models in clinical care. An appropriate governance committee should include clinicians, managers, patient group representatives, and technical and ethics experts so that appropriate deliberations are held about the efficacy and effectiveness of the AI models in addition to oversight of privacy, safety, quality, and ethical factors. Such a governance body should also ensure that an appropriately resourced team and plan is in place to monitor for data drift, input-output variation, unexpected outcomes, data reidentification risk, and clinical practice impacts. These efforts should be reported back up to the clinical owner and it should be the responsibility of the governance to enforce. As with fairness and transparency, the governance components of trustworthiness and accountability in the design and deployment of AI are essential for



Figure 2. Integration of governance in the clinical workflow. AI: artificial intelligence; EHR: electronic health record.

ensuring trust in health care, and for safeguarding the fiduciary relationship between practitioners and patients. In turn, such trust is necessary for meeting the moral demands of the remaining 2 classic ethical principles, namely nonmaleficence and beneficence.²² Ensuring that patients (and the wider public) are not harmed by AI and machine learning, and are, moreover, benefited more by their presence than by their absence, are pivotal reasons for our governance recommendations.

CONCLUSION

While there is some way to go before AI models become a regular feature of healthcare delivery, the path for their use has been already set. AI medical products are already on the market and there is increasing evidence of the efficacy of AI medical software in clinical decision making.^{1,37} Despite some discussion of the morality of AI in health care, very few investigations have moved beyond the ethics to consider the legal and governance aspects. To address this gap, we proposed a governance model that covers the introduction and implementation of AI models in health care. Our model by no means purports to cover every eventuality that may emerge due from the

use of AI in healthcare delivery. Nonetheless, by incorporating basic elements essential to the safe and ethically responsive use of AI in health care, it is designed to be flexible enough to accommodate changes in AI technology. Clearly, a wider discussion about the regulation of AI in health care is needed, a discussion we hope to trigger through our recommendations for a governance framework.

FUNDING

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

AUTHOR CONTRIBUTIONS

SR conceived the early version of the governance framework and the manuscript including figures. SA, SC, and PC then reviewed the manuscript for accuracy, relevance, and grammar. SR then revised and finalized the manuscript.

CONFLICT OF INTEREST STATEMENT

None declared.

