Journal of the American Medical Informatics Association, 28(7), 2021, 1582–1590
doi: 10.1093/jamia/ocab065
Advance Access Publication Date: 25 April 2021
Perspective

Perspective

Enhancing trust in AI through industry self-governance

Joachim Roski,[1] Ezekiel J. Maier,[1] Kevin Vigilante,[1] Elizabeth A. Kane,[1] and Michael E. Matheny[2,3]

[1]Booz Allen Hamilton, Washington, DC, USA, [2]Departments of Biomedical Informatics, Biostatistics, and Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA, and [3]Geriatric Research Education and Clinical Care Center, Tennessee Valley Healthcare System VA, Nashville, Tennessee, USA

Corresponding Author: Joachim Roski, PhD, MPH, Booz Allen Hamilton, 2941 Fairview Park Drive, Falls Church, VA 22042, USA (Roski_Joachim@bah.com)

Received 23 October 2020; Revised 17 March 2020; Editorial Decision 18 March 2021; Accepted 26 March 2021

ABSTRACT

Artificial intelligence (AI) is critical to harnessing value from exponentially growing health and healthcare data. Expectations are high for AI solutions to effectively address current health challenges. However, there have been prior periods of enthusiasm for AI followed by periods of disillusionment, reduced investments, and progress, known as "AI Winters." We are now at risk of another AI Winter in health/healthcare due to increasing publicity of AI solutions that are not representing touted breakthroughs, and thereby decreasing trust of users in AI. In this article, we first highlight recently published literature on AI risks and mitigation strategies that would be relevant for groups considering designing, implementing, and promoting self-governance. We then describe a process for how a diverse group of stakeholders could develop and define standards for promoting trust, as well as AI risk-mitigating practices through greater industry self-governance. We also describe how adherence to such standards could be verified, specifically through certification/accreditation. Self-governance could be encouraged by governments to complement existing regulatory schema or legislative efforts to mitigate AI risks. Greater adoption of industry self-governance could fill a critical gap to construct a more comprehensive approach to the governance of AI solutions than US legislation/regulations currently encompass. In this more comprehensive approach, AI developers, AI users, and government/legislators all have critical roles to play to advance practices that maintain trust in AI and prevent another AI Winter.

Key words: artificial intelligence/ethics, artificial intelligence/organization and administration, certification, accreditation, policy making

INTRODUCTION

Artificial intelligence (AI) has been touted as critical to harnessing value from exponentially growing health and healthcare data. AI can be used for information synthesis, clinical decision support, population health interventions, business analytics, patient self-care and engagement, research, and many other use cases. Clinician, patient, and investor expectations are high for AI technologies to effectively address contemporary health challenges.

However, prior periods of AI enthusiasm were followed by periods of disillusionment, known as "AI Winters," where AI investment and adoption withered.[1] We are now at risk of another AI Winter if current heightened expectations for AI solutions are not met by commensurate performance. Recent examples that highlight the growing concern over inappropriate and disappointing AI solutions include racial bias in algorithms supporting healthcare decision-making,[2,3] unexpected poor performance in cancer diagnostic

support,[4] or inferior performance when deploying AI solutions in real-world environments.[5] Such AI risks may be considered a "public risk," denoting threats to human health or safety that are "centrally or mass-produced, broadly distributed, and largely outside the risk bearers' direct understanding and control."[6] The public's concerns about such risks that could contribute to a "techlash" or AI Winter have recently been documented.[7]

In a seminal report by the National Academy of Medicine (NAM), the authors detailed early evidence for promising AI solutions for use by patients, clinicians, administrators, public health officials, and researchers.[1,8–12] In this article, we expand on that work by identifying 10 groups of widespread AI risks and 14 groups of recently identified mitigation strategies aligned to NAM's AI implementation life cycle.

While AI governance efforts have been proposed previously,[13,14] it remains unclear who (eg, government vs private sector/industry) is best positioned or likely to take specific actions to manage AI risks and ensure continued trust across a broad spectrum of AI solutions. The need for industry self-governance, which refers to the collective, voluntary actions of industry members, typically arises from broad societal concerns and public risks that governments may not be adequately addressing in their legislative or regulatory efforts.[15] In this manuscript, we describe how AI risk mitigation practices could be promulgated through strengthened industry self-governance, specifically through certification and accreditation of AI development and implementation organizations. We also describe how such self-governance efforts could complement current government regulations and tort law to maintain trust in a broad spectrum of AI solutions for clinical, population health, research, healthcare management, patient self-management, and other applications.

AI risks and mitigation practices across the AI implementation life cycle

The recent NAM report on AI & Health described an AI implementation life cycle that can serve as an organizing schema to understand specific AI risks and mitigation practices. Figure 1 illustrates the 4-phase NAM AI implementation life cycle. Phase 1 defines clinical and operational requirements, documents the current state, and identifies critical gaps to be filled by AI development. Phase 2 encompasses the development and validation of AI algorithms for a specific use case and context. Phase 3 focuses on organizational AI implementation. Phase 4 focuses on continued maintenance and sustainment of implemented AI.

We have summarized evidence for 10 groups of AI risks and 14 groups of associated evidence-based mitigation practices aligned to each phase of the NAM Life cycle in Table 1. While it is beyond the scope of this manuscript to provide an exhaustive summary of the relevant literature, Table 1 can serve as a convenient summary for stakeholders interested in translating evidence-based practices into future performance standards.

STRENGTHENING INDUSTRY SELF-GOVERNANCE TO PROMOTE TRUST-ENHANCING PRACTICES

Evidence-based AI risk mitigation practices should be more widely implemented by AI developers and implementers. Wider implementation could be ensured through government regulation of AI. However, such regulation is largely lacking in the US and elsewhere.[66] Additionally, an initial group of AI developers, implementers, and other stakeholders could create new market expectations through collective, voluntary actions—industry self-governance—to identify, implement, and monitor adherence to risk mitigation practice standards.[67]

Industry self-governance can be contrasted with organizational self-governance. Organizational self-governance refers to the policies and governance processes that a single organization relies on to provide overall direction to its enterprise, guide executive actions, and establish expectations for accountability. Many prominent organizations have publicly declared their adoption of select, trust-enhancing AI risk mitigation practices that we described in the previous section. At the same time, there is divergence between these organizations about both what constitutes "ethical AI" and what should be considered best practices for its realization.[68] Poor execution of organizational self-governance can result in damage to the institutional brand—and potentially open the organization to liability.[69,70] It has been argued that a society's exclusive reliance on organizational self-governance processes is unlikely to effectively ameliorate AI risks.[71,72]

Relying on industry self-governance in defining and monitoring adherence can offer several advantages. It has the potential to act faster and with greater technical expertise than government in

Figure 1. NAM AI/ML implementation life cycle.

Adapted and reproduced from: National Academy of Medicine. 2020. NAM Special Publication: *Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril*. Reproduced with permission from the National Academy of Sciences, Courtesy of the National Academies Press, Washington, DC.

Table 1. AI risks and mitigation practices across the AI implementation cycle

NAM Life cycle	Risks	Evidence-based practices
Phase 1: Needs Assessment	• Lack of integration of stakeholder perspectives & considerations[16–22] • Lack of clearly defined organizational values & ethics[23,24]	• User-centered design[25,26] • Organizational readiness assessment[27–29] • Organizational prioritization process[1] • User-centered workflow/change management process[5,30–32]
PHASE 2: Development	• Data bias[33–38] • Lack of representative & equitable population[33,39] • Lack of data management[37,40] • No accounting for causal pathways[41]	• Data transparency & reporting[32,37,40,42–46] • Model provenance records[40] • Promoting trust & explainability[32,47–51] • Distributed model development[52]
PHASE 3: Implementation	• Lack of data encryption & privacy protections[53,54] • Lack of secure hardware • Lack of oversight for responsible AI adoption[39,55]	• Equitable/diverse workforce • Organizational implementation[38,46,47,56,57] • Organizational governance[13,47,58,59] • Promote "human in the loop" practices[60,61]
PHASE 4: Maintenance	• Lack of algorithmic accountability[47,62]	• Performance surveillance[33,63,64] • Organization surveillance governance[65]

defining and enforcing standards for products and services. It may also be more insulated from partisan politics, which can lead to legislative or regulatory deadlocks. Increased reliance on "regulatory oversight" through self-governance that is monitored by regulators has been proposed as a modernized approach to regulation in the age of rapidly evolving health technologies.[73] Finally, in contrast to most government regulation, industry standards and enforcement mechanisms can reach across national jurisdictions to define and transparently enforce standards for products and services with global reach, such as AI.[67]

There is precedence for industry self-governance in the US healthcare sector. For example, a number of private sector healthcare accreditation and certification programs (eg, Joint Commission [JC] and National Committee for Quality Assurance [NCQA] accreditation, ISO9000 certification, Baldridge awards, etc) independently define and verify adherence to practice standards by hospitals, health plans, and other healthcare organizations, with accountability for patient safety and healthcare quality. In these efforts, private sector independent organizations, collaborate with healthcare industry organizations (eg, health plans or hospitals) and other experts to define relevant standards and performance metrics to improve healthcare safety and quality performance. These standards and metrics are based on research evidence, when available, or expert consensus when evidence is lacking or impractical to obtain. Additionally, these organizations also assess adherence to standards and measure performance through established, industry-vetted metrics. Due to the rigor and widespread use of these standards throughout the private-sector healthcare industry, government-run healthcare facilities (eg, Military Health Treatment facilities or Veterans Affairs Medical Centers) have adopted the same industry-defined standards and performance metrics. Similarly, the Centers for Medicaid and Medicare Services (CMS) condition payment/reimbursement of Medicare Advantage plans or healthcare facilities on the adherence to NCQA and JC standards and performance metrics. CMS's deeming authority grants JC and NCQA the ability to demonstrate that their hospital and health plan clients meet or exceed CMS's own standards for safety/quality. Once that has been demonstrated, JC or NCQA accreditation/certification is accepted by CMS in lieu of the agency inspecting these health organizations itself.

To counter growing mistrust of AI solutions,[65,74] the AI/health industry could implement similar self-governance processes, including certification/accreditation programs targeting AI developers and implementers. Such programs could promote standards and verify adherence in a way that balances effective AI risk mitigation with the need to continuously foster innovation. Moreover, as described above in the instances of JC and NCQA, adherence to these standards could be equally expected of private and government-run AI developers and implementers.