REFERENCES

- Salathé M, Wiegand T, Wenzel M. *Focus Group on Artificial Intelligence for Health*. Geneva, Switzerland: World Health Organization; 2018.
- Senate of Canada. Challenge ahead-integrating robotics, AI and 3D printing technologies into Canada's Healthcare Systems. 2017. https://senca-nada.ca/content/sen/committee/421/SOCI/reports/RoboticsAI3DFinal_Web_e.pdf Accessed June 28, 2019.
- Whittlestone J, Nyrop R, Alexandrova A, Dihal K, Cave S. Ethical and societal implications of algorithms, data, and artificial intelligence: a road-map for research. 2019. <http://www.nuffieldfoundation.org/sites/default/files/files/Ethical-and-Societal-Implications-of-Data-and-AI-report-Nuffield-Foundat.pdf> Accessed July 1, 2019.
- National Health Service. *Accelerating AI in Health and Care: Results from a State of the Nation Survey*. London, United Kingdom: Department of Health and Social Service; 2018.
- Cath C. Governing artificial intelligence: ethical, legal and technical opportunities and challenges. *Philos Trans A Math Phys Eng Sci* 2018; 376 (2133): 20180080.
- Cheatham B, Javanmardian K, Samandari H. Confronting the risks of artificial intelligence. *McKinsey Quarterly*. April 2019. <https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/confronting-the-risks-of-artificial-intelligence#> Accessed July 1, 2019.
- Reddy S, Fox J, Purohit MP. Artificial intelligence-enabled healthcare delivery. *J R Soc Med* 2019; 112 (1): 22–8.
- Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: addressing ethical challenges. *PLoS Med* 2018; 15 (11): e1002689.
- Angwin J, Larson J, Mattu S, Kirchner L. Machine Bias. *ProPublica*. 2016. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> Accessed October 8, 2019.
- Char DS, Shah NH, Magnus D. Implementing machine learning in healthcare-addressing ethical challenges. *N Engl J Med* 2018; 378 (11): 981–3.
- Dawson D, Schlieger E, Horton J, et al. *Artificial Intelligence: Australia's Ethics Framework*. Data61 CSIRO, Australia; 2019. https://consult.industry.gov.au/strategic-policy/artificial-intelligence-ethics-framework/suppor-ting_documents/ArtificialIntelligenceethicsframeworkdiscussionpa-per.pdf Accessed July 1, 2019.
- Powles J, Hodson H. Google DeepMind and healthcare in an age of algorithms. *Health Technol* 2017; 7 (4): 351–67.
- Food and Drug Administration. Proposed regulatory framework for modifications to artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD)-discussion paper. <https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm514737.pdf> Accessed July 1, 2019.
- Parikh RB, Obermeyer Z, Navathe AS. Regulation of predictive analytics in medicine. *Science* 2019; 363 (6429): 810–2.
- World Health Organization. Legal frameworks for eHealth. In: *Global Observatory for eHealth series*. Geneva: World Health Organization; 2011: 5.
- Luxton DD. Should Watson be consulted for a second opinion? *AMA J Ethics* 2019; 21 (2): E131–7.
- Froomkin AM, Kerr I, Pineau J. When AIs outperform doctors: confronting the challenges of a tort-induced over-reliance on machine learning. *Ariz Law Rev* 2019; 61 (33): 33–99.
- Loukides M. The ethics of artificial intelligence. 2016. https://www.oreilly.com/ideas/the-ethics-of-artificial-intelligence?imm_mid=0ea9bf&cmp=em-data-na-na-newsletter_ai_20161114 Accessed July 23, 2019.
- Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K. Artificial intelligence, bias and clinical safety. *BMJ Qual Saf* 2019; 28 (3): 231–7.
- Gustavsson SM, Andersson T. Patient involvement 2.0: experience-based co-design supported by action research. *Action Res* 2017 Aug 7.
- Scott J, Heavey E, Waring J, Jones D, Dawson P. Healthcare professional and patient codesign and validation of a mechanism for service users to feedback patient safety experiences following a care transfer: a qualitative study. *BMJ Open* 2016; 6 (7): e011222.
- Gillon R. Four principles plus attention to scope. *BMJ* 1994; 309 (6948): 184–8.
- Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. *Data Min Knowl Discov* 2019; 9 (4): e1312.
- Adadi A, Berrada M. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access* 2018; 6: 52138–60.
- National Science and Technology Council. The National Artificial Research and Development Strategic Plan: 2019 update. 2019. <https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf> Accessed September 4, 2019.
- Lundberg SM, Erion G, Chen H, et al. Explainable AI for trees: from local explanations to global understanding. *arXiv* 2019 May 11.
- Lamy JB, Sekar B, Guezennec G, Bouaud J, Séroussi B. Explainable artificial intelligence for breast cancer: a visual case-based reasoning approach. *Artif Intell Med* 2019; 94: 42–53.
- Lee H, Yune S, Mansouri M, et al. An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets. *Nat Biomed Eng* 2019; 3 (3): 173–82.
- Wakabayashi D. Google and the University of Chicago are sued over data sharing. *The New York Times*. June 26, 2019.
- Strickland E. How IBM Watson overpromised and underdelivered on AI health care. *IEEE Spectrum*. April 2019. <https://spectrum.ieee.org/bio-medical/diagnostics/how-ibm-watson-overpromised-and-underdelivered-on-ai-health-care> Accessed July 7, 2019.
- Harvard University Laboratory of Medical Imaging and Computation. Artificial Intelligence in Healthcare Accelerated Program. 2019. <http://aihap.mgh.harvard.edu/program-info/> Accessed July 23, 2019.
- Kolachalama VB, Garg PS. Machine learning and medical education. *NPJ Digit Med* 2018; 1 (1): 54.
- Nielsen-Bohlman L, Panzer AM, David A. Health literacy: a prescription to end confusion. *Choice Rev Online* 2013; 42: 4059.
- Schiff D, Borenstein J. How should clinicians communicate with patients about the roles of artificially intelligent team members? *AMA J Ethics* 2019; 21 (2): 138–45.
- Jones ML, Kaufman E, Edenberg E. AI and the ethics of automating consent. *IEEE Secur Privacy* 2018; 16 (3): 64–72.
- Adjekum A, Ienca M, Vayena E. What is trust? Ethics and risk governance in precision medicine and predictive analytics. *OMICS* 2017; 21 (12): 704–10.
- Blake K, Hickman J, Huang E, et al. Current state and near-term priorities for AI-enabled diagnostic support software in health care. 2019. <https://healthpolicy.duke.edu/sites/default/files/atoms/files/dukemargolisaiena-bledxss.pdf> Accessed July 1, 2019.
- Lamph S. Regulation of medical devices outside the European Union. *J R Soc Med* 2012; 105 (Suppl 1): 12–21.
- Lupton M. Some ethical and legal consequences of the application of artificial intelligence in the field of medicine. *Trends Med* 2018; 18 (4): 100147.
- Collins GS, Reitsma JB, Altman DG, Moons K. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *Eur Urol* 2015; 67 (6): 1142–51.
- Halligan A, Donaldson L. Implementing clinical governance: turning vision into reality. *BMJ* 2001; 322 (7299): 1413–7.
- Bowens FM, Frye PA, Jones WA. Health information technology: integration of clinical workflow into meaningful use of electronic health records. *Perspect Health Inf Manag* 2010; 7: 1d.