PROMOTING AI RISK-MITIGATING PRACTICES THROUGH CERTIFICATION/ACCREDITATION

Based on other certification and accreditation programs referenced earlier, we next describe essential steps for the implementation of an AI industry self-governed certification or accreditation program. These steps are summarized in Figure 2 and explained in more detail below:

- *Multistakeholder participation:* Self-governance efforts requiring trust by a broad set of stakeholders must incorporate multiple perspectives. Stakeholders may include consumers/patients, clinicians and institutional providers, healthcare administrators, payors, AI developers, and relevant governmental agencies. Stakeholders could be effectively convened by an independent third-party organization (eg, a nonprofit organization) that has expertise in the field and enjoys the trust of all stakeholders. For example, the Consumer Technology Association has suggested potential standards for AI health solutions.[75] A governing board of this organization should include representatives of all critical stakeholder groups in order to be credible and ensure that all perspectives are appropriately represented in a certification/accreditation program. Moreover, the organization's governing board should also provide guidance to multiple committees for specific, detailed elements of the overall program (eg, standard

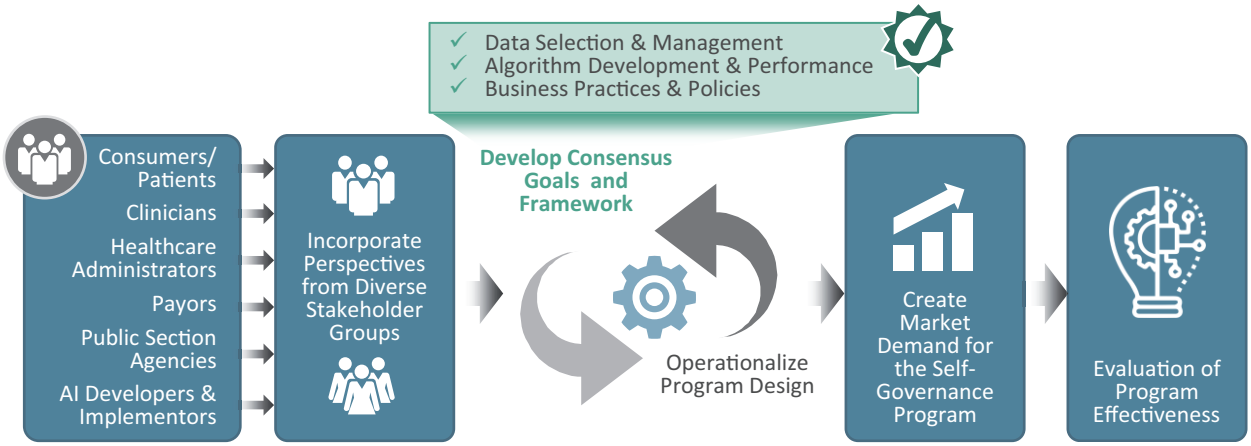

Figure 2. Steps to implement an accreditation/certification program.

development, performance metrics development, assessment/accreditation decisions, etc).

Such an independent third-party organization could be a well-known, already established organization in a particular country, or an international organization with significant expertise that is able to operate in multiple jurisdictions. For example, the Institute for Electrical and Electronics Engineers (IEEE) has more than 417 000 members in over 160 countries and has long-standing experience in defining internationally adopted standards. It recently launched a Global Initiative on Ethics of Autonomous and Intelligent Systems and issued an iterative playbook of standards and best practices called, "Ethically Aligned Design," which is intended to inform governments, organizations, businesses, and stakeholders around the world.[14] To date, IEEE has not established a certification/accreditation program for AI developers and implementers. In addition, the World Economic Forum has also issued a model AI governance framework and assessment guide to be piloted around the globe.[76]

- *Develop consensus goals and framework:* A stakeholder-consented framework to enhance trust in AI and certification/accreditation program goals must be developed to promote and verify effective implementation of risk-mitigation practices. Table 1 describes potential elements of such a framework that identifies AI risks and mitigation practices along an AI implementation life cycle. The formulation of an enduring framework and overarching program goals will allow for a careful and regular evolution of specific standards and assessment methods that is synchronized with the framework and program goals.
- *Operationalize program design:* Accreditation typically ensures adherence to a wide range of diverse standards, whereas certification may refer to a smaller, narrower group of standards. For example, AI accreditation could refer to adherence to all standards of a comprehensive framework, whereas certification could be achieved for only a subset. In either case, several elements will require careful consideration by an accreditation/certification entity, including the following:
  - *Determine the certifiable/accreditable entity*. Clear definitions of the certifiable/accreditable entity must be identified. Should an organization, a specific program within the organization, or a product developed by the organization, be certified or accredited? Should both AI developers and implementers be certified/accredited and based on what group of standards? Moreover, the definition of the accreditable entity should be clearly operationalized and have reasonable stability over time. For example, defining the certifiable/accreditable entity at a product level may be challenging, as certain AI products may evolve in relatively short periods of time. Fundamental product change over short periods may run counter to rendering meaningful certifications/accreditation decisions, which typically are meant to be valid for much longer periods (eg, 2–3 years) and based on an assessment at a particular point in time.
  - *Define standards*. A range of standards should be defined in accordance with an overarching framework and program goals. In Figure 1, we have identified a framework which aligns evidence for groups of standards for each phase of the AI implementation life cycle. Within each phase, individual standard groups can be identified based on evidence that makes up the "group" of standards for that phase. When defining standards, it is also important to define specific elements that an assessor must verify to determine if that standard has been met. It is plausible that different sets of standards might apply to AI development organizations and AI-implementing organizations, respectively, based on their different range of activities along the AI implementation life cycle. Organizations that both develop and implement AI solutions (eg, a large health system with resources and know-know to both develop and implement AI solutions) might be subject to a combined set of standards.
  - *Measure adherence to standards and practices*. A measurement system must be developed that allows for an independent verification of whether entities have met the standards. For instance, it must be determined what "evidence" is required to measure how a standard has been met (eg, review of submitted documents, calculation of submitted performance measures, onsite observation, etc). Additionally, processes must be implemented to ensure measurement methods are (1) valid (eg, assessment accurately verifies adherence to a standard/practice); (2) reliable (eg, different reviewers reach the same result); and (3) the least burdensome.
  - *Establish periodicity for recertification or accreditation*. AI organizations, programs, methods, and products advance rapidly. A viable certification/accreditation program must measure adherence to standards of a rapidly evolving industry. It also must strike the right balance between ensuring

meaningful adherence standards without stifling ongoing innovation and improvements over time.
    - *Continuously review standards and methods.* Standards and assessment methods should be dynamic and adapt to evolving practices. Additionally, certification/accreditation programs may become more stringent and rigorous over time as experience increases with standards, assessment methods, and shifting practices.
- *Create market demand:* The likelihood of effective industry self-governance depends on several factors. This includes, but is not limited to, the extent to which demand for firms' products or services relies on their brand quality or the probability of collective action by stakeholders to exert pressure on an industry to address perceived risk.[15] Verified adherence to best practices through certification/accreditation can improve AI developers' and implementers' brand through the ability to publicize adherence to a "good housekeeping" seal of approval. For example, being branded as a trusted developer and user of AI products or services may increase demand from customers, including hospitals, health systems, health plans, physician practices, and individuals. A similar approach helped establish health plan accreditation in the mid-1990s, when some large employers began demanding that health plan products they intended to purchase on behalf of their employees meet the criteria or standards for best practices established by NCQA.[77]

  The public sector (ie, federal, state, and local entities), in their roles as either payors or regulators, can similarly promote market demand by giving preferential treatment to AI developers and implementers adhering to private sector defined and implemented accreditation/certification programs. To accomplish this, US government agencies could exercise deeming authority by recognizing private sector certification/accreditation programs that ensure adherence to AI best practices, in lieu of submitting their products or services separately to a public sector review. For example, US hospitals accredited by a private-sector organization, such as the JC, can elect to be "deemed" as meeting CMS requirements by submitting to the review process of that private sector accrediting entity. The public sector can also gradually increase the expectations of what private sector accrediting organizations must address to be deemed.[78]
- *Evaluation of program effectiveness.* Finally, certification/accreditation programs should be evaluated to ensure they meet their objective of increasing trust and adherence to best practices. Such evaluation can help determine if the program continues to meet critical private and public sector policy goals for more responsible AI development and implementation. If it is determined that the certification/accreditation program is not effective in managing AI risks, industry or government can decide to strengthen the program or market conditions that would make the program more effective.

INDUSTRY SELF-GOVERNANCE, REGULATION, AND LIABILITY

To date, the rise of AI has largely occurred in a regulatory and legislative vacuum. Apart from a few US states' legislation regarding autonomous vehicles and drones, few laws or regulations exist that specifically address the unique challenges raised by AI.[66]

Industries across the globe have at times defined, adopted, and verified their adherence (eg, certification/accreditation) to beneficial standards in lieu of or as a complement to government regulation. When effective, industry efforts of defining, adopting, and verifying adherence to needed standards, can reduce the urgency of regulation through the public sector and afford the opportunity to invest limited public resources otherwise.[14] Industry self-governance has the additional advantage of being able to establish standards for globally distributed products and services across jurisdictions, reducing the potential of inconsistent regulations, as well as the need and resources potentially required to achieve international harmonization of government regulations at a later point.

If industry self-governance is lacking or relevant legislation or political will already exists, and resources are available, government agencies can reserve the right to institute their own AI programs. One example of a government-implemented program that incorporates several of the aforementioned elements is the US Food and Drug Administration's (FDA) software as a medical device (SaMD) certification program.[13] In this voluntary program, SaMD developers who rely on AI in their software are assessed and certified by demonstrating an organizational culture of quality and excellence and a commitment to ongoing monitoring of software performance in practice.[79–82] However, AI-enabled SaMD represents only a small portion of AI solutions deployed in health and healthcare. Others have suggested that additional legislation or efforts may be needed to manage AI risks across a broader range of AI health solutions. For example, it has been suggested that an Artificial Intelligence Development Act (AIDA) is needed to task an organization or government agency with certifying the safety of a broad range of AI products/systems across industry sectors.[66]

Approaches towards establishing greater accountability for AI developers and implementers through industry self-governance programs or regulation do not obviate the need for addressing legal liability. Unlike an accrediting organization or regulatory agency which would typically become active before harm from AI products occurs, courts are reactive institutions as they apply tort law and adjudicate liability in individual cases of alleged harm. To date, courts have not developed standards to specifically address who should be held legally responsible if an AI technology causes harm.[66] Consequently, established legal theory would likely hold providers who rely on AI liable for malpractice in individual cases if it is proven that they owed (1) a professional duty to a patient; (2) that they were in breach of such duty; (3) that that breach caused an injury; and (4) that there were resulting damages.[83] In order to establish legal links between certification and liability, AIDA could stipulate a certification scheme under which designers, manufacturers, sellers, and implementers of certified AI programs would be subject to limited tort liability, while uncertified programs that are offered for commercial sale or use would be subject to stricter joint and severable liability.[66] A more in-depth exploration of legal liability is beyond the scope of this article, but both liability and self-governance can promote greater accountability for ameliorating AI risks.

CRITICAL CONSIDERATIONS FOR EFFECTIVE SELF-GOVERNANCE

There are a number of critical success-factors, as well as risks, or potentially unintended consequences that need to be considered and mitigated when relying on industry self-governance as a complement to other legislative or regulatory efforts to foster responsible use of AI.

In the US, the FDA is, as described earlier, currently offering certification for AI solutions, such as medical devices.[13] However, the FDA's current authority does not extend to most types of AI solutions supporting health/healthcare needs such as population health management, patient/consumer self-management, research/development, healthcare operations, etc. At the same time, some of the most prominent failures of AI solutions to deliver on their promise, therefore jeopardizing trust, pertain to AI solutions not covered by the FDA.[2–5] This large segment of highly visible AI solutions in health/healthcare may be an appropriate focus for self-governance efforts to maintain trust.

While self-governance efforts in health/healthcare have proven to be successful in complementing legislative or regulatory efforts, several risks to effective self-governance should be managed carefully. Generally speaking, self-governance will fall short when the costs of self-governance to industry are higher than the alternatives. For example, success of self-governance may be less likely if the following conditions aren't present or are not being created: a) the public sector signaling pending legislative actions to establish greater accountability for AI health solutions (eg, through expanded regulatory authority), and that government would accept self-governance programs in lieu of implementing its own programs to ensure accountability; b) perceived public pressure (eg, through public media) on industry to create more trustworthy products; c) private and public sector commitment to preferentially purchase AI solutions that have been certified/accredited; and d) a prominent initial (small) set of organizations (AI developers/users) willing to collaborate under the auspices of an independent organization to define standards and hold themselves accountable to them, thereby creating a market expectation for certification/accreditation for AI health solution developers or implementers. Since many private companies, research institutions, and public sector organizations have issued principles and guidelines for ethical AI, there may be a significant number of organizations interested in initiating such self-governance efforts.[68]

Importantly, self-governance is likely only successful if all stakeholders have confidence that standards and verification methods were developed by appropriately balancing perspectives of consumers/patients, clinicians, AI developers, AI users, and others. To that end, as described earlier, it is imperative that a third party, independent organization (eg, rather than a trade organization representing 1 stakeholder group), is charged with the development of standards and verification methods. Balanced development/oversight processes, resulting in meaningful and operationally "achievable" performance standards, avoid the risk of standards/verification methods being perceived as self-serving for industry. However, standards need to be created that don't stifle innovation by being unnecessarily restrictive or by creating "high-costs" for accreditation/certification that may deter some AI developers from continuing to develop valuable AI health solutions.

To initiate the self-governance processes through an independent organization, start-up funding by the public sector or private-sector foundations or a group of organizations may be necessary. Such funding could support the independent organization in convening stakeholders and defining an initial set of standards and verification methods. Ongoing maintenance of standards and certification/accreditation program operations would likely need to be funded by fees levied on those organizations seeking certification/accreditation. Such a model is analog to the funding/business models of other health/healthcare certification/accreditation efforts.

CONCLUSION

The advancement of AI is actively being promoted by the US government,[84–86] governments and policy makers of other countries,[87] and supranational entities (eg, the European Union).[88] However, signs of a "techlash" and the acknowledgment of disconcerting AI-related risks and challenges are also abundant.

Governmental management of public risks such as AI risks typically occurs in democratic societies through actions of the legislative, executive, and judicial branches of government. However, as described, AI-specific legislation, regulation, or established legal standards or case law largely do not exist worldwide—or they apply only to a narrow subset of AI health solutions. At the same time, many countries are hesitant to create national industrial policy approaches that may risk disadvantaging its industries during an intense global "competition" as the Fourth Industrial Revolution unfolds, dominated by smart technologies, AI, and digitalization.[89] In 2020, the US government issued a report on AI that directed federal agencies to avoid regulatory or nonregulatory actions that needlessly hamper AI innovation and growth. The report identified ensuring trust in AI as the #1 principle of stewardship of AI while encouraging reliance on voluntary frameworks and consensus standards.[90]

The AI and healthcare industry could step in to manage AI risks through greater self-governance.[14] We presented a framework to increase trust in AI that maps known AI risks and their associated, mitigating, evidence-based practices to each phase of the AI implementation life cycle. We also described how this framework could inform the standard development for certification/accreditation programs for a broad spectrum of AI health solutions that is not covered through current regulation.

Potential future legislation and regulation across the globe will, in the coming years, likely differ in terms of managing specific AI risks. However, encouraging the use of evidence-based risk mitigation practices, promulgated through self-governance and certification and accreditation programs, could be effective and efficient across national jurisdictions in promoting and sustaining user trust in AI, while staving off another AI Winter.

FUNDING

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

AUTHOR CONTRIBUTIONS

JR and MEM developed the concept and designed the manuscript; KV and EJM provided key intellectual support, and EAK provided research support and helped edit the manuscript.

DATA AVAILABILITY STATEMENT

There are no new data associated with this article. No new data were generated or analyzed in support of this research.

CONFLICT OF INTEREST STATEMENT

None declared.

REFERENCES

1. Roski J, Chapman W, Heffner J, *et al*. How artificial intelligence is changing health and health care. In: Matheny M, Israni ST, Ahmed M, Whicher D, eds, *Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril*. NAM Special Publication. Washington, DC: National Academy of Medicine; 2019. https://nam.edu/wp-content/uploads/2019/12/AI-in-Health-Care-PREPUB-FINAL.pdf Accessed 22 Oct. 2020
2. Obermeyer Z, Powers B, Vogeli C, *et al*. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 2019; 366 (6464): 447–53.
3. Johnson CY. Racial bias in a medical algorithm favors white patients over sicker black patients. *The Washington Post*. October 24, 2019. https://www.washingtonpost.com/health/2019/10/24/racial-bias-medical-algorithm-favors-white-patients-over-sicker-black-patients/ Accessed October 22, 2020
4. Ross C. IBM's Watson supercomputer recommended 'unsafe and incorrect' cancer treatments, internal documents show. *Stat+ News* June 25, 2018. https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-treatments/ Accessed 22 October, 2020
5. Beede E, Baylor E, Hersch F, *et al*. A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Honolulu, HI: Association for Computing Machinery; 2020: 1–12.
6. Huber P. Safety and the second best: the hazards of public risk management in the courts. *Columbia Law Rev* 1985; 85 (2): 277–337.
7. Knight Foundation. *Techlash? America's Growing Concern with Major Technology Companies*. 2020. https://knightfoundation.org/reports/techlash-americas-growing-concern-with-major-technology-companies/ Accessed March 16, 2021
8. Kasthurirathne SN, Vest JR, Menachemi N, *et al*. Assessing the capacity of social determinants of health data to augment predictive models identifying patients in need of wraparound social services. *J Am Med Inform Assoc* 2018; 25 (1): 47–53.
9. Contreras I, Vehi J. Artificial intelligence for diabetes management and decision support: literature review. *J Med Internet Res* 2018; 20 (5): e10775.
10. Zieger A. Will Payers Use AI to Do Prior Authorization? And Will These AIs Make Things Better? *Healthcare IT Today*. December 27th, 2018. https://www.healthcareittoday.com/2018/12/27/will-payers-use-ai-to-do-prior-authorization-and-will-these-ais-make-things-better/ Accessed 22 October, 2020
11. Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics* 2018; 34 (13): i457–66.
12. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR Ment Health* 2017; 4 (2): e19.
13. Reddy S, Allan S, Coghlan S, Cooper P. A governance model for the application of AI in health care. *J Am Med Inform Assoc* 2020; 27 (3): 491–7.
14. Butcher J, Beridze I. What is the state of artificial intelligence governance globally? *RUSI J* 2019; 164 (5-6): 88–96.
15. Mayer F, Gereffi G. Regulation and economic globalization: prospects and limits of private governance. *Bus Polit* 2010; 12 (3): 1–25.
16. Johnson KB, Wei WQ, Weeraratne D, *et al*. Precision medicine, AI, and the future of personalized health care [published online ahead of print, 2020 Sep 22]. *Clin Transl Sci* 2021; 14 (1): 86–93.
17. Wall J, Krummel T. The digital surgeon: how big data, automation, and artificial intelligence will change surgical practice. *J Pediatr Surg* 2020; 55S: 47–50.
18. Pedersen M, Verspoor K, *et al*. Artificial intelligence for clinical decision support in neurology. *Brain Commun* 2020; 2 (2): fcaa096.
19. Ting DSW, Peng L, Varadarajan AV, *et al*. Deep learning in ophthalmology: the technical and clinical considerations. *Prog Retin Eye Res* 2019; 72: 100759.
20. Solomon DH, Rudin RS. Digital health technologies: opportunities and challenges in rheumatology. *Nat Rev Rheumatol* 2020; 16 (9): 525–35.
21. Graham S, Depp C, Lee EE, *et al*. Artificial intelligence for mental health and mental illnesses: an overview. *Curr Psychiatry Rep* 2019; 21 (11): 116.
22. Liyanage H, Liaw ST, Jonnagaddala J, *et al*. Artificial intelligence in primary health care: perceptions, issues, and challenges. *Yearb Med Inform* 2019; 28 (1): 41–6.
23. van de Poel I. *Embedding Values in Artificial Intelligence (AI) Systems. Minds and Machines*. 2020. https://link.springer.com/article/10.1007/s11023-020-09537-4 Accessed October 5, 2020
24. Gerhards H, Weber K, Bittner U, Fangerau H. Machine Learning Healthcare Applications (ML-HCAs) are no stand-alone systems but part of an ecosystem - a broader ethical and health technology assessment approach is needed. *Am J Bioeth* 2020; 20 (11): 46–8.
25. Filice RW, Ratwani RM. The case for user-centered artificial intelligence in radiology. *Radiology* 2020; 2 (3): e190095.
26. Barda AJ, Horvat CM, Hochheiser H. A qualitative research framework for the design of user-centered displays of explanations for machine learning model predictions in healthcare. *BMC Med Inform Decis Mak* 2020; 20 (1): 257.
27. Miake-Lye IM, Delevan DM, Ganz DA, *et al*. Unpacking organizational readiness for change: an updated systematic review and content analysis of assessments. *BMC Health Serv Res* 2020; 20 (1): 106.
28. Alami H, Lehoux P, Denis J-L, *et al*. Organizational readiness for artificial intelligence in health care: insights for decision-making and practice. *J Health Organ Manag*2020; 35 (1): 106–14.
29. Williams I. Organizational readiness for innovation in health care: some lessons from the recent literature. *Health Serv Manage Res* 2011; 24 (4): 213–8.
30. Cai CJ, Reif E, Hegde N, *et al*. Human-Centered Tools for Coping with Imperfect Algorithms During Medical Decision-Making. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. New York: Association for Computing Machinery; 2019: 1–14; Glasgow, Scotland, Uk. doi:10.1145/3290605.3300234
31. Cai CJ, Winter S, Steiner D, Wilcox L, Terry M. "Hello AI": uncovering the onboarding needs of medical practitioners for human-AI collaborative decision-making. *Proc Acm Hum-Comput Interact* 2019; 3 (CSCW): 1–24. Article 104.
32. Asan O, Bayrak AE, Choudhury A. Artificial intelligence and human trust in healthcare: focus on clinicians. *J Med Internet Res* 2020; 22 (6): e15154.
33. Kelly CJ, Karthikesalingam A, Suleyman M, *et al*. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med* 2019; 17 (1): 195.
34. Ntoutsi E, Fafalios P, Gadiraju U, *et al*. Bias in data-driven artificial intelligence systems—An introductory survey. *WIREs Data Mining Knowl Discov* 2020; 10 (3): e1356.
35. Gianfrancesco MA, Tamang S, Yazdany J, *et al*. Potential biases in machine learning algorithms using electronic health record data. *JAMA Intern Med* 2018; 178 (11): 1544–7.
36. Lee NT, Resnick P, Barton G. Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms. Center for Technology Innovation, Brookings. 2019. https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/ Accessed 22 October, 2020
37. Hernandez-Boussard T, Bozkurt S, Ioannidis JPA, Shah NH. MINIMAR (MINimum Information for Medical AI Reporting): Developing reporting standards for artificial intelligence in health care. *J Am Med Inform Assoc* 2020; 27 (12): 2011–5.
38. DeCamp M, Lindvall C. Latent bias and the implementation of artificial intelligence in medicine. *J Am Med Inform Assoc*. 2020; 27 (12): 2020–3.
39. Tzachor A, Whittlestone J, Sundaram L, hÉigeartaigh SÓ. Artificial intelligence in a crisis needs ethics with urgency. *Nat Mach Intell* 2020; 2 (7): 365–6.

40. Norgeot B, Quer G, Beaulieu-Jones BK, *et al*. Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist. *Nat Med* 2020; 26 (9): 1320–4. doi:10.1038/s41591-020-1041-y
41. Richens JG, Lee CM, Johri S. Improving the accuracy of medical diagnosis with causal machine learning. *Nat Commun* 2020; 11 (1): 3923.
42. Crowley RJ, Tan YJ, Ioannidis JPA. Empirical assessment of bias in machine learning diagnostic test accuracy studies. *J Am Med Inform Assoc* 2020; 27 (7): 1092–101.
43. Rivera SC, Liu X, Chan AW, Denniston AK, Calvert MJ, SPIRIT-AI and CONSORT-AI Working Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI Extension. *BMJ* 2020; 370: m3210.
44. Liu X, Cruz RS, Moher D, Calvert MJ, Denniston AK, SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Lancet Digit Health* 2020; 2 (10): e537–48.
45. Andaur Navarro CL, Damen J, Takada T, *et al*. Protocol for a systematic review on the methodological and reporting quality of prediction model studies using machine learning techniques. *BMJ Open* 2020; 10 (11): e038832.
46. Mongan J, Moy L, Charles E, Kahn J. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): a guide for authors and reviewers. *Radiology* 2020; 2 (2): e200029.
47. Hunt R, McKelvey F. Algorithmic regulation in media and cultural policy: a framework to evaluate barriers to accountability. *J Inform Policy* 2019; 9: 307–35.
48. Payrovnaziri SN, Chen Z, Rengifo-Moreno P, *et al*. Explainable artificial intelligence models using real-world electronic health record data: a systematic scoping review. *J Am Med Inform Assoc* 2020; 27 (7): 1173–85.
49. Zednik C. Solving the black box problem: A normative framework for explainable artificial intelligence. *Philos. Technol* 2019; https://doi.org/10.1007/s13347-019-00382-7.
50. Ribeiro MT, Singh S, Guestrin C. "Why should i trust you?": explaining the predictions of any classifier. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, CA: Association for Computing Machinery; 2016:1135–44.
51. Phillips PJ, Hahn AC, Fontana PC, *et al*. (2020). *Four Principles of Explainable Artificial Intelligence (Draft)*. https://doi.org/10.6028/NIST.IR.8312-draft Accessed 22 October, 2020
52. Jiang X, Wu Y, Marsolo K, *et al*. Development of a web service for analysis in a distributed network. *EGEMS (Wash DC)* 2014; 2 (1): 1053.
53. Rastogi N, *et al*. Security and privacy of performing data analytics in the cloud: a three-way handshake of technology, policy, and management. *J Inform Policy* 2015; 5: 129–54.
54. Shi S, He D, Li L, *et al*. Applications of blockchain in ensuring the security and privacy of electronic health record systems: a survey. *Comput Secur* 2020; 97: 101966–20.
55. Mudgal KS, Das N. The ethical adoption of artificial intelligence in radiology. *BJR Open* 2020; 2 (1): 20190020.
56. Strohm L, Hehakaya C, Ranschaert ER, Boon WPC, Moors EHM. Implementation of artificial intelligence (AI) applications in radiology: hindering and facilitating factors. *Eur Radiol* 2020; 30 (10): 5525–32.
57. Petitgand C, Motulsky A, Denis JL, Régis C. Investigating the barriers to physician adoption of an artificial intelligence- based decision support system in emergency care: an interpretative qualitative study. *Stud Health Technol Inform* 2020; 270: 1001–5.
58. Roski J, Gillingham BL, Just E, Barr S, Sohn E, Sakarcan K. Implementing and scaling artificial intelligence solutions: considerations for policy makers and decision makers. *Health Aff Blog*. September 18, 2018; doi: 10.1377/hblog20180917.283077. https://www.healthaffairs.org/do/10.1377/hblog20180917.283077/full/ Accessed April 8, 2021.
59. Sohn E, Roski J, Escaravage S, Maloy K. Four lessons in the adoption of machine learning in health care. *Health Aff Blog*. *May 9*, 2017; doi: 10.1377/hblog20170509.059985. https://www.healthaffairs.org/do/10.1377/hblog20170509.059985/full/ Accessed April 8, 2021.
60. Holzinger A, Plass M, Kickmeier-Rust M, *et al*. Interactive machine learning: experimental evidence for the human in the algorithmic loop. *Appl Intell* July 2019; 49 (7): 2401–14.
61. Lee D, *et al*. A human-in-the-loop perspective on AutoML: milestones and the road ahead. *IEEE Data Eng. Bull* 2019; 42: 59–70.
62. Diakopoulos N. Algorithmic Accountability Reporting: on the Investigation of Black Boxes. 2014. http://academiccommons.columbia.edu, doi:10.7916/D8ZK5TW2
63. Subbaswamy A, Saria S. From development to deployment: dataset shift, causality, and shift-stable models in health AI. *Biostatistics* 2020; 21 (2): 345–52.
64. Davis SE, Greevy RA, Lasko TA, Walsh CG, Matheny ME. Comparison of prediction model performance updating protocols: using a data-driven testing procedure to guide updating. *AMIA Annu Symp Proc* 2019; 2019: 1002–10.
65. Eaneff S, Obermeyer Z, Butte AJ. The case for algorithmic stewardship of artificial intelligence and machine learning technologies. *JAMA* 2020; 324 (14): 1397.
66. Scherer MU. Regulating artificial intelligence systems: risks, challenges, competencies, and strategies. *Harvard J Law Technol* 2015; 29: 353.
67. Maurer SM. The new self-governance: a theoretical framework. *Bus Polit* 2017; 19 (1): 41–67.
68. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. *Nat Mach Intell* 2019; 1 (9): 389–99.
69. "Exclusive: Google Cancels AI Ethics Board in Response to Outcry". Vox. 2019. https://www.vox.com/future-perfect/2019/4/4/18295933/google-cancels-ai-ethics-board Accessed September 23, 2020
70. D'Onfro J. Google Scraps Its AI Ethics Board Less Than Two Weeks After Launch in the Wake of Employee Protest. *Forbes*. https://www.forbes.com/sites/jilliandonfro/2019/04/04/google-cancels-its-ai-ethics-board-less-than-two-weeks-after-launch-in-the-wake-of-employee-protest/ Accessed January 8, 2021
71. Putting Responsible AI Into Practice. https://sloanreview.mit.edu/article/putting-responsible-ai-into-practice/ Accessed January 8, 2021
72. Tiku N. Google hired Timnit Gebru to be an outspoken critic of unethical AI. Then she was fired for it. *Washington Post*. 2020. https://www.washingtonpost.com/technology/2020/12/23/google-timnit-gebru-ai-ethics/ Accessed January 8, 2021
73. Clark J, Gillian K. Hadfield. "Regulatory Markets for AI Safety." *arXiv preprint arXiv:2001.00078* 2019.
74. Scott M. "In 2020, Global 'Techlash' Will Move from Words to Action." POLITICO. 2019, https://www.politico.eu/article/tech-policy-competition-privacy-facebook-europe-techlash/ Accessed 22 Oct 2020
75. The Use of Artificial Intelligence in Health Care: Trustworthiness (ANSI/CTA-2090). Consumer Technology Association. https://shop.cta.tech/products/the-use-of-artificial-intelligence-in-healthcare-trustworthiness-cta-2090 Accessed March 16, 2021
76. Model Artificial Intelligence Governance Framework and Assessment Guide. World Economic Forum. https://www.weforum.org/projects/model-ai-governance-framework/ Accessed March 16, 2021
77. When Employers Choose Health Plans: Do NCQA Accreditation and HEDIS Data Count | Commonwealth Fund. 1998. https://www.commonwealthfund.org/publications/fund-reports/1998/aug/when-employers-choose-health-plans-do-ncqa-accreditation-and Accessed January 8, 2021
78. CMS to Strengthen Oversight of Medicare's Accreditation Organizations. CMS Newsroom. 2018. https://www.cms.gov/newsroom/press-releases/cms-strengthen-oversight-medicares-accreditation-organizations Accessed October 14, 2020
79. US Department of Health and Human Services, US Food and Drug Administration, Center for Devices & Radiological Health. Developing the Software Precertification Program: Summary of Learnings and Ongoing Activities. 2020. https://www.fda.gov/media/142107/download Accessed October 22, 2020
80. US Department of Health and Human Services, US Food and Drug Administration, Center for Devices & Radiological Health, Digital Health Program. Digital Health Innovation Action Plan. 2017. https://www.fda.gov/media/106331/download Accessed Oct 22, 2020

81. US Department of Health and Human Services, US Food and Drug Administration, Center for Devices & Radiological Health. Developing a Software Precertification Program: A Working Model (v1.0 – January 2019). 2019. https://www.fda.gov/media/119722/download Accessed October 22, 2020
82. US Department of Health and Human Services, US Food and Drug Administration, Center for Devices & Radiological Health. Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) - Discussion Paper and Request for Feedback. 2019. https://www.fda.gov/media/122535/download Accessed 22 October, 2020
83. Bal BS. An introduction to medical malpractice in the United States. *Clin Orthop Relat Res* 2009; 467 (2): 339–47.
84. Office of the President of the United States. Maintaining American leadership in artificial intelligence. Executive Order 13859. 2019. https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/ Accessed 22 October, 2020
85. Intelligence SCoA. The National Artificial Intelligence Research and Development Strategic Plan: 2019 Update. 2019. https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf Accessed Oct 22, 2020
86. Executive Office of the President of the United States; Artificial Intelligence Research & Development Interagency Working Group. 2016–2019 Progress Report: Advancing Artificial Intelligence R&D. 2019. https://www.nitrd.gov/pubs/AI-Research-and-Development-Progress-Report-2016-2019.pdf Accessed October 22, 2020
87. KI Strategie. https://www.ki-strategie-deutschland.de/home.html Accessed September 23, 2020
88. White Paper on Artificial Intelligence: A European Approach to Excellence and Trust. European Commission - European Commission. 2020. https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en Accessed September 23, 2020
89. The Fourth Industrial Revolution: what it means and how to respond. World Economic Forum. 2016. https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/ Accessed March 16, 2021
90. Memorandum M-21-06, Guidance for Regulation of Artificial Intelligence Applications. Executive Office of the President, Office of Management Budget (OMB). 2020. https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf Accessed March 16, 2021.